

Masarykova univerzita
Přírodovědecká fakulta
Geografický ústav

Geostatistika

doc. RNDr. Petr Dobrovolný, CSc.

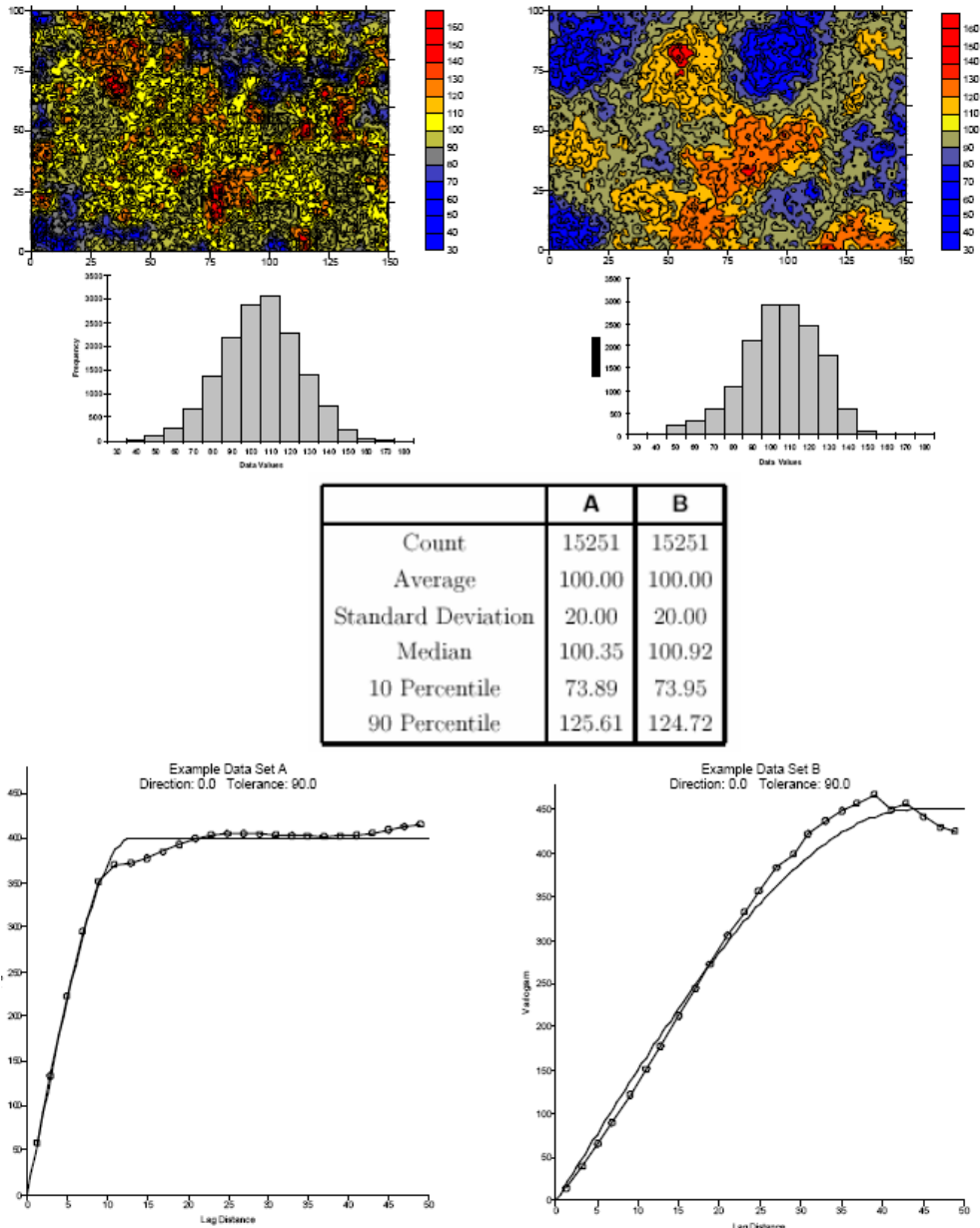
(pracovní texty)

Brno 2010

1. 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

2. Metody prostorové interpolace

2.1 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.

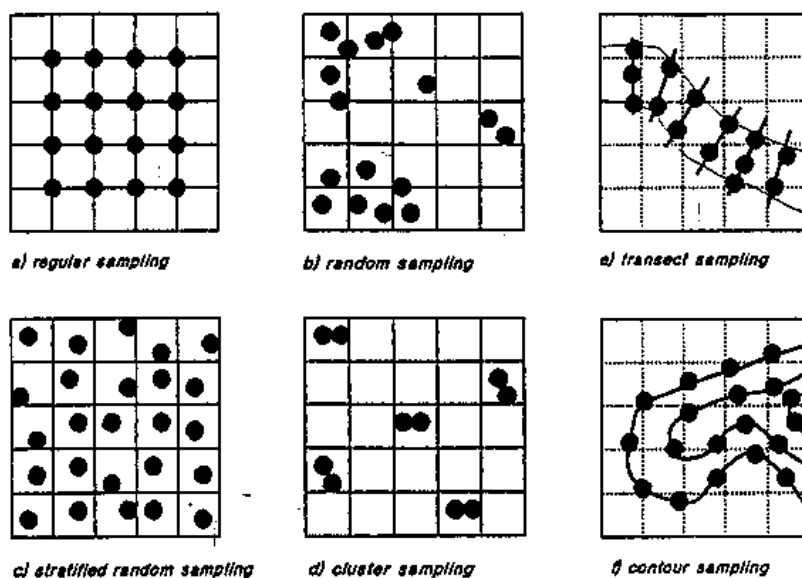
Prostorová interpolace – skupina metod, které slouží k vytváření spojitých povrchů (polí) z bodových měření. Body mohou být lokalizovány v 1, 2 i 3 rozměrném prostoru. Interpolace se může týkat nejenom bodů, ale i linií a ploch. V rámci interpolace je často řešen také problém **extrapolace** – tedy odhad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

2.1.1 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. 2.1 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 spojených 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

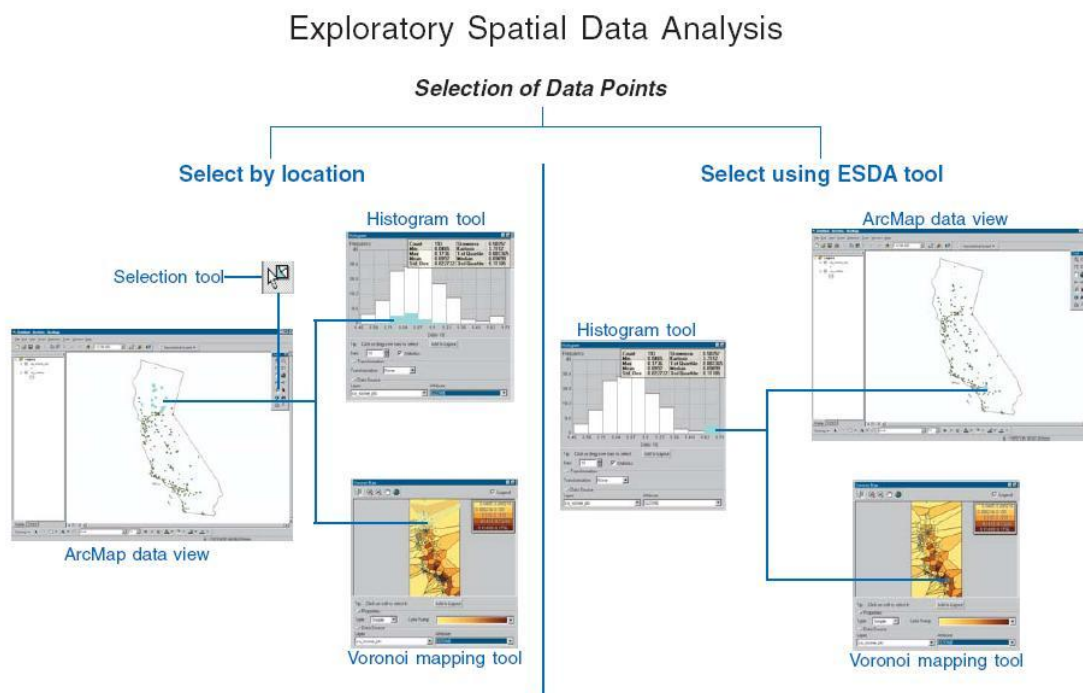
2.1.2 Průzkumová analýza dat (EDA – Exploratory Data Analysis)

- **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).



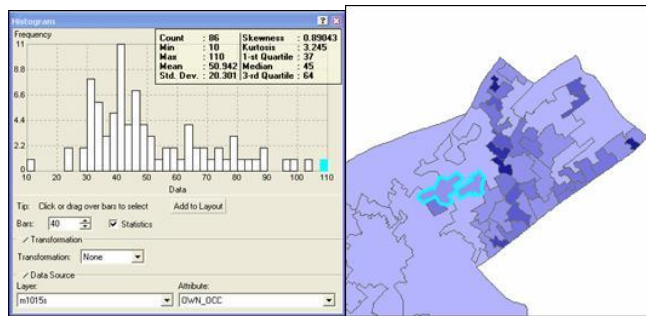
Obr. 2.2 Nástroje EDA jsou často propojeny s vlastní mapou (ESRI, Using ArcGIS Geostatistical analyst).

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

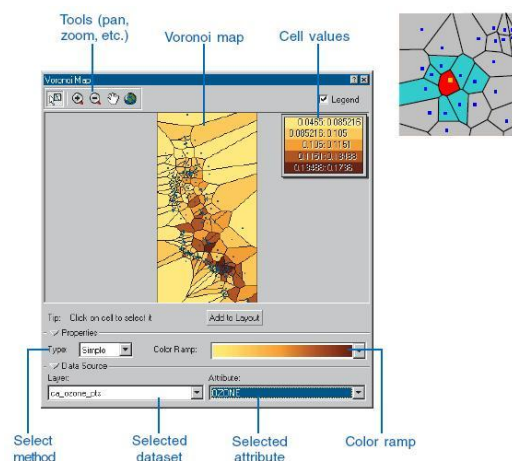


Obr. 2.3 Příklad histogramu

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.



Obr. 2.4 Příklad Voronoi mapy

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

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

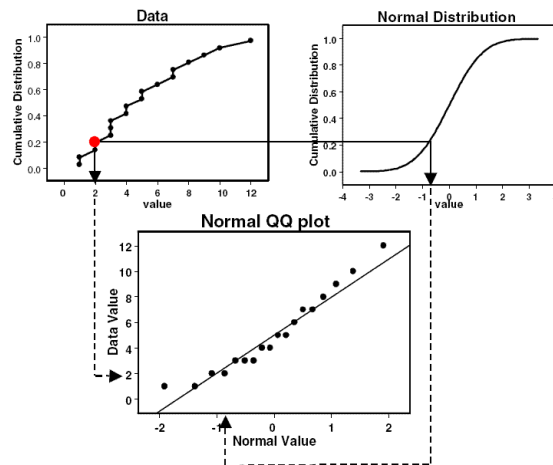
kde p_i je poměr polygonů náležejících do dané třídy z celkového počtu polygonů

Minimální entropie – všechny buňky patří do stejné třídy

Maximální entropie – každá z buněk náleží k jiné třídě.

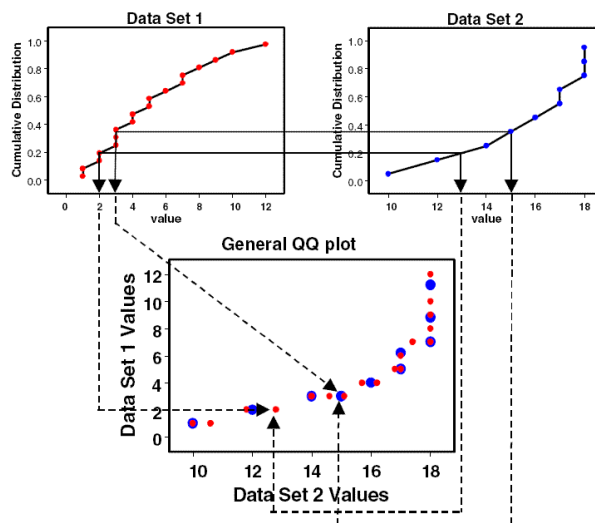
Kvantilové grafy - grafy zobrazující kvantily dvou rozdělení

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.



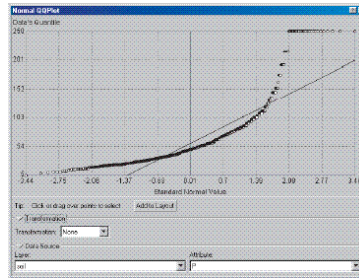
Obr. 2.5 Normální Q-Q graf

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



Obr. 2.6 Obecný Q-Q graf

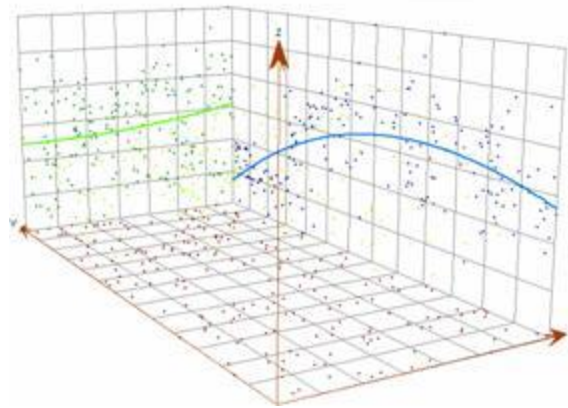
Shodu v obou případech indikují v grafech body přimykající se k přímce.



Takováto data vyžadujú transformáciu. Základní typy transformácií:

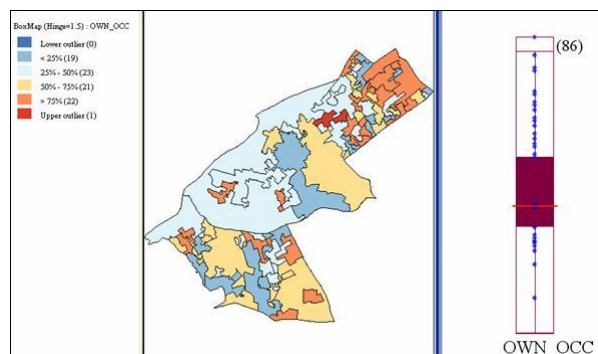
- Box-Cox
- Arcsine
- Logaritmická

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.

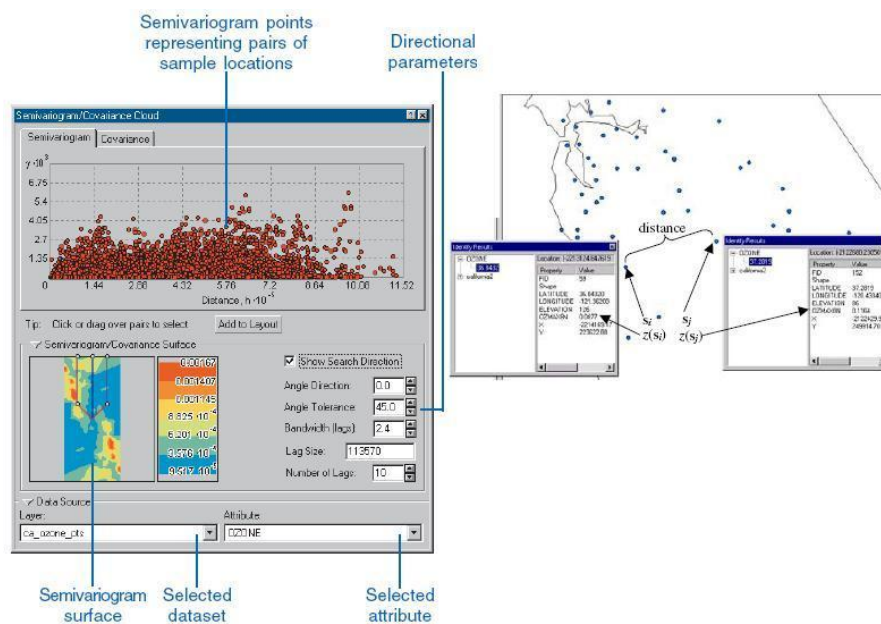
Krabicové grafy (box plots) a detekce odlehlých či extrémních hodnot – odlehlé hodnoty ve smyslu jejich polohy – odlehlé v porovnání s hodnotami okolními (local spatial outliers), nemusejí být odlehlé absolutně (global outliers) Klasickými metodami bez zahrnutí prostorového aspektu je lze těžko identifikovat – 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 odhlaení 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 :

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

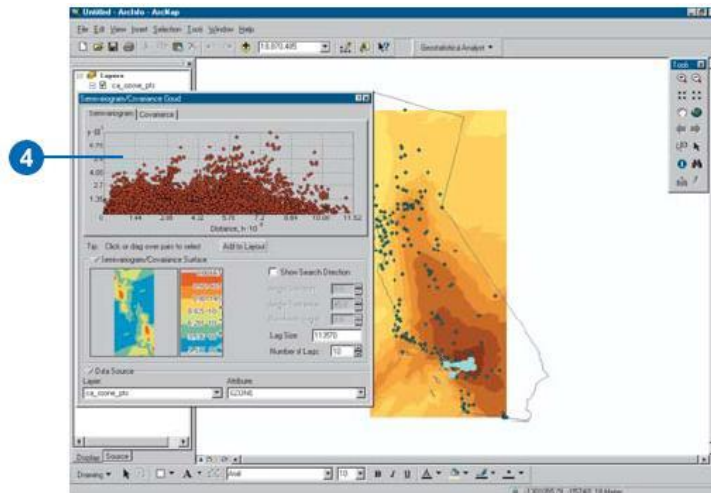
Hodnota empirické **covariance**

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

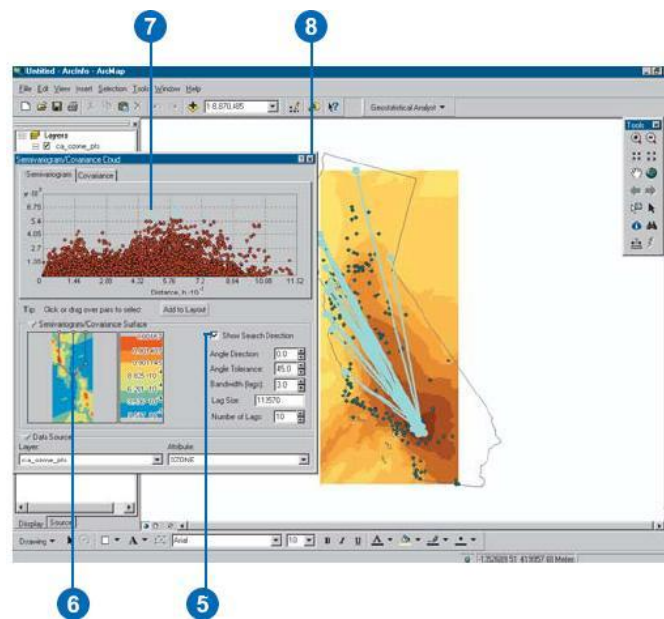
Hodnota empirické **crosscovariance**

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

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.



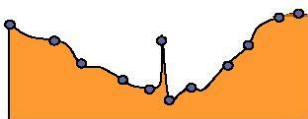
Z různých důvodů nemusí být hodnoty prostorové autokorelace obdobné s ohledem na orientaci spojnice vyšetřovaných bodů. Vysoké hodnoty semivariance mohou být vázány na body nacházející se vzájemně v určitém směru. To potom svědčí o tzv. anizotropii. Pro interpolaci to znamená asymetricky definované okolí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



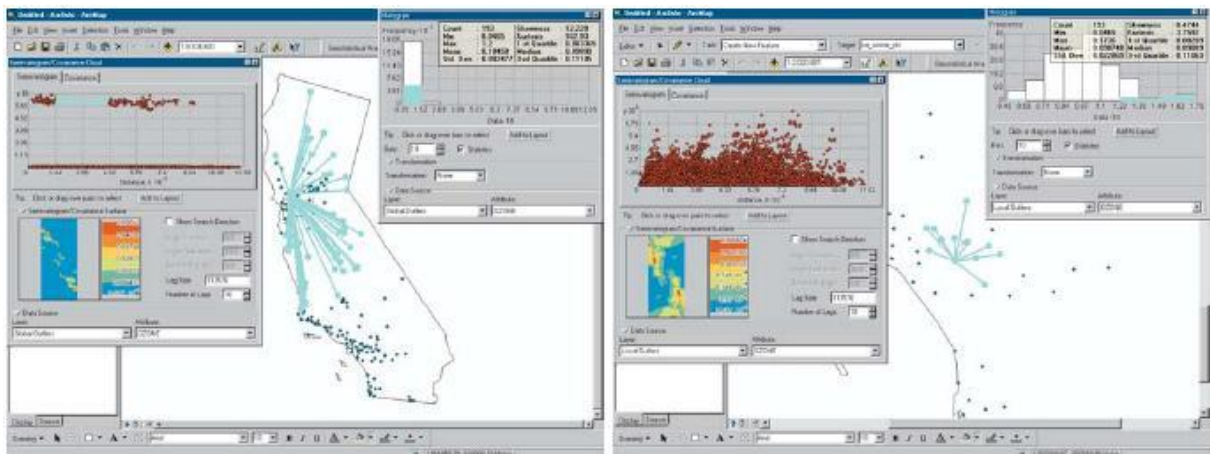
Povrch semivariogramu – povrch tvořený četnostmi (bins) hodnot semivariogramu daného směru a dané vzájemné vzdálenosti.

Directional influences - vzájemná orientace vyhledávaných bodů (search direction)

Detekce odlehlých hodnot (outliers) – globalních i lokálních



Základní nástroje: histogram
semivariogram/ covariance cloud
Voronoi map

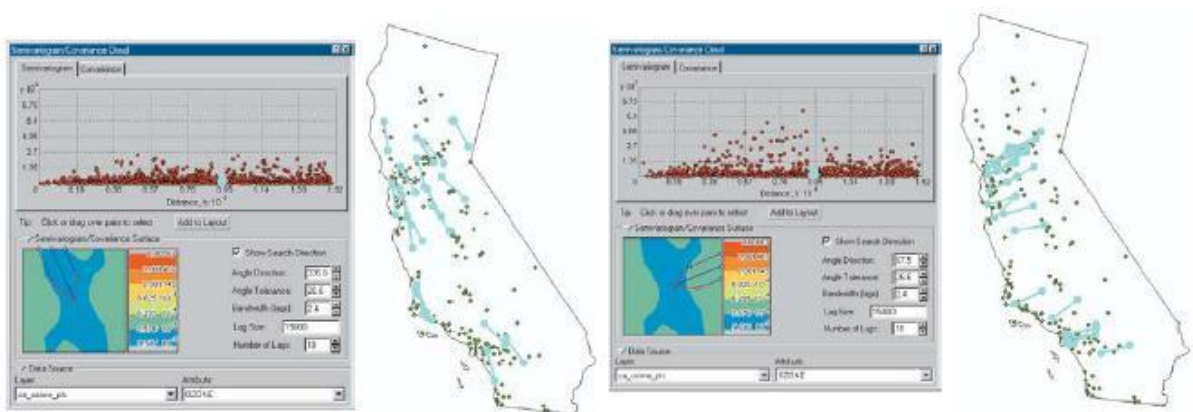


Detekce globální (vlevo) a lokální (vpravo) odlehlé hodnoty. Globální – vysoké hodnoty semivariance bez ohledu na vzdálenost a navíc se budou v mapě separovat s jedním bodem, který právě obsahuje danou odlehlou hodnotu (viz. mapa). Všechny body semivariogramu se rozdělí do dvou shluků (clouds). Globální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

2.2 Rozdělení metod prostorové interpolace

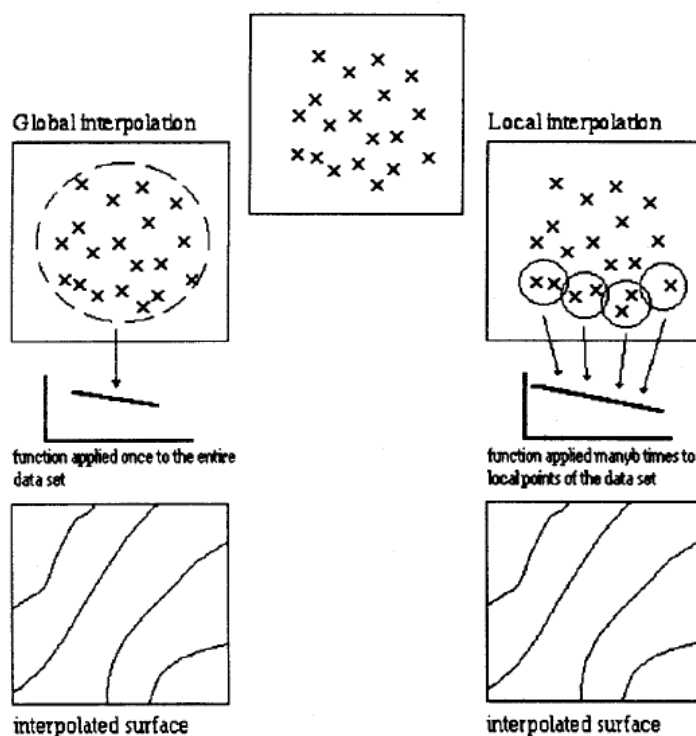
2.2.1 Podle prostorových entit, na které jsou aplikovány:

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

2.2.2 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, redukuje 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.

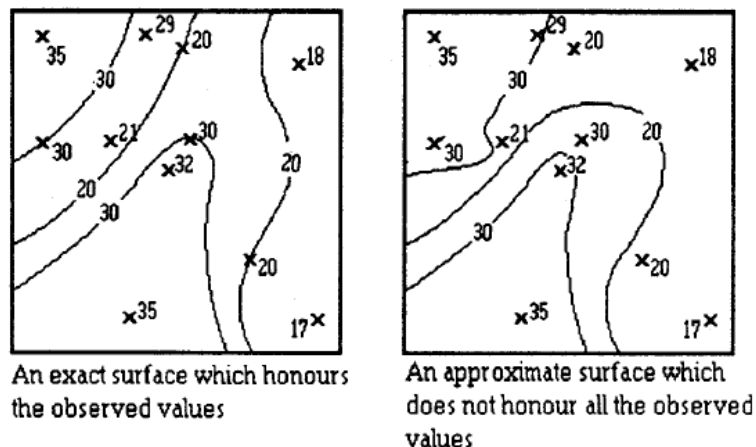


Obr. 2.7 Globální a lokální metody interpolace

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.

2.2.3 Metody exaktní a aproximující



Obr. 2.8 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.

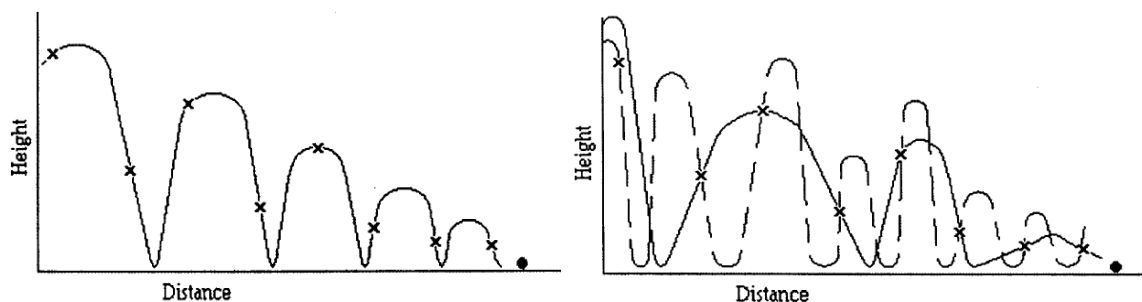
2.2.4 Metody spojité a zlomové (abrupt)

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

2.2.5 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 měřenými 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. 2.9 Deterministické a stochastické metody interpolace

2.3 Přehled interpolačních metod

2.3.1 Metody analogové interpolace (line threading or eye balling)

Jedná se o metody vytváření izoliní na základě spojování míst s obdobnými hodnotami jevu založené na expertním odhadu (geolog, synoptik). Dále využívají empirie, obecné teorie a znalosti místních zvláštností (budování expertních systémů). Základní omezení (s ohledem na počítačové zpracování):

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

2.3.2 Globální interpolátory využívající klasifikačních modelů a ANOVA analýzy

Jsou založeny na podmínce **stacionarity** výběru. To je předpoklad, že míry úrovně a variability výběrového souboru nezávisí na velikosti výběru a rozmístění jednotlivých měřených bodů.

Za výše uvedeného předpokladu lze k interpolaci v rámci studovaného území využít externě definovaných prostorových jednotek (regionů). **Klasifikace homogenními polygony** předpokládá, že rozptyl hodnot interpolovaného atributu v rámci externě definovaného regionu je menší jak mezi dvěma regiony. Tohoto přístupu se často používá při mapování půd či landscape units, ekotopů, kde jednotlivé objekty (půdní jednotky, říční terasy, dílčí povodí, svahy, ...) jsou vhodné pro definování jiných (interpolovaných) atributů o daném území.

ANOVAR 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

Nevýhody: hodnoty mohou v rámci jedné třídy kolísat více (měně) jak v rámci jiné třídy. Nelze mapovat prostorové změny ve větším měřítku. Data často nemají normální rozdělení. Potom je nutná normalizace např. pomocí transformace přirozenými logaritmy.

2.3.3 Globální interpolátory využívající analýzy trendu

Jestliže se určitá vlastnost v prostoru mění kontinuálně a je spojitá (teplota, nadmořská výška, apod.), lze body z tohoto povrchu interpolovat polynomickeou funkcí. Body v neměřených lokalitách lze vypočítat z koeficientů, vypočtených na základě bodů měřených a souřadnic bodů neměřených (interpolovaných).

Nejjednodušší způsob - mnohonásobná regrese hodnot atributu vs. geografické souřadnice. Metodou nejmenších čtverců lze nalézt nejvhodnější koeficienty pro daný polynom n -tého řádu. Předpokládá se normální rozdělení.

Předpokládejme měření studované veličiny v transektu (profilu). Jestliže hodnoty obecně rostou či klesají (zanedbáme-li náhodná kolísání) - lze hodnoty interpolovat pomocí lineárního regresního modelu:

$$z(x) = b_0 + b_1x + \varepsilon$$

b_0 a b_1 - koeficienty

ε - náhodný šum - nezávislý na hodnotách x s normálním rozdělením

Není-li povrch rovinou, ale složitějším tvarem - lze ho interpolovat polynomem vyššího řádu, např. kvadrátem:

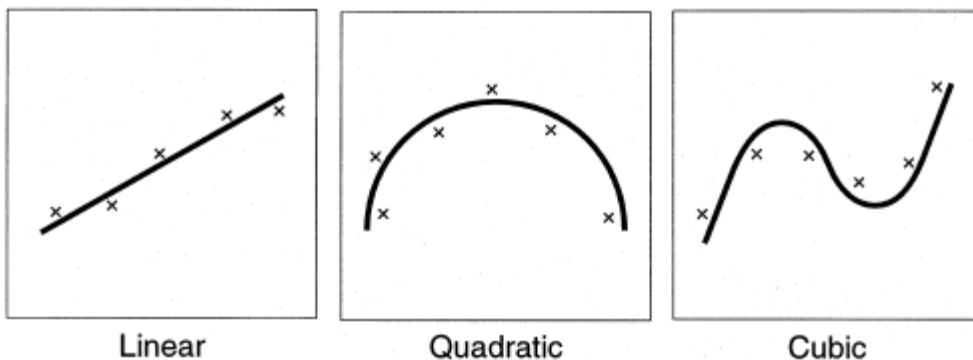
$$z(x) = b_0 + b_1x + b_2x^2 + \varepsilon$$

Zvyšováním stupně polynomu lze vystihnout složitější tvary a redukuje se náhodná složka. Uvedené rovnice platí pro 1D, ve dvourozměrném prostoru budou v rovnici začleněny obě souřadnice x, y :

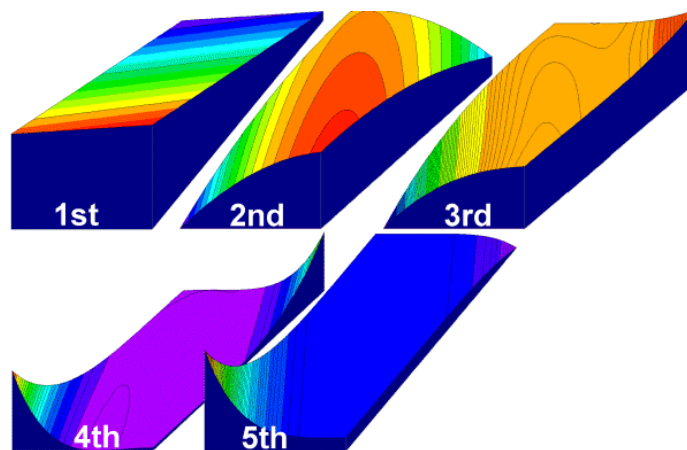
lineární trend: $z = b_0 + b_1x + b_2y$

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

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



Obr. 2.10 Proložení polynomu 1 až 3 stupně množinou měřených bodů



Obr. 2.11 Interpolace trendové složky polynomu 1 až 5 stupně

Trendový povrch prezentovaný polynomem vyššího řádu vykazuje značné chyby na okrajích zpracovávaného povrchu (edge effects). Mimo zpracovávané území může nabývat extrémních či dokonce záporných hodnot interpolované vlastnosti (nemajících fyzikální význam- např. záporná hodnota atmosférických srážek).

Jde o globální interpolátor, který zřídka prochází měřeními body a který shazuje lokální odchylky. Protože lokální odchylky jsou prostorově závislé, často se tohoto postupu využívá

k definování částí povrchu, které se významně odlišují od obecného trendu. Druhý častý způsob využití je odfiltrování obecného trendu a aplikace lokálních interpolátorů na reziduální složku prostorových změn studovaného jevu. Vypočtený trend lze testovat z hlediska jeho významnosti.

2.3.4 Globální interpolátory využívající regresní analýzy

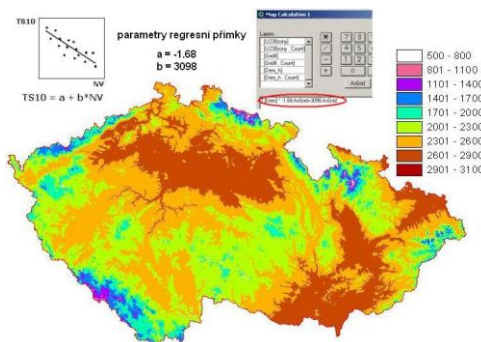
V řadě případů existuje zřejmá vazba mezi hodnotami interpolované veličiny a vybranými jinými atributy studovaného prostoru (teplota a nadmořská výška, srážky a vzdálenost od moře, koncentrace znečištění a vzdálenost od zdroje apod.). Lze tedy sestavit empirický model závislosti interpolované veličiny na hodnotách jedné či několika veličin nezávislých. Tento model má následující obecnou formu:

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

$b_0 \dots b_n$ - regresní koeficienty

$P_1 \dots P_n$ - nezávisle proměnné

Sestavení regresní závislosti je založeno na metodě nejmenších čtverců, výsledný model může být lineární i nelineární a jako nezávisle proměnné lze kombinovat geografické souřadnice s jinými atributy.



Obr. 2.12 Příklad sestavení regresního modelu závislosti teplotních sum na nadmořské výšce, zápis modelu v prostředí Map Calculator a vytvořená mapa teplotních sum pro ČR

2.4 Metody lokální interpolace (lokální interpolátory)

Výše uvedené globální interpolátory považovaly lokální efekty za náhodný šum. Lokální interpolátory využívají k výpočtu hledané hodnoty pouze 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 regulární sítě (grid)

Uvedený postup je opakován do té doby, dokud nejsou vypočteny hodnoty interpolované veličiny pro všechny uzly (buňky) gridu. Pro každý konkrétní postup lokální interpolace jsou důležité následující skutečnosti:

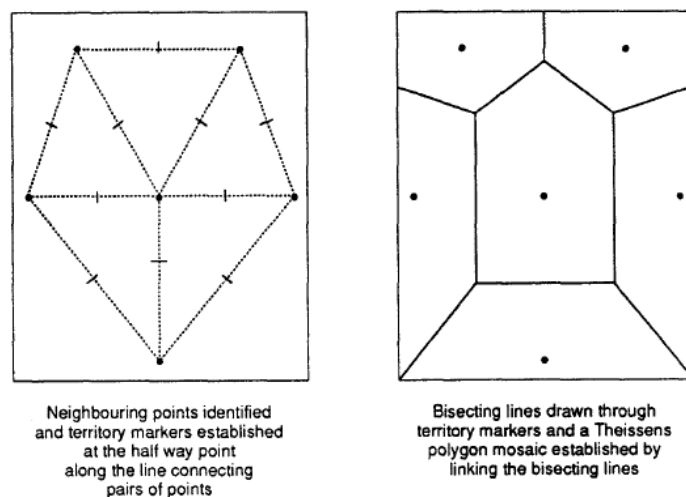
- druh použité interpolační funkce

- 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

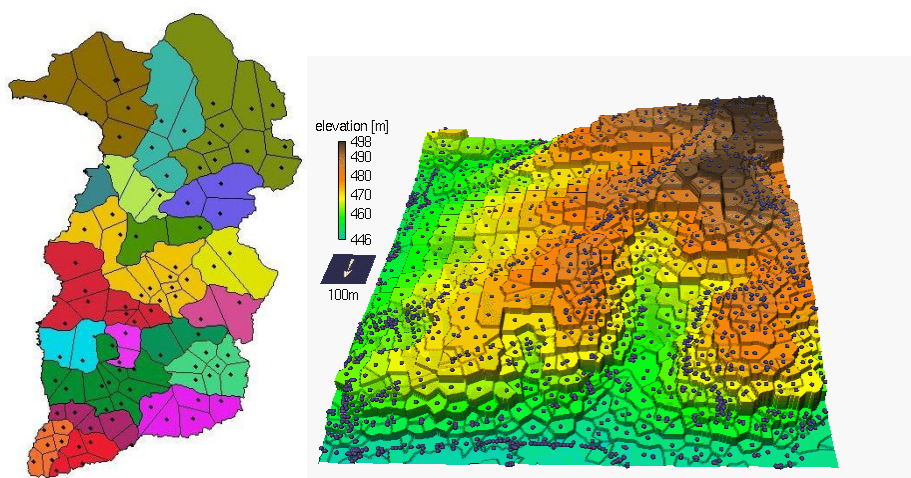
Většina lokálních interpolátorů pracuje na principu „filtrovacího okénka“, do jisté míry počítají průměrnou hodnotu z bodů v okolí či v definované vzdálenosti.

2.4.1 Metoda nejbližšího souseda (*thiessenovy polygony*)

Hodnoty atributů v neměřených místech jsou určeny z hodnot nejbližšího místa měřeného. Podle schématu uvedeného na obrázku je zpracovávané území rozděleno na nepravidelné trojúhelníky (Delaunay triangulace). Z nich jsou poté definovány tzv. thiessenovy polygony. V závislosti na rozmístění měřených dat mohou tyto polygony být pravidelné či nepravidelné. V GIS se často využívají jako rychlý prostředek pro vztažení bodu k určitému okolí. Celá metoda je založena na předpokladu např., že meteorologická data z určité oblasti mohou být určena z nejbližší meteorologické stanice. Tato metoda je nevhodná pro spojitě se měnící jevy (srážky, teplota, ...).



Obr. 2.13 Konstrukce thiessenových polygonů na pravidelně rozmístěných bodech



Obr. 2.14 Příklady interpolace množiny nepravidelně rozmístěných bodů v ploše metodou thiessenových polygonů

Lokální, exaktní metoda interpolace. Metoda původě využívaná pro plošné odhady srážek. Je to metoda robustní, vždy produkuje stejný povrch ze stejné množiny vstupních dat. Nelze při ní však použít externí informace o faktorech, které mohou ovlivňovat hodnoty v místech měření. Je vhodná k vymezení teritoria (oblasti vlivu). Forma výsledného povrchu (mapy) je determinována rozdělením původních měřených bodů. Změny v hodnotách atributů se dějí skokem, na hranicích každého polygonu. Postup však lze použít na kvalitativní data.

2.4.2 Metody konstrukce sítě nepravidelných trojúhelníků (TIN)

Exaktní metoda vhodná pro nepravidelně rozmístěné body měření. Tyto body jsou spojeny liniemi a vytváří síť nepravidelných trojúhelníků. Protože 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. Pro některé druhy povrchů je vhodná – obecně pro povrchy které se vyznačují náhlými změnami spádu (fluviálně erodované povrchy).

Proces vytváření spojitého povrchu metodou nepravidelné trojúhelníkové sítě zahrnuje:

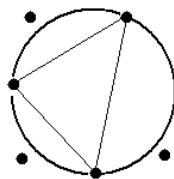
- výběr charakteristických bodů (ne z jakékoliv množiny nepravidelně rozmístěných bodů lze vytvořit TIN)
- způsob propojení bodů do trojúhelníkové sítě
- způsob modelování povrchu uvnitř trojúhelníků

Výběr bodů – body by především měly reprezentovat významné rysy terénu – zlomy, údolnice, hřbetnice. V závislosti na komplexnosti terénu může být hustota bodů značně proměnlivá. Algoritmy pro výběr bodů:

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

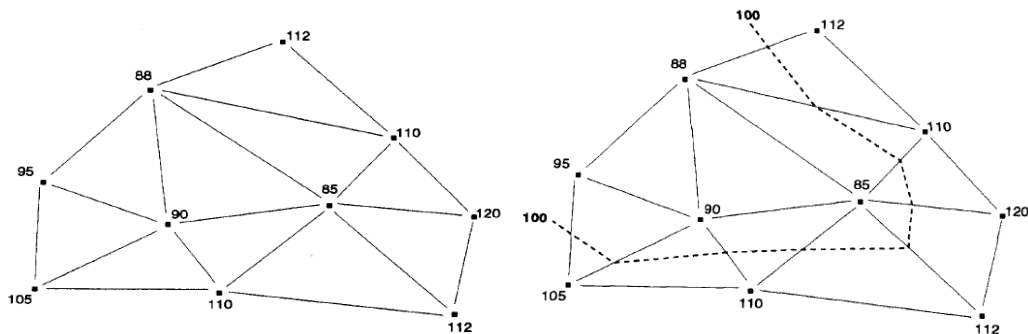
Princip algoritmů – viz <http://www.ncgia.ucsb.edu/giscc/units/u056/>

Způsob propojení bodů do trojúhelníkové sítě se řeší metodou **Delaunay triangulace**: Tři body tvoří tzv. Delaunay trojúhelník pouze v případě, pokud kružnice, která je těmito třemi body opsána neobsahuje žádný další bod. Tato podmínka zaručuje, že trojúhelníky jsou přibližně rovnostranné a jakýkoliv vnitřní bod trojúhelníka je co možná nejbližší jednomu z vrcholů – tedy bodu měření. Delaunay triangulace může být také vytvořena z Thiessenových polygonů (viz. výše).

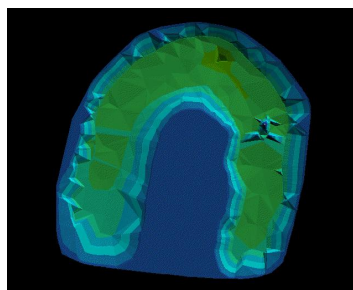


Obr. 2.15 Podmínka tzv. Delaunay triangulace

TIN je model vhodný k následné konstrukci izolinií. Nejprve se zvolí krok, se kterým budou izolinie interpolovány, poté jsou identifikovány všechny linie, které bude protínat izolinie s danou hodnotou. Poté se podél všech těchto linií vypočtou souřadnice x, y bodu „přechodu“ izočáry. Následně se body spojí. Pro „hladký“ průběh izolinií se body spojují nelineární funkcí. Metody není možné použít k extrapolaci – výsledný povrch má plochu, která vznikne spojením vnějších měřených bodů („hull“).



Obr. 2.16 Vytvoření TIN a konstrukce izolinií



Obr. 2.17 Příklad povrchu vytvořeného metodou TIN

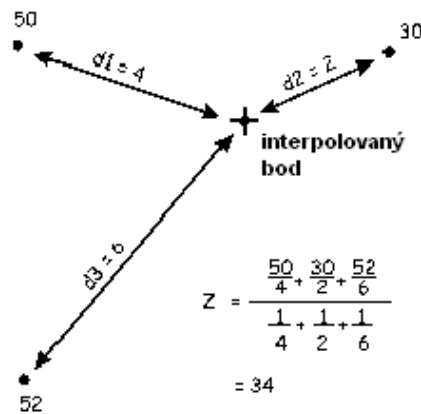
2.4.3 Metoda inverzní vzdálenosti

Tato metoda kombinuje ideu vzdálenosti využívanou v thiessenových polygonech a ideu postupných změn trendových povrchů. Je založena na předpokladu, že hodnota atributu v určitém bodě je váženým aritmetickým průměrem hodnot okolních měřených bodů. Váhy jsou určeny pro každý bod například jako inverzní vzdálenost měřeného bodu od bodu interpolovaného (čím bližší bod, tím má větší váhu). Nejjednodušším je lineární interpolátor. Jde většinou o exaktní interpolátor. Forma výsledného interpolovaného povrchu závisí na shlcích bodů a na odlehlých měřeních. Dává nejlepší výsledky při dostatečném množství měřených bodů pravidelně rozmístěných v interpolovaném prostoru.

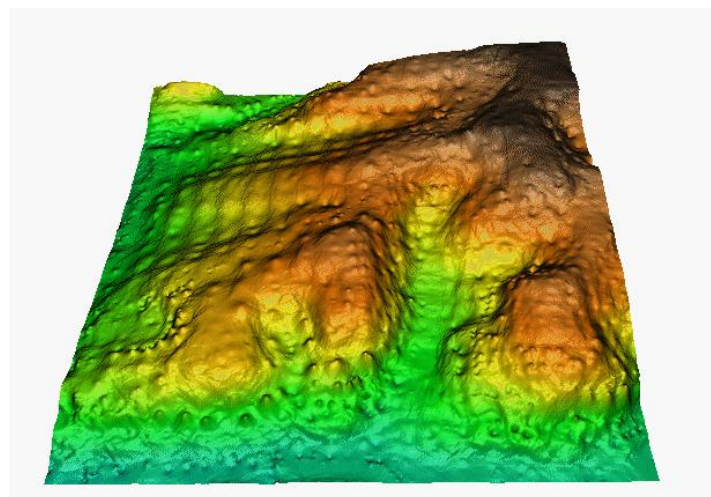
Obecný vzorec pro odhad hodnoty Z:

$$\hat{Z} = \frac{\sum_{i=1}^n w_i z_i}{\sum_{i=1}^n w_i} \quad \text{kde váhy se nejčastěji určují ze vztahu } w = \frac{1}{d^k} \text{ a nebo } w = e^{-kd}$$

hodnoty vah w_i představují funkci vzdálenosti d . Hodnota exponentu k se nejčastěji volí 1 či 2 a ovlivňuje, v jakém poměru klesá hodnota váhy měřeného bodu s rostoucí vzdáleností od bodu interpolovaného.

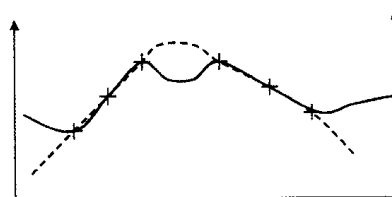


Obr. 2.18 Odhad hodnoty v bodě metodou inverzní vzdálenosti



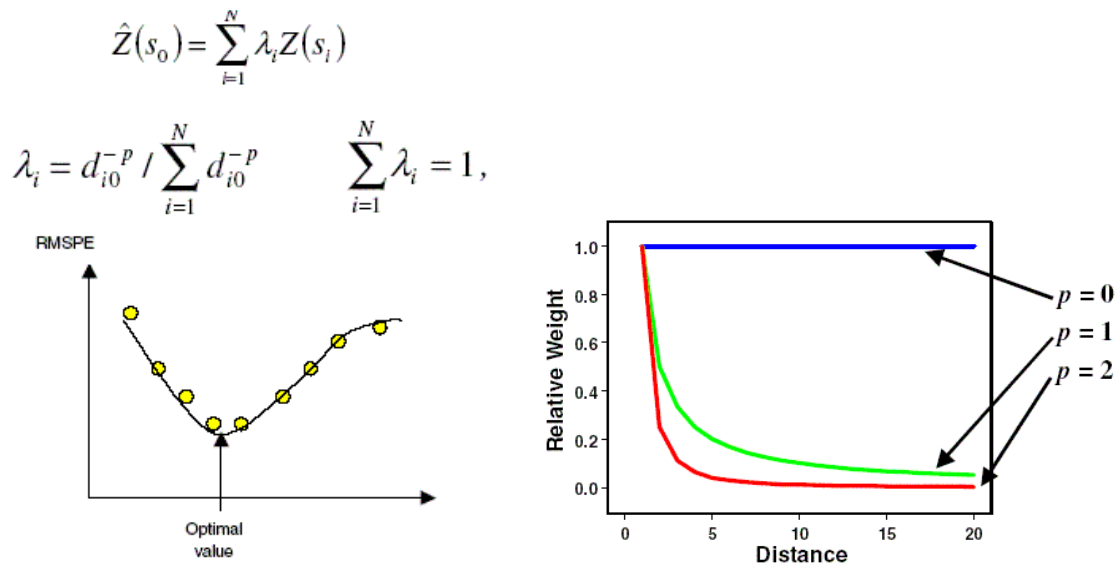
Obr. 2.19 Příklad interpolace metodou inverzní vzdálenosti

Metoda IDW často produkuje povrch, který je charakteristický koncentrickými strukturami kolem interpolovaných bodů (tzv. „bulls eyes“). Protože IDW je založena na lokálním průměrování, neposkytuje odhady mimo rozsah hodnot měřených bodů. Výsledkem jsou často nereálné tvary výsledného povrchu (viz. následující obr).



Obr. 2.20 Metoda inverzní vzdálenosti efekt „průměrování“- potlačení lokálních extrémů

Modifikace metody inverzní vzdálenosti implementovaná např. v ArcGIS je založena na následujícím modelu:



Obr. 2.21 Modifikace metody inverzní vzdálenosti- způsob odhadu optimální hodnoty exponentu vah p výpočtem RMSPE. Závislost mezi hodnotou váhy (λ) a vzdáleností bodu měřeného bodu od bodu interpolovaného pro různé hodnoty exponentu p

Váhy (λ) jsou v tomto případě definovány podle výše uvedeného vzorce, ve kterém exponent p vyjadřuje jejich změnu v závislosti na vzdálenosti interpolovaného bodu od bodu měřeného. Tuto závislost ukazují obr. vlevo. Metoda dále umožňuje prostřednictvím minimalizace tzv. RMSPE – root mean square prediction error – nalézt optimální hodnotu p .

Způsob definování velikosti okolí – ve většině případů se uvažuje kruhové okolí interpolovaného bodu a pro odhad hodnoty se berou všechny body bez ohledu na směr, ve kterém se nachází (povrch se považuje za izotropní). Pokud však existuje reálný předpoklad, že body v jistém směru mohou mít na interpolovanou hodnotu jinou váhu než ve směru jiném, potom může mít okolí tvar elipsy. Je-li například takovýmto vlivem převládající směr větru, potom okolí interpolovaného bodu je definováno jako elipsa, jejíž hlavní osa je rovnoběžná s tímto směrem. Předpokládáme, že v tomto směru si budou hodnoty interpolované veličiny více podobné na větší vzdálenost než ve směru kolmém.

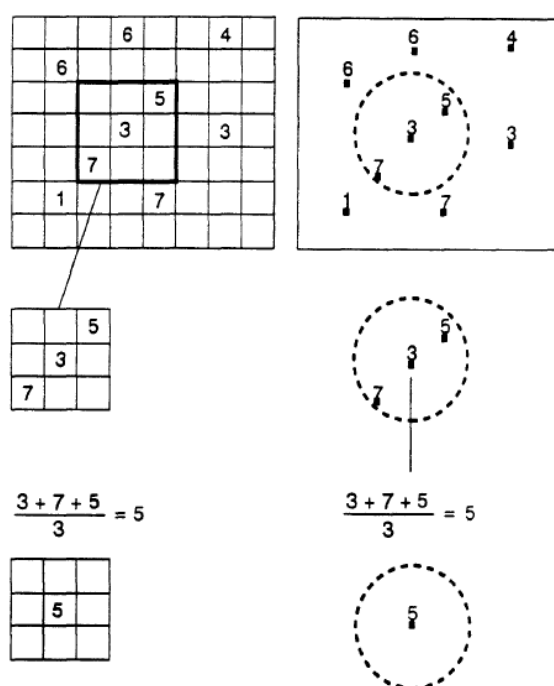
Dále je řešena otázka počtu bodů (minimální a maximální počet bodů uvažovaných pro výpočet nové hodnoty) a také jejich rozmístění v rámci definovaného okolí. To bývá děleno na kvadranty či oktanty a takovém případě je min. a max. počet vztažen k těmto sektorům.

Metoda IDW je senzitivní na shluky měřených bodů a také na odlehle hodnoty. Jistou nevýhodou také je, že minimální a maximální hodnota interpolované veličiny se může nacházet pouze v bodech měření (viz. dále - porovnání s metodami RBF).

Jistou modifikací výše popsané metody je tzv. **Shepardova metoda**. Ta navíc provádí vyrovnání interpolovaných hodnot metodou nemenších čtverců. Výsledkem je potlačení efektu koncentrických izolinií.

2.4.4 Prostorové klouzavé průměry

Za modifikaci metody inverzní vzdálenosti lze považovat metodu prostorových klouzavých průměrů. Nová hodnota může být prostým průměrem či váženým průměrem ale též např. modální hodnotou. Stěžejní úlohou této metody je definování velikosti, tvaru a charakteru okolí. Okolí je nejčastěji navrhováno ve tvaru kruhu či pravoúhelníka. Jako váhy se nejčastěji využívá vzdálenosti od středu definovaného okolí a váhy se mohou měnit lineárně i nelineárně. Vzhledem k často omezenému počtu bodů měření je vedle velikosti okolí důležitá otázka také počtu bodů v okolí (minimálního i maximálního). Borrough (1986) navrhuje použít 4 až 12 bodů s optimem 6 až 8 bodů. Větší počet bodů produkuje značně shlazený povrch, u malého počtu bodů dominují extrémní hodnoty.



Obr. 2.22 Příklad interpolace metodou prostorových klouzavých průměrů

Metody je vhodné použít za těchto podmínek:

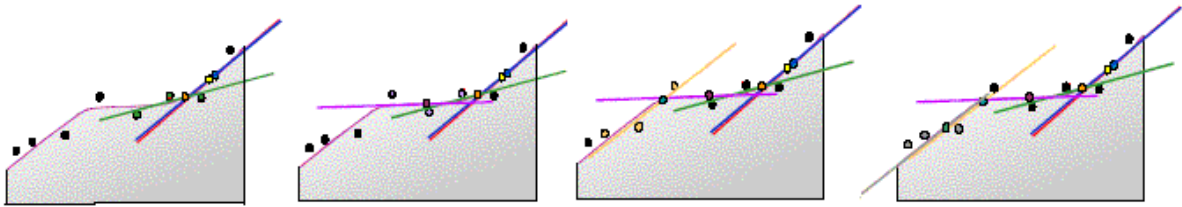
- existuje nejistota s ohledem na reprodukovatelnost výsledků opakovaných měření v daném bodě (vlastní proměnlivost pole hodnot měření)
- samotná technická stránka měření je zatížena jistou chybou
- je známo, že skutečné prostorové pole daného jevu vykazuje kromě obecného trendu také lokální variabilitu.

Příkladem může být měření rychlosti větru.

2.4.5 Interpolace metodou lokálních polynomů

Polynom n-tého stupně je aplikován ne na celý interpolovaný povrch, ale vždy na část povrchu definovanou jako okolí interpolovaného bodu přičemž tato okolí se překrývají. Stejně jako v případě IDW je specifikován tvar okolí, min. a max. počet bodů v okolí resp. rozdělení

okolí na sektory. Body definovaného okolí je proložen polynom n-tého stupně a interpolovaná hodnota je použita pro střední bod okolí. V následném kroku se okolí posouvá po interpolované ploše stejně jako v případě klouzavých průměrů. Jedná se o aproximativní metodu interpolace, která však více zohledňuje lokální vlivy než metoda „globálních“ polynomů. Obrázek ukazuje v transektu čtyři kroky postupného prokládání přímky třemi nejbližšími body.



Obr. 2.23 Interpolace metodou lokálních polynomů

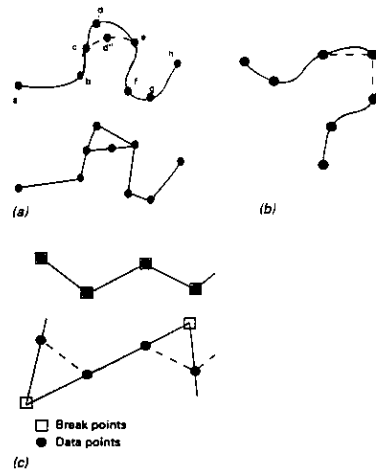
Model lokálních polynomů je optimalizován výpočtem RMSPE a může počítat s efektem anizotropie stejně jako v případě metody inverzní vzdálenosti. Metoda je závislá na správné volbě velikosti okolí interpolovaného bodu.

2.4.6 Lokální interpolátory využívající regresní analýzy

Spočívají v sestavení empirického modelu závislosti interpolované veličiny na hodnotách jedné či několika veličin nezávislých a to pro jisté okolí interpolovaného bodu. Regresní vztah je tedy na rozdíl od globální varianty této metody sestaven pouze pro body v předem definovaném okolí bodu. Interpolovaná hodnota je použita pro střední bod okolí, které se posouvá stejně jako v případě klouzavých průměrů.

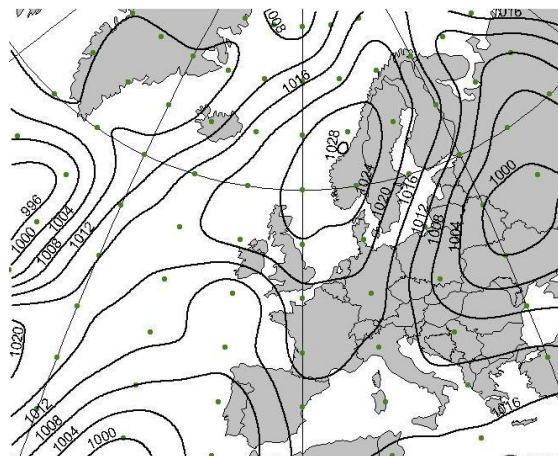
2.4.7 Splinové funkce (*piece wise polynomial function*)

Splinové funkce jsou matematicky definované křivky, které po částech interpolují jednotlivé body povrchu a to exaktně, přitom navíc zajišťují kontinuální spojení jednotlivých částí interpolovaného povrchu. Se spliny lze modifikovat část povrchu aniž bychom museli přepočítávat celý povrch (toto například neumožňují trendy). Pro interpolování linií se používá tzv. **kubických splinů**, pro interpolování povrchů se využívá jejich 2D analogie označované jako „**thin plate splines**“



Obr. 2.24 Interpolace splinovými funkcemi

Kubické splíny používané ke shlazování čar dávají v případě interpolovaných povrchů značný počet chyb (výrazně malých či velkých hodnot), ať již v důsledku chyb měření či v důsledku komplexnosti interpolovaného povrchu. V tomto případě se na místo přesných splínů používá tzv. „**thin plate splines**“. Ty nahrazují části povrchů interpolované přesným splínem lokálně shlazenou průměrnou hodnotou. Povrch je interpolován tak, aby procházel co nejbližší měřeným bodům a také aby zachoval **podmínku minimální křivosti**. Splíny jsou tedy lokálním interpolátorem - používají v daném čase pouze několika málo bodů, na rozdíl od trendových funkcí a povrchů interpolovaných metodou vážené inverzní vzdálenosti splíny zachovávají řadu lokálních rysů interpolované proměnné. Splíny interpolované povrchy jsou často značně shlazené a jsou tedy vhodné pro interpolaci jevů, které se mění spojitě (např. tak vzduchu). Jistou nevýhodou splinových funkcí je, že produkují „falešná“ lokální minima a maxima.



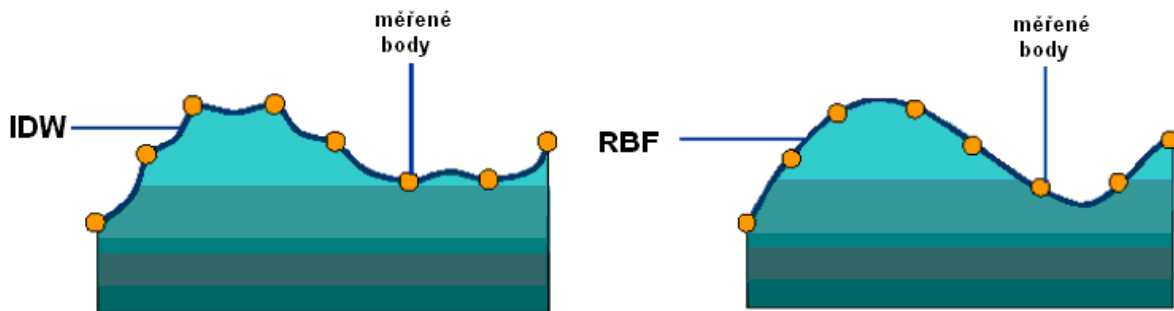
Obr. 2.25 Příklad izolinií vytvořených interpolací gridových hodnot přízemního pole tlaku vzduchu splinovými funkcemi

Následující stránka prezentuje názorně způsob interpolace metodou splínů:

<http://www.math.ucla.edu/~baker/java/hoefer/Spline.htm>

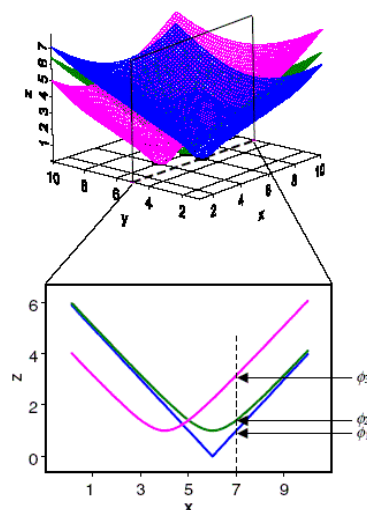
Metody radiálních funkcí (RBF)

V prostředí ArcGIS jsou tyto metody interpolace označovány jako „**radial basis functions**“ (RBF). Jedná se o skupinu pěti exaktních interpolátorů označovaných: thin plate splines, spliny s tenzí, regularizované spliny, multikvadratické spliny, inverzní multikvadratické spliny. Tyto postupy k interpolaci využívají m.j. umělých neuronových sítí za podmínky minimalizování křivosti povrchu (analogie „přetažení“ gumové membrány přes body v prostoru). Obr. XX uvádí porovnání RBF metod s metodou inverzní vzdálenosti. Jak je z obrázku patrné, výsledkem interpolace metodou inverzní vzdálenosti nikdy nejsou body, které by byly větší než maximální hodnota v měřeném bodě resp. menší než minimální hodnota v měřeném bodě.



Obr. 2.26 Porovnání výsledků interpolace metodami splinových funkcí (RBW) a metodou inverzní vzdálenosti (IDW).

Parametry konkrétní interpolující funkce jsou optimalizovány výpočtem RMSPE. RBF jsou exaktní metodou a jsou vhodné pro hladké povrchy generované z velkého počtu bodů (např. modely terénu). Naopak se nehodí pro interpolaci jevů, které se v prostoru mění skokem (abrupt) a dále pro interpolaci jevů, u nichž existuje jistá míra nejistoty ohledně přesnosti měřených bodů.



Obr. 2.27 Princip interpolace metodou multiquadratic RBF (vysvětlivky viz text)

RBF jsou funkce které se mění se vzdáleností od interpolovaného bodu. Jsou konstruovány pro každý měřený bod. Na obrázku jsou vykresleny různou barvou tři RBF funkce pro tři body v prostoru. V tomto případě jsou RBF jednoduchou funkcí vzdálenosti od měřeného

bodů a mají tvar obráceného kužele. Budeme uvažovat řez v rovině os X a Z pro bod $y = 5$. Předpokládejme, že budeme interpolovat bod o souřadnicích $x = 7$ a $y = 5$. Hodnotu každé RBF v predikovaném bodě můžeme odečíst z grafu jako ϕ_1, ϕ_2, ϕ_3 . Prediktor má podobu váženého průměru, tedy:

$$w_1\phi_1 + w_2\phi_2 + w_3\phi_3 + \dots$$

Doposud nebylo využito hodnot v měřených bodech. Proto váhy w_1, w_2, w_3 jsou nalezeny na základě podmínky, že pokud je odhadován bod v bodě měření, je interpolován přesně. Tato podmínka umožňuje sestavit soustavu N rovnic o N neznámých, která má jednoznačné řešení. Všechny metody interpolace využívající RBF dávají velmi podobné výsledky. Metody mají možnost nastavit parametr, který ovlivňuje shlazení výsledného povrchu. U všech metod RBF platí, že čím větší hodnota vyhlazovacího parametru, tím více shlazený je povrch. Opačně je tomu pouze pro tzv. inverzní multiquadric RBF.

Nejčastěji využívanou je multikvadrková metoda (multiquadric RBF), která vychází z řešení následující rovnice:

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

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

$d_i(x, y)$ – relativní vzdálenost měřeného bodu v místě x_i, y_i 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.

2.4.8 Kriging

Je to lokální interpolátor, který optimalizuje výběr bodů okolí, ze kterých je odhadována nová hodnota. K této optimalizaci se provádí tzv. **strukturní analýza** založená na studiu tzv. semivariogramu a konstrukci teoretického modelu. Parametry tohoto modelu jsou použity ve vlastním krigování. Kriging je založen na odhadu závislosti průměrné změny v hodnotách studované veličiny a vzdálenosti měřených bodů. Strukturní analýze a metodě korigování je věnována zvláštní kapitola.

2.4.9 Metody prostorové interpolace ploch (area based)

Mnoho jevů se vztahuje k plošným jednotkám spíše než k bodům (hustota obyvatelstva států, kvalita pitné vody...). Metody řeší způsob, jakým lze odhadnout hodnoty jistého jevu na základě hodnot jiného jevu vázaných na plošné jednotky. Přitom mohou nastat dvě situace:

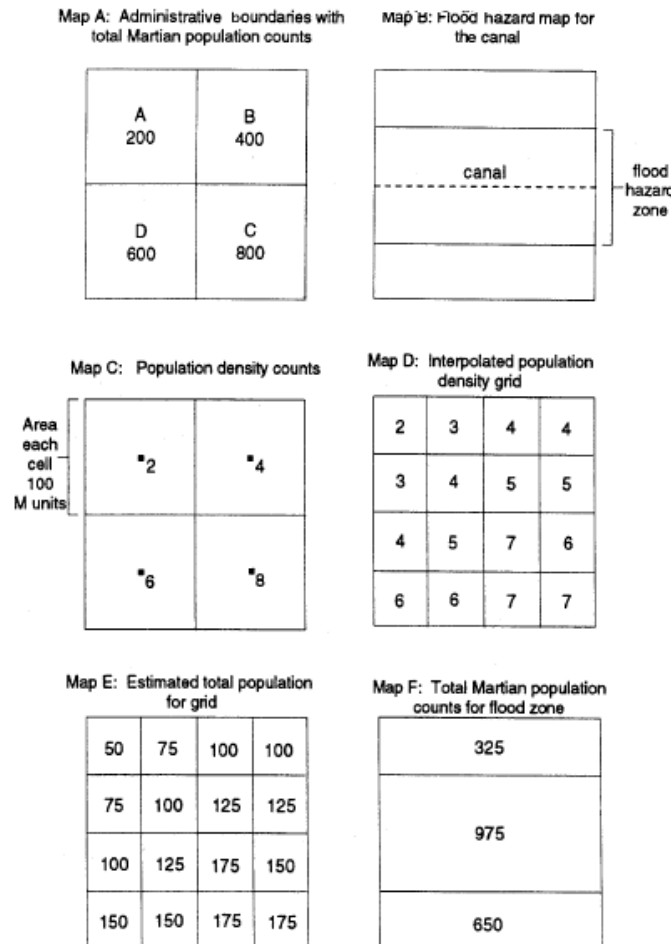
1. plošné jednotky se shodují
2. zdrojové jednotky jsou podmnožinou (nested) jednotek výstupních

Metody lze dělit do dvou skupin:

1. metody zachovávající objem studovaného jevu (volume preserving)
2. metody nezachovávající objem studovaného jevu (non-volume preserving)

Metody nezachovávající objem studovaného jevu (non-volume preserving)

Příklad – mapa A vyjadřuje celkový počet obyvatel ve čtyřech administrativních jednotkách určitého území. Mapa B potom záplavovou oblast kolem vodního toku. Cílem je zjistit, jaká je hustota obyvatelstva uvnitř záplavové zóny.



Obr. 2.28 Princip metody prostorové interpolace ploch nezachovávající objem studovaného jevu

Postup:

- 1) výpočet hustoty obyvatelstva pro každou plochu
- 2) určení centroidu každé plochy
- 3) interpolace hustoty obyvatelstva výše popsanými metodami

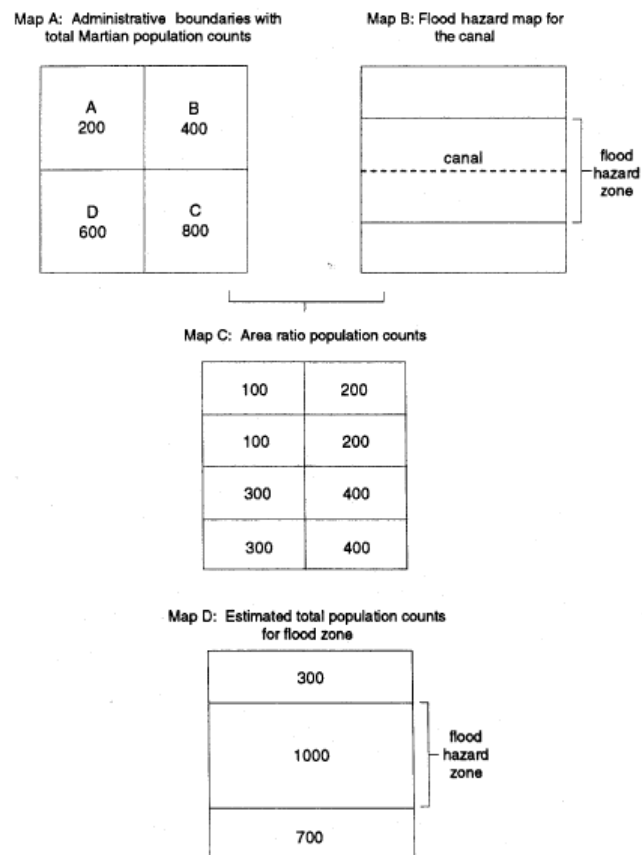
Metody zachovávající objem studovaného jevu (volume preserving)

Provede se překrytí cílových zón (oblastí) přes oblasti zdrojové a určí se poměrná část cílové zóny, která spadá do zóny zdrojové. Celková hodnota atributu v cílové zóně je určena v závislosti na plošném zastoupení zón zdrojových.

2.4.10 Pycnophylatic method

Jeden z hlavních problémů metody Thiessenových polygonů je, že měřený prvek se považuje za homogenní v rámci jedné třídy, veškeré prostorové změny jsou vázány na hranice. V případě spojitých či relativně spojitých prvků jde o naprosto nevhodný způsob interpolace.

Modifikací je metoda, která v rámci každé třídy zachovává sumu studovaného prvku, avšak dovoluje kontinuální změnu směrem k hranicím každé třídy. Metoda bere v úvahu hodnotu atributu sousedních tříd. Nepředpokládá se existence bariér a hodnoty sousedních tříd jsou shlazeny bez skokových změn v hodnotách daného atributu pomocí pravděpodobnostní funkce (density function). Metoda byla použita na demografická data. Jde o neexaktní interpolátor. Minimální i maximální hodnoty vypočtené touto metodou jsou výrazně menší resp. větší než skutečně naměřené hodnoty.



Obr. 2.29 Princip metody prostorové interpolace ploch zachovávající objem studovaného jevu

Radial basis interpolation – skupina metod podobných geostatistickým metodám krigování, nelze ale využít metody modelování variogramu, nejsou vyžadovány vstupní podmínky jako u korigování.

Splines

SPLINE

Cubic – polynom 3 řádu

Bicubic – 2 souřadnice

$$f(x,y)=a_1+a_2x+a_3y+a_4x^2+a_5y^2+a_6xy+a_7x^2y+a_8xy^2+a_9x^3+a_{10}y^3$$

Remember that linear interpolation uses a linear function for each of intervals $[x_k, x_{k+1}]$. Spline interpolation uses low-degree polynomials in each of the intervals, and chooses the polynomial pieces such that they fit smoothly together. The resulting function is called a [spline](#).

For instance, the [natural cubic spline](#) is [piecewise](#) cubic and twice continuously differentiable. Furthermore, its second derivative is zero at the end points. The natural cubic spline interpolating the points in the table above is given by

3. Geostatistické metody interpolace

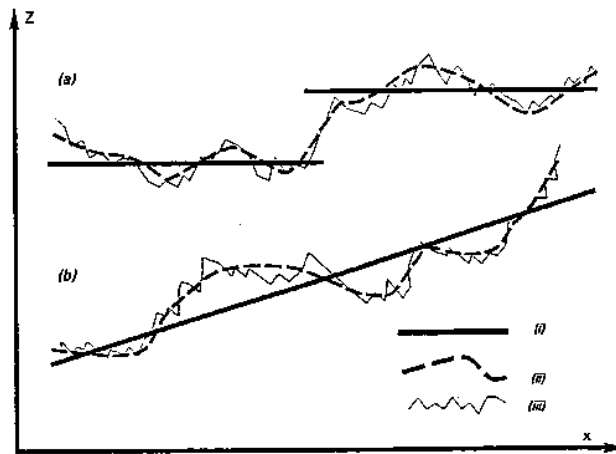
Při použití většiny dosud uvedených interpolačních algoritmů nemáme *a priori* objektivní informaci o tom, zda způsoby definování okolí interpolovaného bodu, body použité k interpolaci a také jejich váhy jsou zvoleny vhodně. V případě kvalitních vstupních dat dává většina interpolačních technik podobné výsledky (Borrough, McDonnell, 1998). Vstupních dat musí být především dostatečný počet a musí být také vhodně rozmístěna ve studovaném území. V opačném případě hraje velkou roli volba vhodného interpolačního algoritmu.

Žádná z metod dosud neřešila následující problémy:

- počet bodů nutných k výpočtu lokálního průměru
- velikost orientaci a tvar okolí
- zda neexistuje jiná cesta k definování vah než funkce vzdálenosti bodů
- jaké jsou chyby a nejistoty spojené s interpolovanými hodnotami

Odpovědi na tyto otázky poskytují geostatistické postupy založené na tzv. **strukturní analýze**. Její výsledky jsou poté mimo jiné využitelné v interpolačních postupech **krigingu**. Metodu krigingu uvedli G. Matheron a D.G. Krige a je založena na skutečnosti, že interpolovaný povrch lze lépe vystihnout stochastickou funkcí než shlazující matematickou funkcí. Předchozí metody interpolace lze označit jako deterministické. Metoda krigingu odhaduje interpolační váhy bodů tak, že optimalizuje interpolační funkci a lze ji například využít také k optimalizaci sítí apod.

Kriging jako interpolační metoda byla vyvinuta v geologii a je vhodná pro interpolování proměnných, které se v prostoru mění s jistou kontinuitou, ale nelze je popsat jednoduchou shlazující funkcí některého z globálních interpolátorů. V případě krigingu je interpolovaný povrch tvořen ze **tří složek** (obr. 1). První složku představuje tzv. strukturální komponenta s konstantním průměrem či trendem - **obecný trend** (tzv. drift). Druhou složku povrchu představují kolísání (drobné sníženiny či vyvýšeniny), jejichž podstatu lze vyjádřit určitou matematickou funkcí jako v případě trendu, ale která vyjadřují určitou prostorovou korelaci (tzv. **regionalizovaná proměnná**) Třetí složku potom představuje **náhodný šum**.



Obr. 3.1 Základní komponenty spojitého povrchu (i – trendová složka – drift; ii – tzv. regionalizovaná proměnná; iii – náhodná složka)

V případě krigingu jsou všechny tři složky analyzovány separovaně. První složka je odhadována za pomoci obecné trendové funkce. Druhá složka - náhodná, ale prostorově korelovaná kolísání jsou analyzována metodou tzv. variogramu.

Nechť x je poloha bodu v 1, 2 či 3 dimenzích, potom hodnota náhodné proměnné Z v bodě x bude

$$Z(x) = \mu(x) + \varepsilon'(x) + \varepsilon''$$

x - pozice v 1, 2 či 3 rozměrném prostoru

Z - interpolovaná proměnná

$Z(x)$ - hodnota proměnné v bodě x

$\mu(x)$ - deterministická složka (trend)

$\varepsilon'(x)$ - stochastická složka (regionalizovaná proměnná) - lokálně proměnné, ale prostorově závislé reziduum od $\mu(x)$

ε'' - náhodná, prostorově nezávislá složka, gaussovský šum s nulovým průměrem a s rozptylem σ^2 . Velké písmeno Z značí, že se jedná o náhodnou funkci a ne o měřenou hodnotu proměnné z .

Prvním krokem je zvolení vhodné funkce vystihující složku $\mu(x)$. V nejjednodušším případě, když se v hodnotách nenachází žádný trend (drift), potom se $\mu(x)$ rovná průměrné hodnotě v ploše a průměr či očekávaný rozdíl mezi dvěma místy x a $x+h$ vzdálenými od sebe o hodnotu vektoru h , bude nula:

$$E[Z(x) - Z(x+h)] = 0$$

kde $Z(x)$ a $Z(x+h)$ jsou hodnoty náhodné proměnné Z v poloze x , $x+h$.

Dále také předpokládáme, že rozptyl rozdílů závisí pouze na vzdálenosti mezi místy, tedy:

$$E\left\{[Z(x) - Z(x+h)]^2\right\} = E\left\{[\varepsilon'(x) - \varepsilon'(x+h)]^2\right\} = 2\gamma(h)$$

kde hodnota $\gamma(h)$ se označuje jako semivariance.

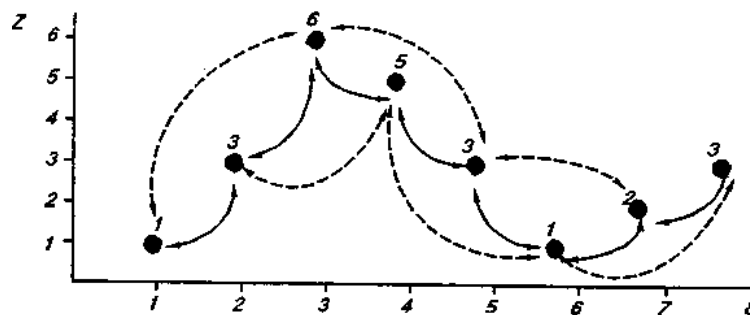
To značí, že jakmile jsme odhadli příspěvek strukturní proměnné $\mu(x)$, zbývající kolísání má konstantní rozptyl a difference mezi dvěma místy jsou pouze funkcí jejich vzdálenosti. Výše uvedený vztah lze přepsat:

$$Z(x) = \mu(x) + \gamma(h) + \varepsilon$$

tedy mezi $\varepsilon'(x)$ a $\gamma(h)$ je ekvivalence

3.1 Strukturní analýza

Geostatistická strukturní analýza (variografie) je procedura zahrnující výpočet strukturálních funkcí, výběr a konstrukci odpovídajících teoretických modelů a jejich aplikace, interpretaci průběhu strukturálních funkcí. Cílem je popsat takové vlastnosti jako jsou **kontinuita**, **homogenita**, **stacionarita** či **anizotropie** pole studovaných prostorových proměnných veličin. Tyto vlastnosti jsou popisovány prostřednictvím měr prostorové autokorelace a prostorové variability. Strukturní analýza je výchozím krokem geostatistického modelování. Sama o sobě ale poskytuje řadu velmi důležitých informací o struktuře náhodného pole jako modelu konkrétního objektu v krajinné sféře.



Obr. 3.2 Příklad výpočtu měr prostorové variability pro 1D (řadu hodnot)

Ke kvantifikaci prostorové autokorelace, která vyjadřuje skutečnost, že objekty blízké si jsou více podobné než objekty vzdálenější slouží **strukturní funkce**. Strukturní funkce měří sílu korelačního vztahu jako funkci vzdálenosti. Na obr. 2 je pro jednoduchost odvozena strukturní funkce pro řadu pravidelně rozmístěných bodů na linii (tedy pro 1D).

Charakteristiky, které popisují prostorovou variabilitu lze odvodit z měr úrovně a variability následovně:

$$\text{průměr} = (1+3+6+5+3+1+2+3)/8=3,0$$

$$\text{rozptyl} = [(1-3)^2 + (3-3)^2 + (6-3)^2 + (5-3)^2 + (3-3)^2 + (1-3)^2 + (2-3)^2 + (2-3)^2]/8=2,75$$

$$\text{kovariance}(1) = [(1-3)*(3-3) + (3-3)*(6-3) + (6-3)*(5-3) + (5-3)*(3-3) + (3-3)*(1-3) + (1-3)*(2-3) + (2-3)*(3-3)]/7=1,14$$

$$\text{semivariance}(1) = [(1-3)^2 + (3-6)^2 + (6-5)^2 + (5-3)^2 + (3-1)^2 + (1-2)^2 + (2-3)^2]/7=3,43$$

$$\text{semivariance}(2) = [(1-6)^2 + (3-5)^2 + (6-3)^2 + (5-1)^2 + (3-2)^2 + (1-3)^2]/6=9,83$$

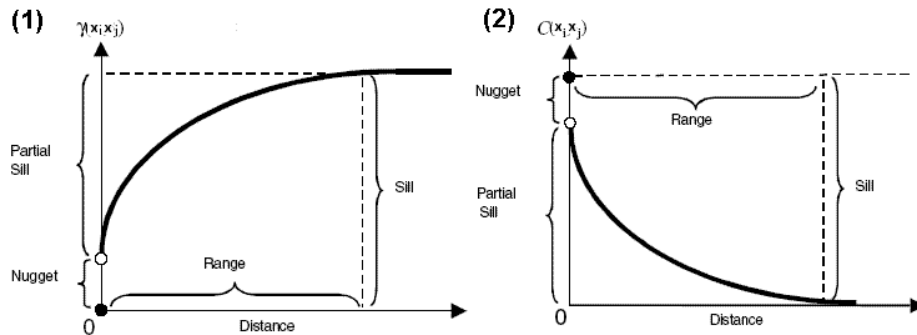
$$\text{semivariance}(3) = [(1-5)^2 + (3-3)^2 + (6-1)^2 + (5-2)^2 + (3-3)^2]/5=12,50$$

...

Semivariance může být definována jako.

$$\gamma(x_i, x_j) = 1/2 \text{var}(Z(x_i) - Z(x_j))$$

kde *var* značí rozptyl. Jestliže jsou dva body x_i a x_j blízko sebe, bude rozdíl hodnot studované veličiny $Z(x_i)$ a $Z(x_j)$ těchto bodech malý. S růstem vzdálenosti si budou hodnoty méně podobné. Grafickým vyjádřením závislosti semivariance na vzdálenosti je strukturální funkce nazývaná semivariogram, jejíž typický průběh je na obr. 3.



Obr. 3.3 Vztah mezi semivariogramem (1) a kovarianční funkcí (2). Vysvětlení jednotlivých termínů viz. dále Semivariogram je pouze jednou z mnoha strukturálních funkcí, i když nejpoužívanější. Semivariogram je **mírou nepodobnosti**. K dalším takovýmto funkcím patří kovarianční funkce:

$$C(x_i, x_j) = \text{cov}(Z(x_i), Z(x_j))$$

kde *cov* značí kovarianci. **Kovariance** je stejně jako korelace **mírou podobnosti**. Budou-li dva body blízko sebe, budou hodnoty v těchto bodech podobné a jejich kovariance bude velká. S růstem vzdálenosti bude klesat podobnost bodů a budou klesat i hodnoty kovariance. Hodnota kovariance je při nulové vzdálenosti rovna rozptylu množiny zpracovávaných dat.

Vztah mezi semivariogramem a kovarianční funkcí lze vyjádřit následovně:

$$\gamma(x_i, x_j) = \text{sill} - C(x_i, x_j)$$

Veličina označená *sill* značí tzv. práh a je to hodnota, na níž má semivariogram vodorovný průběh (viz. dále).

3.2 Strukturální analýza v 2D

Výše uvedený příklad určení semivariance platí pro řadu pravidelně rozmístěných bodů. Strukturální analýzy a následného krigování se však ve většině případů používá pro charakterizování vztahů prostorové autokorelace a pro následné odhady a interpolace v rovině, ve kterém máme množinu pravidelně, častěji však nepravidelně rozmístěných bodů měření.

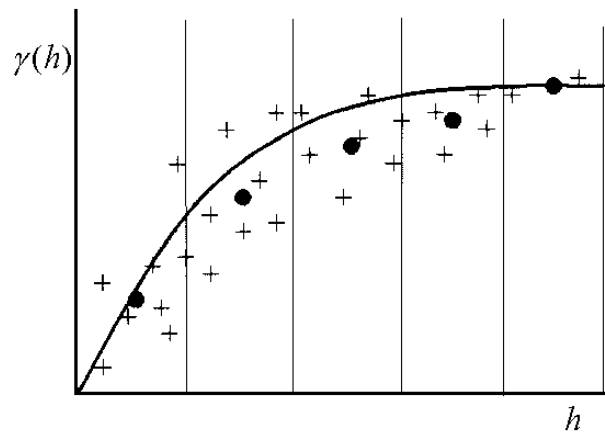
Abychom popsali závislost hodnot studovaného jevu v prostoru, vyneseme hodnoty semivariancí pro všechny dvojice bodů do semivariogramu obdobně jako ve výše uvedeném případě. **Semivariogram** je strukturální funkce, která popisuje závislost průměrné kvadratické difference hodnot prostorové proměnné veličiny Z na vzdálenosti h . Semivarianci lze odhadnout z naměřených dat podle následujícího vztahu:

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

kde:

n - počet dvojic bodů pozorování proměnné z vzdálených o hodnotu h

h - tzv. lag - vzdálenost dané dvojice bodů.

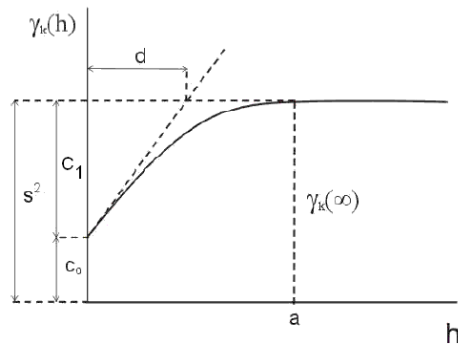


Obr. 3.4 Experimentální semivariogram (+) s charakteristickými hodnotami pro vzdálenosti h (•) a proložený teoretický model semivariogramu (plná čára)

Graf hodnot $\hat{\gamma}(h)$ a h se označuje jako **experimentální semivariogram** a je prvním krokem ke kvantitativní deskripci regionalizované proměnné. I když je semivariogram vektorová funkce, sestavuje se často jako **izotropní** (tj. bez ohledu na směr) pro celkové charakterizování hodnoceného náhodného pole nebo tehdy, je-li k dispozici omezený počet pozorování. Experimentálním semivariogramem se v následujícím kroku prokládá **teoretický model**.

3.2.1 Prvky semivariogramu

Na Obr. 3.5 je uveden často používaný tzv. sférický model s vysvětlením používané terminologie. Na horizontální ose je vynášena vzdálenost h mezi jednotlivými vstupními body interpolovaného povrchu (tzv. lag), na vertikální ose potom rozptyl zkoumané proměnné jako funkce vzájemných vzdáleností jednotlivých měřených bodů. Je mírou, která vyjadřuje, jak velké je okolí, daného bodu, ve kterém se nacházejí body sousední, jejichž hodnota interpolovaného atributu závisí (koreluje) s hodnotou v tomto bodě. Takto vnesenými body je proložena křivka mající charakteristický tvar. Je-li vzdálenost mezi dvěma body malá, jejich hodnoty jsou podobné a hodnota semivariance je také malá. Se zvětšující se vzdáleností hodnota semivariance roste. Při určité vzdálenosti dvou bodů je možno říci, že jejich hodnoty (např. výšky) spolu již nekorelují a hodnota semivariance i se zvyšováním vzdálenosti již neroste, ale zůstává konstantní. Plochá část semivariogramu určuje tzv. **práh (sill)** a je rovna rozptylu zpracovávaných dat. Existence prahu značí, že se zvětšující se vzdáleností se hodnoty semivariance nemění. Kritická hodnota vzdálenosti, na níž se křivka semivariogramu stává rovnoběžnou s vodorovnou osou se označuje jako **dosah (range)**. Dosah definuje pro daný bod velikost okolí, které je nutné uvažovat při interpolaci hodnoty v daném bodě.



Obr. 3.5 Příklad teoretického semivariogramu – sférický model. Parametry semivariogramu: a - dosah (range), d – rozpětí, c_0 - zbytkový rozptyl (nugget), $c=c_0 + c_1$ - práh (sill), h – lag (krok vzdálenosti)

Z výše uvedeného by také mělo platit, že proložená křivka semivariogramu by měla procházet počátkem souřadné soustavy (nulová vzdálenost mezi body znamená zákonitě také nulovou hodnotu rozptylu). Velmi často nenabývají experimentální semivariogramy v počátku nulové hodnoty; protínají osu y v nenulové hodnotě, která je nazývána **zbytkový rozptyl (nugget variance)**. To může ukazovat na rozptyl menší než je "vzorkovací" vzdálenost, nebo na malou přesnost měření, kdy např. jsou v datech obsaženy dva vzorky ze stejného místa, pokaždé s jinou hodnotou. Zbytkový rozptyl je tak vyjádřením náhodného šumu ε' a sestává se jednak z chyb měření (dvě různé hodnoty pro jeden bod) a jednak z tzv. microscale variation – ty jsou vyjádřením rozptylu hodnot složky ε' . V případě nulového zbytkového rozptylu a tedy v případě nulové chyby měření je krigování exaktním interpolátorem.

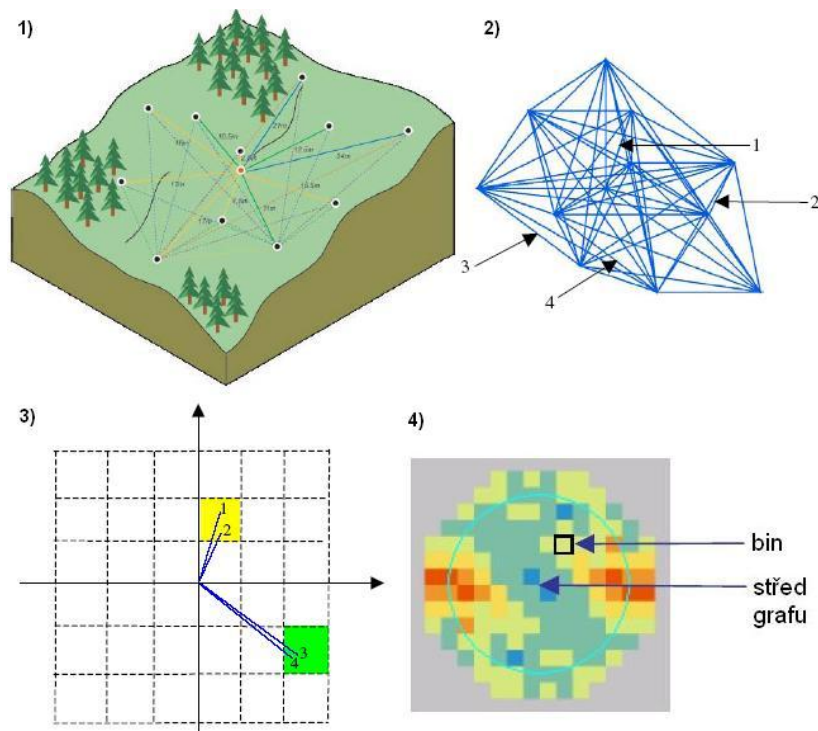
Na tvar semivariogramu má značný vliv velikost hodnoty lag (vzdálenosti h). Velikost h se volí např. jako průměrná minimální vzdálenost mezi sousedními body. Velká hodnota h dává hladší průběh semivariogramu, může však zamaskovat efekt autokorelace studovaných dat na menší vzdálenosti. Naopak malá hodnota h má vliv na malý a tedy často nereprezentativní počet bodů v rámci každé hodnoty násobku h. Nevhodná délka kroku se může dále projevit v oscilaci hodnot semivariogramu. Výpočet semivariací se většinou provádí do vzdálenosti rovné polovině maximální vzdálenosti bodů v prostoru. Tedy násobíme-li velikost kroku počtem kroků, měli bychom dostat hodnotu rovnou zhruba polovině maximální vzdálenosti mezi interpolovanými body.

3.2.2 Efekt anizotropie

V případě velkého množství nepravidelně rozmístěných bodů je vhodné hodnotu semivariance vyjádřit pro skupiny bodů přibližně stejně vzdálených. V tomto případě se do výpočtu hodnoty semivariance pro dané h berou všechny body padnoucí do mezikruží určeného **tolerancí délky kroku**. Toleranci délky kroku je nutné volit v případě nerovnoměrného rozmístění měřených bodů v rovině. Tato hodnota se volí v mezích od 10 - 50% z délky kroku h. Grupování hodnot semivariací na základě podobné vzdálenosti (tzv. **binning**) dovoluje konstruovat druhý typ grafu, často využívaného pro studium prostorové autokorelace a pro snadnější interpretaci hodnot semivariogramu – tzv. **plošný graf semivariance**. Ten navíc umožňuje posoudit eventuelní rozdíly v hodnotách semivariance

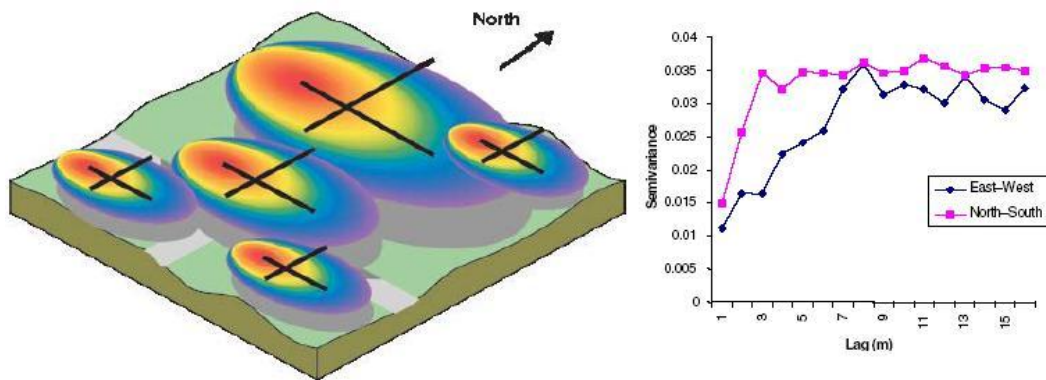
v závislosti na směru – tedy definovat efekt **anizotropie**. Proto se tento typ grafu také označuje jako povrch anizotropie.

Plošný graf semivariance představuje grid s buňkami o velikosti strany rovné vzdálenosti h (lag). V grafu určujeme vzdálenost směrem od středu. Jednotlivé buňky grafu nesou hodnotu semivariance vypočtenou ze všech dvojic bodů, které jsou od sebe vzdáleny o právě o vzdálenost od středu grafu a které se navíc nacházejí v určitém směru - viz. obr. 2) a 3). Hodnoty semivariací jsou potom vyjádřeny např. barvou.



Obr. 3.6 Princip grupování hodnot semivariací na základě podobné vzdálenosti a plošný graf semivariance.

Hodnoty semivariací obecně rostou směrem od středu grafu, protože podobnost hodnot studované veličiny s růstem vzdálenosti obecně klesá – tedy roste jejich nepodobnost vyjádřená semivariací. V případě, že se hodnota semivariance mění s rostoucí vzdáleností (směrem od středu grafu) stejně ve všech směrech, potom hovoříme izotropii studovaného pole. V opačném případě tvoří hodnoty semivariance na plošném grafu tvar elipsy. Výsledkem je, že i tvar semivariogramu bude jiný ve směru hlavní a vedlejší poloosy. Semivariogram sestavený z bodů ve směru kratší poloosy elipsy anizotropie se bude vyznačovat strmějším průběhem. Tento směr odpovídá směru maximální variability minimálního dosahu (viz. dále). Směr hlavní osy elipsy je naopak směrem minimální variability.

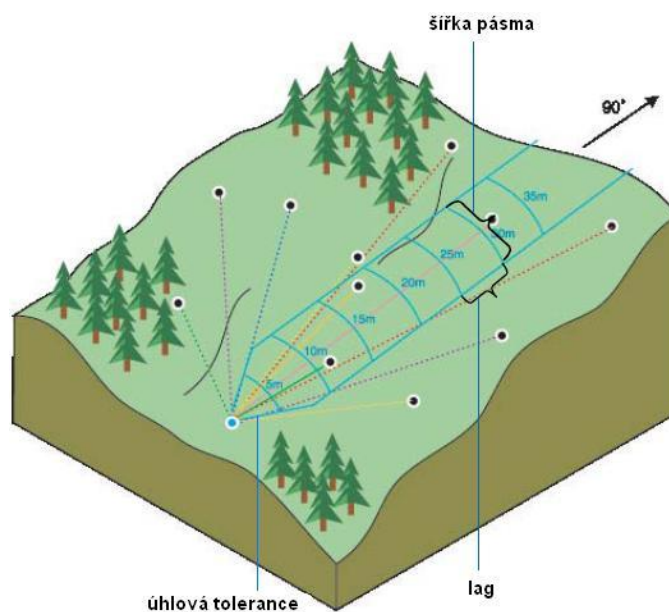


Obr. 3.7 Povrch vykazující efekt anizotropie a odpovídající empirické semivariogramy

Tzv. izotropní semivariogram tedy neuvažuje odchylky v závislosti na směru, naproti tomu anizotropní semivariogram se liší především odlišnou hodnotou dosahu pro specifické směry, další charakteristiky semivariogramu (typ, práh, zbytkový rozptyl) se většinou nemění. Takovouto anizotropii označujeme jako **geometrickou**. V případě, že nelze použít stejný model semivariogramu resp. stejné hodnoty prahu a zbytkového rozptylu hovoříme o tzv. **zonální** anizotropii. K modelování zonální anizotropie lze využít konstrukce tzv. složených modelů semivariogramů.

Ke konstrukci směrových semivariogramů je nutné řešit otázku vhodného výběru bodů. Ve většině případů se volí 4 až 8 směrů a k jejich vymezení je nutné stanovit následující parametry (jejich význam je zřejmý z následujícího obrázku):

- úhlovou toleranci
- šířku pásma
- délkovou toleranci (lag)



Obr. 3.8 Parametry tzv. směrových semivariogramů

Efekt anizotropie je vyjádřením náhodného procesu chování studované veličiny. Nelze ho zaměňovat s trendovou složkou. Ta by měla být ve zpracovávaných datech předem definována a před vlastní strukturní analýzou odstraněna (viz. ESDA).

Pokud k tomu není pádný důvod, daný fyzikální podstatou zpracovávaných dat, není vhodné používat anizotropního modelu s poměrem os elipsy anizotropie větším než 3 ku 1. Pokud experimentální semivariogram ukazuje na takovou anizotropii, je to způsobeno zřejmě trendem obsaženým v datech. Potom je vhodné nejprve z dat trend odstranit.

Teoretický semivariogram

Jedná se o model, který nejlépe aproximuje průběh experimentálního semivariogramu v okolí počátku a prahu (viz. dále). Právě proces hledání teoretického semivariogramu se někdy označuje jako strukturní analýza. Modely semivariogramů se dělí podle chování v okolí počátku a v „nekonečnu“ do několika skupin:

- modely přechodového typu - tj. s prahem (sférický, kvadratický, gaussovský, exponenciální),
- modely bez přechodu (lineární, logaritmický),
- modely s oscilujícím prahem (sinový, cosinový),
- čistě náhodný model.

U prvních tří skupin se může objevit tzv. efekt zbytkového rozptylu (nugetový efekt), který se odráží v posunu grafu semivariogramu o hodnotu c_0 ve směru osy $\hat{\gamma}(h)$.

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" semivariogramu - tzv. prahu; v rovnicích semivariogramů je dán konstantou C
- vzdálenost, ve které semivariance dosáhne prahové hodnoty - dosah (range); v rovnicích semivariogramů je dán konstantou a
- chováním v počátku (tj. semivariance mezi velmi blízkými body)
- dosah je mírou korelace uvnitř množiny dat; "dlouhý" dosah indikuje vysokou korelaci, "krátký" dosah korelaci nízkou.
- hodnota prahu je rovna celkovému rozptylu.

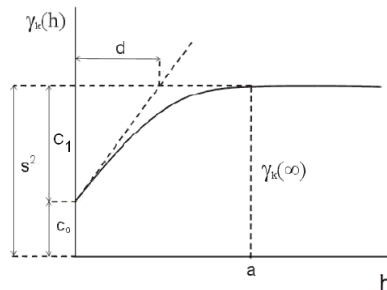
3.3 Přehled teoretických semivariogram

3.3.1 Modely přechodového typu

Modely přechodového typu (transitivní) – prostorová autokorelace kolísá s hodnotu h . U těchto klasických modelů je vyjádřena skutečnost, že při malých vzdálenostech je shoda mezi zjištěnými hodnotami vysoká (a tedy variabilita nízká), s rostoucí vzdáleností se „neshoda“ zvyšuje až do určité vzdálenosti (dosah), kde se úroveň neshody stabilizuje kolem hodnoty

statistického rozptylu. Za touto vzdáleností se již neuplatňuje prostorová vazba mezi zkoumanými místy a variabilita je plně určována statistickým rozptylem.

Sférický model – zbytkový rozptyl je důležitý, ale malý. Je jasně vyjádřen dosah (range) a prahová hodnota (sill). Je typický pro pole, ve kterém dominuje jeden zdroj variability.

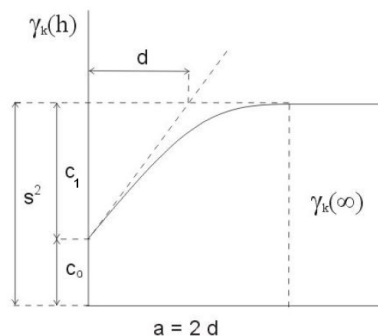


Obr. 3.9 Sférický model semivariogramu

$$\gamma(h) = c_0 + c_1 * \left[\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \dots\dots\dots \text{pro } h \leq a$$

$$\gamma(h) = c_0 + c_1 \dots\dots\dots \text{pro } h > a$$

Kvadratický model

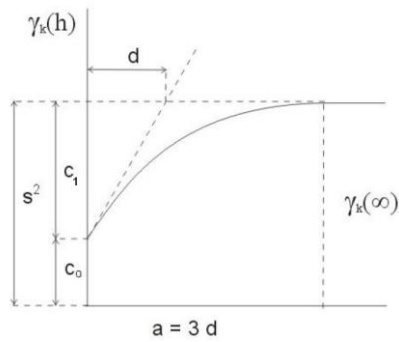


Obr. 3.10 Kvadratický model semivariogramu

$$\gamma(h) = c_0 + c_1 * \left[2 \frac{h}{a} - \left(\frac{h}{a} \right)^2 \right] \dots\dots\dots \text{pro } h \leq a$$

$$\gamma(h) = c_0 + c_1 \dots\dots\dots \text{pro } h > a$$

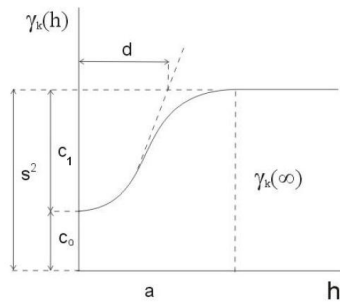
Exponenciální model – dobře vyjádřené hodnoty zbytkového rozptylu a prahu, ale pouze postupná aproximace k hodnotě dosahu (range)



Obr. 3.11 Exponenciální model semivariogramu

$$\gamma(h) = c_0 + c_1 * [1 - \exp(-h/d)], \text{ kde } a = 3d$$

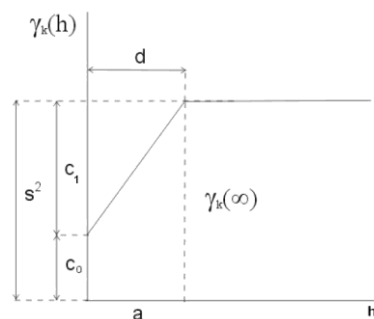
Gaussův model – hladký povrch, hodnota zbytkového rozptylu je velmi malá ve srovnání s regionalizovanou proměnnou. Model má inflexní bod. Je typický plynulými změnami hodnot. Používá se často např. při modelování výškových dat. Je používán u dobře prozkoumaných polí. Často se však vyznačuje nestabilitou.



Obr. 3.12 Gaussův model semivariogramu

$$\gamma(h) = c_0 + c * [1 - \exp(-h^2 / d^2)], \text{ kde } a = d\sqrt{3}$$

Lineární model s prahem - jednoduchý a poměrně často využívaný zvláště programy provádějícími interpolaci pomocí krigování na základě automaticky vypočítaného a vyhodnoceného semivariogramu. Při provádění strukturální analýzy se využívá raději jiných přechodových modelů.



Obr. 3.13 Lineární model semivariogramus prahem

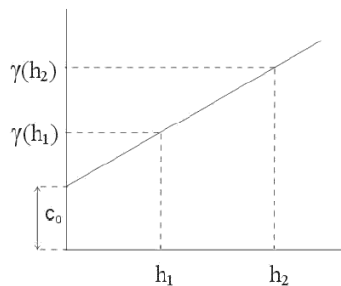
$$\gamma(h) = c_0 + bh \text{ pro } h \leq a$$

$$\gamma(h) = c_0 + c_1 \dots \dots \dots \text{ pro } h > a$$

3.3.2 Modely bez přechodu

Modely bez přechodu (netransitivní) - nemají prahovou hodnotu (sill) v rámci studované plochy a lze je popsat např. lineárním modelem. Výskyt těchto modelů si lze zjednodušeně představit jako určitý extrémní případ klasického přechodového modelu. Představme si, že bychom u něho prováděli výpočet semivariogramu jen do vzdálenosti nepřesahující rozpětí d. Pak bychom při vynesení bodů nenašli žádnou oblast stabilizace křivky semivariogramu a daný případ bychom interpretovali jako model bez přechodu.

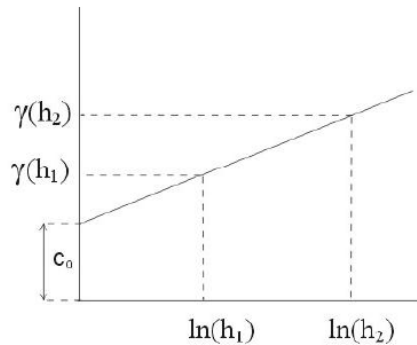
Lineární model:



Obr. 3.14 Lineární model semivariogramu

$$\gamma(h) = c_0 + bh, \text{ kde } b \text{ je směrnice přímky}$$

Logaritmický semivariogram



Obr. 3.15 Logaritmický model semivariogramu

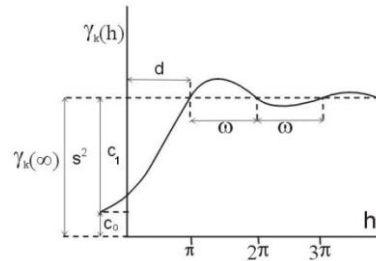
$$\gamma(h) = c_0 + b \ln(h), \text{ kde } b \text{ je směrnice přímky}$$

3.3.3 Oscilační modely

Oscilační modely - oscilační (tj. nehomogenní) charakter má zkoumané pole nejčastěji v důsledku pravidelného střídání pásů s vyššími a nižšími hodnotami. Průměrná šířka pásů se dá odhadnout podle rozměru poloviny periody vlny. U těchto modelů se často projevuje nestabilita. Nepoužívají se pro odvození parametrů potřebných pro krigování (upřednostňují se robustní, jednoduché přechodové modely).

Jev, kdy hodnoty semivariogramu v jisté vzdálenosti delší než dosah začnou opět klesat či vykazují více lokálních minim ukazuje na periodická kolísání v hodnotách atributu a označuje se jako **hole effect**.

Sinový model semivariogramu

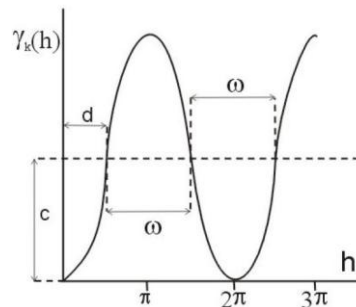


Obr. 3.16 Sinový model semivariogramu

$$\gamma(h) = c_0 + c_1 \left[1 - \frac{\sin(gh)}{gh} \right] \quad \text{kde } g = \pi / \omega$$

Hodnota *sin* se udává v radiánech. Dochází k postupnému tlumení hodnot oscilací. Hodnota ω udává průměrný rozměr bohatších a chudších úseků.

Cosinový model semivariogramu - Nedochozí k postupnému tlumení hodnot oscilací.



Obr. 3.17 Cosinový model semivariogramu

$$\gamma(h) = c_0 + c_1 [1 - \cos(gh)] \quad \text{kde } g = \pi / \omega$$

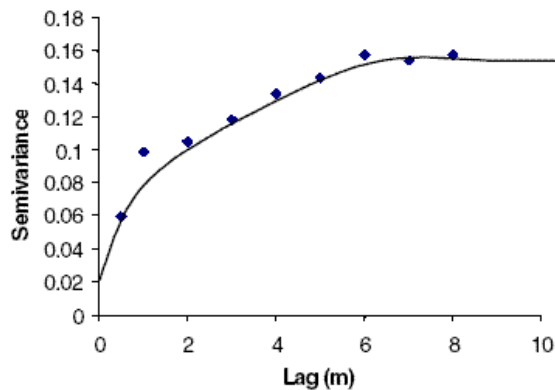
Čistě náhodný model semivariogramu

$$\gamma(h) = c_0$$

Semivariogram nemá žádnou úvodní rostoucí větev, hodnoty často pouze kolísají kolem prahu. K této situaci dochází, když je studované pole příliš variabilní vzhledem ke zvolenému kroku vzorkování (zjišťování hodnot).

3.4 Další druhy semivariogramů

3.4.1 Složené modely (komplexní semivariogram)



Obr. 3.18 Složený model semivariogramu

$$\gamma_T(h) = \gamma_1(h) + \gamma_2(h) + \gamma_3(h) + \dots$$

Prostorová kolísání v závislosti na odlišných typech povrchů (cover classes) – svoji vlastní strukturu prostorového uspořádání a autokorelace hodnot proměnné mohou mít rozdílné kategorie landuse, druhy půd, atd. V tomto případě mohou být modely sestavené pro jednotlivé třídy vhodnější než model globální. Je zde však často problém dostatku dat.

Indikátorové semivariogramy se konstruují a využívají při strukturální analýze nominálních (kvalitativních) dat (barva, druh horniny). Primární data se transformují do hodnot 1 a 0 podle splnění indikační podmínky – např. zda je hornina pískovcem. Často slouží jako vstup pro tzv. indikátorové krigování (viz. dále).

Soft semivariogramy se využívají při v případě nedostatku primárních dat, kdy je možné na základě provedené simulace doplnit další data a usnadnit provedení strukturální analýzy. Interpretace a verifikace je však dosti nesnadná a vyžaduje větší zkušenosti. Soft semivariogramy se často používají při provádění soft krigingu (viz. dále).

3.5 Analýza a interpretace strukturálních funkcí

Pro každý model existují vlastní pravidla interpretace. Konstrukci semivariogramu a odvození teoretického modelu by měla vždy předcházet důkladná analýza vstupních dat založená na metodách popisné statistiky (ESDA – explorační analýza prostorových dat, viz. dále)

Pro korektní odhady vhodného teoretického modelu je důležitý počet bodů uvažovaných pro vyjádření hodnot semivariance pro daný lag (h). Proto se často hodnoty teoretického modelu odhadují za pomoci vážené metody nejmenších čtverců, kdy jako váhy se berou počty párů na dané vzdálenosti h . Značný podíl šumu ve variogramu může být dále způsoben malým rozsahem vzorku použitého k výpočtu $\hat{\gamma}(h)$

K dosažení stabilních hodnot se doporučuje 20 – 30, v některých případech však až až 50-100 hodnot. Je-li jejich počet nízký, stoupá chyba odhadu. Hladší průběh semivariogramu lze docílit zvětšením velikosti vyhledávacího okna (větším h). O velikosti okna vypovídá hodnota dosahu (range). Je-li odhadnutý dosah z variogramu příliš malý a všechny body jsou dále jak dosah, potom nejlepším odhadem je použití celkového průměru. Vzdálenosti mohou být modifikovány efektem anizotropie - potom je nutné měnit tvar okolí. Anizotropie však může být výsledkem i nedostatečného počtu vzorků.

Výpočet experimentálních semivariogramů se doporučuje provádět do vzdálenosti $h \leq L/2$, kde L je maximální vzdálenost míst pozorování v poli.

Vždy je vhodné upřednostňovat jednodušší teoretický model semivariogramu, který dobře vystihuje hlavní rysy experimentálních hodnot, před modelem složitějším.

V případě výpočtu experimentálního semivariogramu z nepravidelně sítě pozorování je nutno počítat s vyšší „rozkolísaností“ stanovených bodů kolem teoretického modelu.

Úroveň prahu se obvykle doporučuje volit podle hodnoty statistického rozptylu.

Je-li hodnota dosahu použitého teoretického semivariogramu malá vzhledem k hodnotám empirickým je možné zmenšit hodnotu kroku h a naopak

Při prokládání tečny počátkem experimentálního semivariogramu pro určení rozpětí musíme respektovat skutečnost, že funkce semivariogramu je vždy kladná. Hodnota rozpětí je důležitá pro aplikaci oscilačních semivariogramů.

Při interpretaci zbytkového rozptylu musíme uvážit i možný vliv chyb měření (technických chyb) výchozích pozorování.

Výběr vhodného teoretického modelu musí vycházet z cíle analýzy. Je-li cílem odhalení strukturálních úrovní a podrobný popis všech charakteristik studovaného pole, pak je nutno podrobně analyzovat chování v celém reálném průběhu experimentálního semivariogramu. Jestliže je interpretace prováděna pro účely návazných krigovacích výpočtů, je účelné zvolit pokud možno jednoduchý a robustní model, vystihující chování a okolí počátku až do úrovně prahu.

Při interpretaci je důležité vycházet z dobré znalosti objektu v krajinné sféře a z využití všech informací o jeho parametrech.

Při analýze anizotropie je podle zkušenosti dobré volit pro všechny směrové semivariogramy – samozřejmě pokud je to možné – stejný teoretický model. Proto je výhodné vyjít z izotropního semivariogramu pole.

V případě anizotropního pole se zpravidla snažíme využít předpokladu geometrické anizotropie, kterou lze snadno eliminovat transformací souřadného systému.

Obecně je účelné postupovat tak, že v počáteční fázi aplikace geostatistických metod na přírodní objekt se provede podrobná interpretace strukturálních funkcí a v následných fázích se podle získaných zkušeností použije zjednodušený základní model.

Analýza semivariogramu je podstatným krokem k určení optimálních vah pro interpolaci. Jestliže ve semivariogramu dominuje náhodná složka (ϵ''), potom data obsahují takový šum, že interpolace nemá smysl. Jako nejlepší odhad $z(x)$ je vhodné použít průměrnou hodnotu.

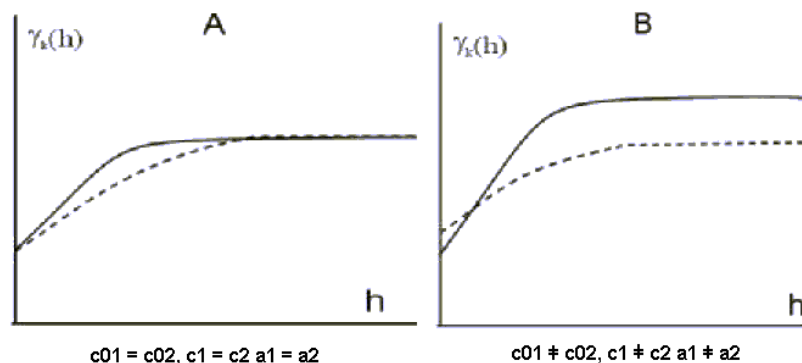
Charakteristiky pole popsané strukturní analýzou:

Kontinuita – je vyjádřena hodnotou dosahu semivariogramu. Pole s větší kontinuitou se vyznačuje vyšší prostorovou autokorelací.

Nehomogenita – projevuje se tzv. oscilací hodnoty prahu. Délka poloviny periody odpovídá průměrnému rozměru elementů nehomogenity. Nehomogenity na dané úrovni pozorování nepostizitelné se projeví jako zbytkový rozptyl.

Nestacionarita - projevuje se zpravidla parabolickým nárůstem křivky semivariogramu. Prokazatelná je případech, kdy dochází k parabolickému růstu křivky až za hodnotou dosahu, tedy na stabilizované části křivky. Nestacionarita pole dokládá změnu průměrné hodnoty proměnné v poli. Ze vzdálenosti, kde se začne deformace křivky semivariogramu projevovat, lze určit vzdálenost, do které jsou změny průměrné hodnoty v poli zanedbatelné.

Anizotropie - lze ji popsat pomocí modelů jednotlivých směrových semivariogramů (tj. semivariogramů vypočtených na různých směrech v poli). Projevuje se změnami parametrů (dosahu, prahu, zbytkového rozptylu), jednak v rozdílech typů směrových semivariogramů. Jak bylo uvedeno výše rozlišujeme geometrickou a zonální anizotropii (viz. obr).



Obr. 3.19 Rozdíl mezi geometrickou (A) a zonální (B) anizotropií semivariogramu

4. KRIGING – geostatistické metody interpolace

Krigování je základní geostatistickou metodou určování lokálního odhadu. Metoda se často označuje akronymem **BLUE** (Best Linear Unbiased Estimator – tedy nejlepší lineární nezkraslený odhad). Toto označení má vystihnout výchozí podmínky krigování:

- odhadovaná hodnota je vypočtena jako lineární kombinace vstupních hodnot:

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i \cdot z(x_i)$$

kde pro váhy platí

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

- nezkreslený (nestranný) odhad značí, že průměrná chyba tohoto odhadu je rovna nule

$$\sum (\hat{z}_i - z_i) = 0$$

- je minimalizován rozptyl odhadu

$$\sum (\hat{z}_i - z_i)^2 = \min.$$

Pokud prostorově závislá náhodná kolísání nejsou překryta nekorelovaným šumem, potom může být semivariogram využit k určení vah λ_i potřebných pro interpolaci. Procedura je podobná jako v případě metody vážených klouzavých průměrů s tím rozdílem, že právě váhy jsou odhadnuty geostatistickými metodami. Váhy λ_i jsou zvoleny tak, aby odhad $\hat{z}(x_0)$ byl nestranný a odhad rozptylu σ_e^2 byl menší, než jakákoliv jiná lineární kombinace pozorovaných hodnot (minimální).

Přitom pro minimální rozptyl hodnot $\hat{z}(x_0)$ platí výraz :

$$\hat{\sigma}_e^2 = \sum_{i=1}^n \lambda_i \gamma(x_i, x_0) + \phi$$

kde:

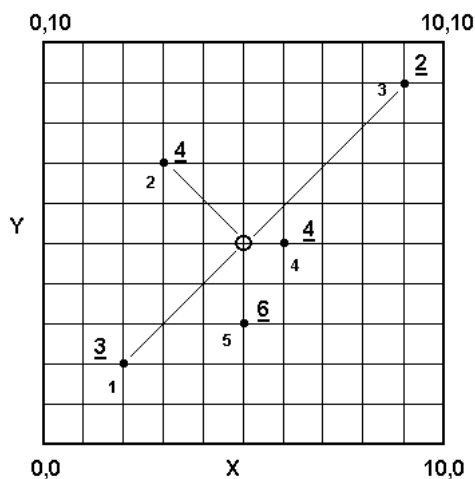
$$\sum_{i=1}^n \lambda_i \gamma(x_i, x_j) + \phi = \gamma(x_j, x_0) \text{ pro všechna } j.$$

Hodnota $\gamma(x_i, x_j)$ je semivariance proměnné z mezi body x_i a x_j . Hodnota $\gamma(x_i, x_0)$ je semivariance proměnné z mezi bodem x_i a bodem x_0 , pro který je hodnota proměnné z zjišťována. Obě hodnoty lze získat z vhodného teoretického modelu semivariogramu. Hodnota ϕ je tzv. Lagrangeův multiplikátor, který zajišťuje požadavek minimalizace odchylek a zároveň podmínku, že suma vah je rovna jedné.

Uvedená metoda se označuje jako **základní (ordinary) kriging** a je možné ji použít pro interpolaci v pravidelné mřížce hodnot, ke konstrukci map (např. izolinií).

PŘÍKLAD: Výpočet neznámé hodnoty v bodě metodou základního krigingu.

Na základě změřených hodnot veličiny Z v pěti bodech ($i=1, \dots, 5$) (viz. obrázek) máme za úkol odhadnout hodnotu Z bodě ($i=0$) o souřadnicích ($x=5, y=5$) metodou krigingu.



Obr. 4.1 Vstupní data pro lokální odhad metodou základního krigování (podtržená čísla značí hodnotu atributu v bodě)

Na základě předem provedené strukturální analýzy použijeme **sférický semivariogram**.

$$\gamma(h) = c_0 + c_1 * \left[\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] \dots \dots \dots \text{pro } h \leq a$$

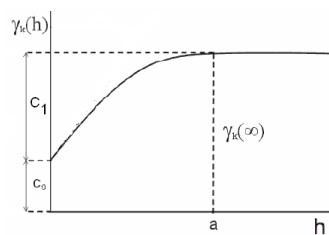
$$\gamma(h) = c_0 + c_1 \dots \dots \dots \text{pro } h > a$$

Parametry použitého teoretického semivariogramu jsou:

$$c_0 = 2,5$$

$$c_1 = 7,5$$

$$a = 10,0 \text{ (dosah)}$$



Data v pěti měřených bodech mají následující souřadnice

i	x	y	z
1	2	2	3
2	3	7	4
3	9	9	2
4	6	5	4
5	5	3	6

Pokud budeme dále značit:

A – matice semivariací mezi všemi dvojicemi bodů
 b – vektor semivariací mezi všemi body a bodem predikovaným
 λ – vektor vah jednotlivých bodů
 Φ – tzv. Lagrangeův člen
 potom základní vztah pro odhad metodou krigování lze psát jako:

$$A \cdot \lambda = b$$

Pro vlastní řešení je nutné vypočítat váhy λ , které musí splňovat podmínku $\sum \lambda = 1$
 Uvedený základní vztah lze vyjádřit jako soustavu rovnic:

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdot & \cdot & \cdot & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \cdot & \cdot & \cdot & \gamma_{2n} & 1 \\ \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & & & & & \cdot \\ \cdot & \cdot & & & & & \cdot \\ \gamma_{n1} & \gamma_{n2} & \cdot & \cdot & \cdot & \gamma_{nn} & 1 \\ 1 & 1 & \cdot & \cdot & \cdot & 1 & 0 \end{bmatrix} * \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_n \\ \phi \end{bmatrix} = \begin{bmatrix} \gamma_{10} \\ \gamma_{20} \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{n0} \\ 1 \end{bmatrix}$$

V tomto zápisu poslední řádek a poslední sloupec v první matici a hodnota Lagrangeova členu Φ jsou použity pro zajištění podmínky sumy vah $\sum \lambda = 1$. Hodnota Lagrangeova multiplikátoru Φ také slouží pro výpočet rozptylu odhadnuté hodnoty. Uvedená soustava rovnic nám poskytne hodnoty všech vah λ a hodnotu Φ . V maticovém zápisu lze tedy psát:

$$A^{-1} \cdot b = \begin{bmatrix} \lambda \\ \phi \end{bmatrix}$$

Aby bylo možné vyčíslit hodnoty semivariací, je v prvním kroku zapotřebí vytvořit matici vzdáleností mezi datovými body:

i	1	2	3	4	5
1	0,000	5,099	9,899	5,000	3,162
2	5,099	0,000	6,325	3,606	4,472
3	9,899	6,325	0,000	5,000	7,211
4	5,000	3,606	5,000	0,000	2,236
5	3,162	4,472	7,211	2,236	0,000

Vektor vzdáleností mezi měřeními body a bodem predikovaným:

i	0
1	4,234

2	2,828
3	5,657
4	1,000
5	2,000

Těchto vzdáleností využijeme k výpočtu semivariací pro sférický model semivariogramu s výše uvedenými parametry c_0 , c_1 , a – tedy k sestavení matice A a vektoru b :

Matice A :

i	1	2	3	4	5	6
1	2,500	7,739	9,999	7,656	5,939	1,000
2	7,739	2,500	8,667	6,381	7,196	1,000
3	9,999	8,667	2,500	7,656	9,206	1,000
4	7,656	6,381	7,656	2,500	4,936	1,000
5	5,939	7,196	9,206	4,936	2,500	1,000
6	1,000	1,000	1,000	1,000	1,000	0,000

Ve výše uvedené matici má řádek navíc ($i=6$) zajistit podmínku, že váhy budou mít sumu rovnu jedné.

Vektor b :

i	0
1	7,151
2	5,597
3	8,815
4	3,621
5	4,720
6	1,000

Inverzní matce A^{-1} :

i	1	2	3	4	5	6

1	-,172	,050	,022	-,026	,126	,273
2	,050	-,167	,032	,077	,007	,207
3	,022	,032	-,111	,066	-,010	,357
4	-,026	,077	,066	-,307	,190	,030
5	,126	,007	-,010	,190	-,313	,134
6	,273	,207	,357	,003	,134	- 6,873

Řešením výše uvedené soustavy rovnic lze pro jednotlivé vzdálenosti získat hodnoty vah λ :

i	λ	
1	0,0175	
2	0,2281	
3	- 0,0891	vypočtené hodnoty vah
4	0,6437	
5	0,1998	
6	0,1182	vypočtená hodnota Φ

Pro váhy $i=1, \dots, 5$ platí, že jejich suma se rovná jedné, v posledním řádku je hodnota Lagrangeova členu Φ .

Vzdálenosti měřených bodů od bodu predikovaného, již přísluší výše určené váhy:

i	0
1	4,234
2	2,828
3	5,657
4	1,000
5	2,000

Potom odhad hodnoty Z v bodě ($i=0$) o souřadnicích ($x=5, y=5$):

$$Z(x_{i=0}) = 0,0175*3 + 0,2281*4 - 0,0891*2 + 0,6437*4 + 0,1998*6 =$$

$$Z(x_{i=0}) = \mathbf{4,560}$$

Rozptyl odhadu:

$$\begin{aligned}\sigma_e^2 &= [0,0175*7,151+0,2281*5,597-0,0891*8,815+0,6437*3621+0,1998*4,720]+ \varphi = \\ \sigma_e^2 &= 3,890 + 0,1182 = \\ \sigma_e^2 &= 4,008\end{aligned}$$

4.1 Typy krigování

Na rozdíl od deterministických metod interpolace nabízí metody krigingu vedle odhadů vlastní interpolované hodnoty také odhady pravděpodobnosti výskytu těchto hodnot a dále odhady chyb predikce. Pro charakterizování jednotlivých typů krigingu předpokládejme jednoduchý model:

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

kde $Z(x_i)$ je proměnná v bodě x_i , která se skládá z deterministické hodnoty trendu $\mu(x_i)$ a autokorelované náhodné proměnné $\varepsilon(x_i)$. Protože ve většině případů hodnotu trendu často pouze odhadujeme a neznáme ji přesně, je určitou chybou zatížena též náhodná složka. Pro tuto platí, že její průměrná chyba je rovna nule a autokorelace hodnot $\varepsilon(x_i)$ $\varepsilon(x_i+h)$ nezávisí na aktuální pozici, ale pouze na hodnotě vzdálenosti h .

Hodnota trendu může nabývat konstantní hodnoty ($\mu(x_i) = \mu$). V závislosti na tom, zda hodnota μ představuje konstantu či zda data obsahují trendovou složku a v závislosti na tom, za hodnota μ představuje zámou hodnotu či zda ji odhadujeme definujeme různé metody (modely) krigování. Mezi základní metody krigování, kterými se provádí odhad na základě přímo naměřených hodnot patří především:

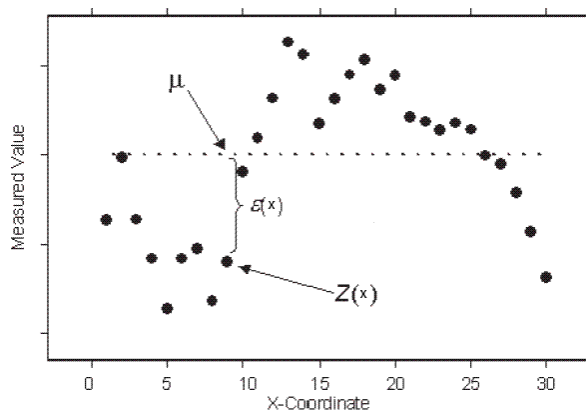
- základní (ordinary) krigování s bodovým odhadem
- základní (ordinary) krigování s blokovým odhadem
- jednoduché (simple) krigování
- univerzální krigování
- pravděpodobnostní krigování
- co-kriging
- lognormální krigování.

4.1.1 Základní krigování (ordinary kriging)

Obecný model základního krigování:

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

kde μ je neznámá hodnota trendu.



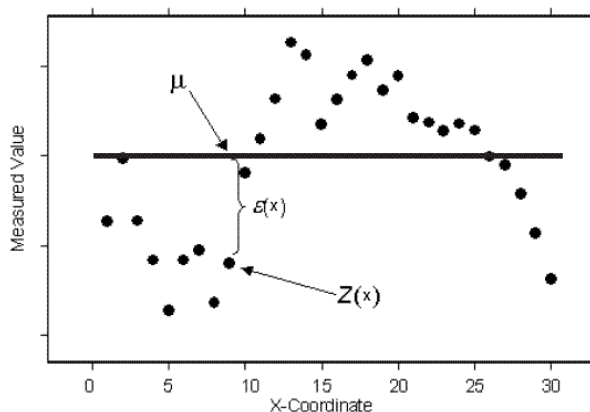
Obr. 4.2 Princip základního krigování

4.1.2 Jednoduché krigování (Simple kriging)

Nejjednodušší variantou krigování je tzv. **jednoduché krigování** (simple kriging). K výpočtu je potřebná znalost průměrné hodnoty veličiny v poli (μ) a dále předpoklad stacionarity. Obecný model jednoduchého krigování má opět tvar:

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

kde μ je známá konstanta.



Obr. 4.3 Princip základního krigování

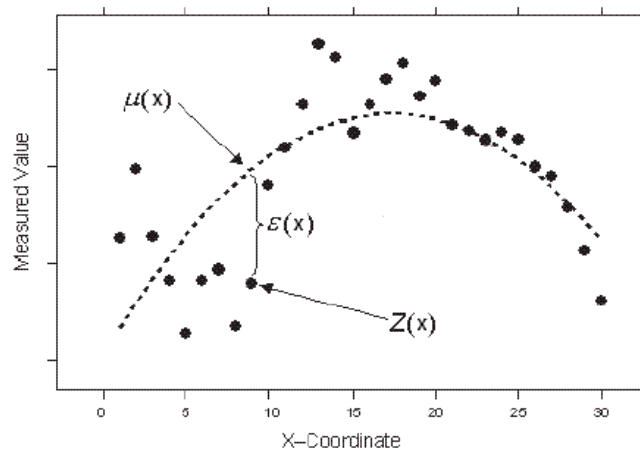
V tomto modelu, protože známe hodnotu μ , potom v bodech měření známe přesně také $\varepsilon(x)$. V případě základního krigování jsou obě hodnoty odhadovány. Základní krigování tedy nabízí přesnější odhad autokorelace, předpoklad znalosti μ je však často nereálný. Často se však používá některá z běžných trendových funkcí, kterou se nejprve vyjádří rezidua. Trend reziduí se potom považuje za nulový a aplikuje se základní krigování.

4.1.3 Univerzální krigování (Universal kriging)

Obecný model univerzálního krigování má opět tvar:

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

kde $\mu(x)$ je deterministická funkce - např. polynom druhého stupně jako na přiloženém obrázku. Pokud odečteme hodnotu polynomu od originálních dat, dostaneme chybovou složku $\varepsilon(x)$, která má charakter náhodné proměnné s průměrem rovným nule. Autokorelace je modelována právě z této náhodné proměnné. Jak je patrné z obrázku, univerzální kriging je v tomto případě analogií regresní závislosti. Na rozdíl od regrese, kde složku $\varepsilon(x)$ považujeme za nekorelovanou, v případě krigování ji modelujeme jako složku autokorelovanou. Stejně jako v případě základního krigingu, vhodná dekompozice na obě výše uvedené složky ze samotných dat nelze provést. Univerzální kriging používá k popisu autokorelované náhodné složky semivariogramu nebo kovariance.



Obr. 4.4 Princip univerzálního krigování

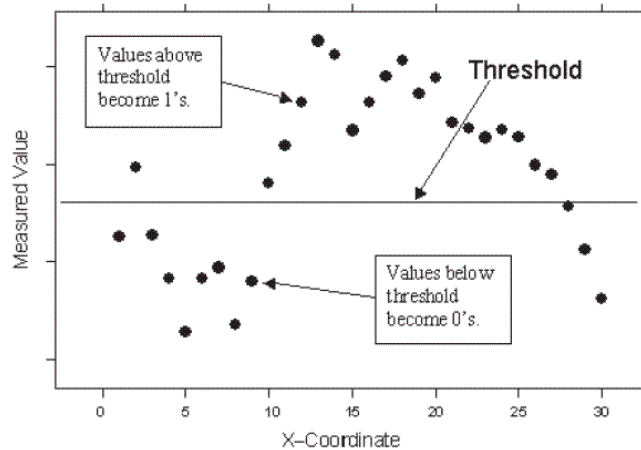
4.1.4 Indikátorové krigování

V řadě úloh nás nezajímá, zda je v daném místě nebo na dané ploše nejpravděpodobnější průměrná koncentrace 0,016 nebo 0,015, ale odhad pravděpodobnosti, s jakou je překročena limitní hodnota např. 0,012. Tyto úlohy řeší tzv. indikátorové krigování, které náleží do skupiny neparametrických geostatistických metod. Dále sem patří také soft kriging a pravděpodobnostní krigování. Pro tyto metody krigingu se zavádí pojem prahové hodnoty, kterou lze spojitou proměnnou převést na binární (viz. obr.)

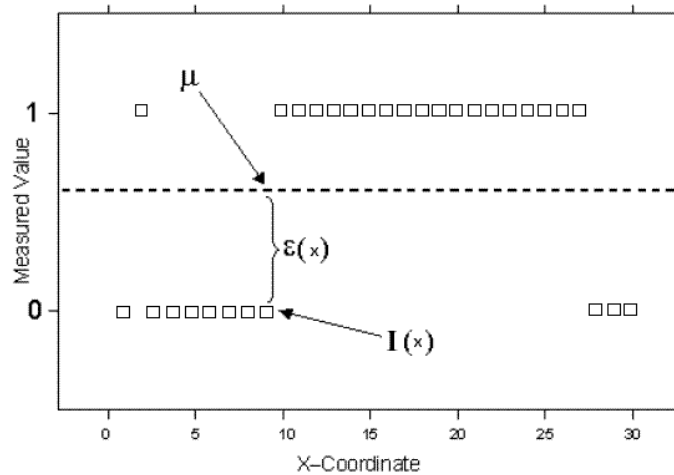
Indikátorové krigování předpokládá model ve tvaru:

$$I(x_i) = \mu + \varepsilon(x_i)$$

kde μ neznámá konstanta, $\varepsilon(x)$ autokorelovaná náhodná proměnná a $I(x)$ je buďto přímo zjištěná binární proměnná a nebo binární proměnná, kterou obdržíme ze spojitých dat prahováním. Jinak se model neliší od základního krigování. Protože indikátorová proměnná nabývá hodnot 0 nebo 1, potom interpolované hodnoty budou nabývat hodnot od 0 do 1 a lze je interpretovat jako pravděpodobnost, že proměnná nabývá hodnoty 1. Pokud bylo pro vytvoření indikátorové proměnné použito postupu prahování, potom lze vytvořenou mapu interpretovat jako pravděpodobnost překročení prahové hodnoty.



Obr. 4.5 Princip prahování hodnot proměnné



Obr. 4.6 Princip indikátorového krigování

Uvedený princip lze obecně rozšířit a použitím dvou více hodnot vytvořit např. dvě indikátorové proměnné (viz. dále - co-kriging).

4.1.5 Pravděpodobnostní krigování (probability kriging)

Pravděpodobnostní krigování předpokládá model ve tvaru:

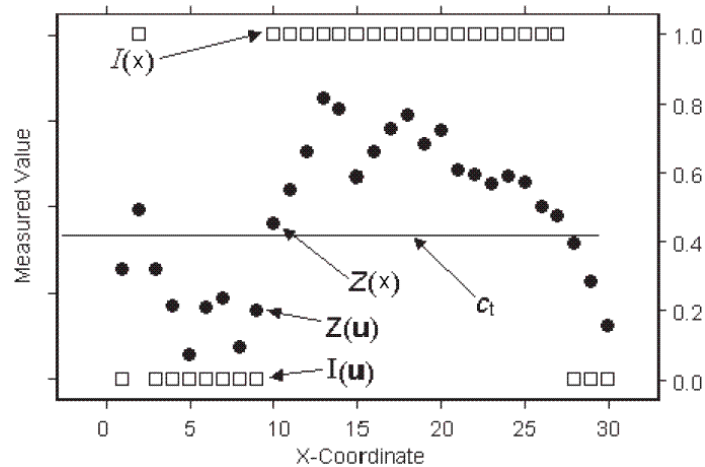
$$I(x) = I(Z(x) > c_1) = \mu_1 + \varepsilon_1(x)$$

$$Z(x) = \mu_2 + \varepsilon_2(x)$$

kde μ_1 a μ_2 jsou neznámé konstanty. $I(x)$ je binární proměnná vytvořená indikátorovým prahováním ($I(Z(x) > c_1)$). V tomto případě dostáváme dvě náhodné chyby $\varepsilon_1(x)$ a $\varepsilon_2(x)$. Cíle pravděpodobnostního krigování jsou stejné jako u krigování indikátorového, jsou však dosaženy využitím konceptu co-krigingu.

Na obrázku 7 má datový bod $Z(u=9)$ hodnotu indikátorové proměnné $I(u)=0$ a bod $Z(x=10)$ hodnotu $I(x)=1$. Pokud bychom chtěli predikovat hodnotu v polovině vzdálenosti mezi oběma body – na x-ové souřadnici 9,5, potom použitím modelu indikátorového krigování bychom obdrželi hodnotu 0,5. Z obrázku je však patrné, že datový bod $Z(x)$ je nepatrně nad hodnotou

prahu, naopak bod $Z(u)$ je výrazně pod prahovou hodnotou. Je tedy reálné předpokládat, že predikovaná proměnná v bodě 9,5 bude méně než 0,5.



Obr. 4.7 Princip pravděpodobnostního krigování

Pravděpodobnostní krigování se tedy snaží využít vedle indikátorové proměnné ještě další extra informace v původních datech. Nevýhodou pravděpodobnostního krigování je nutnost provádět odhady jako autokorelací pro jednotlivé proměnné, tak křížových korelací mezi nimi. Dalšími odhady neznámých autokorelací se vnáší do výsledného modelu větší míra nejistoty.

4.1.6 Nelineární kriging (log-normal)

Pokud nemají vstupní data normální rozdělení, je nutné je před vlastní interpolací transformovat. Nejběžnější je transformace lognormální. Originální data jsou transformována na přirozený logaritmus o základu 10. Tedy modelování variogramu a interpolace probíhá s proměnou $y(u)$:

$$y(u) = \ln z(u)$$

Predikované hodnoty je poté nutno transformovat nazpět, což může působit problémy (viz. Borrough et. al. 1992) a jako alternativa se nabízí indikátorový kriging Pro některá FG data, která vykazují rozdělení s kladnou asymetrií, je však lognormální transformace výhodná (např. obsah chemických látek v půdě).

4.1.7 Kriging s využitím externí informace

K interpolaci kromě hodnot vlastní interpolované proměnné lze využít například:

1. vhodnou stratifikaci dat (**stratifikovaný kriging**)
2. hodnoty jiné proměnné, která koreluje s původní a kterou lze snadno měřit ve větším počtu bodů (např. výškové poměry) – **co kriging**
3. fyzikální či empiricky sestavený model, který podmiňuje rozložení hodnot studované proměnné

Stratifikovaný kriging spočívá v rozdělení oblasti na subregiony. Předpokládá dostatečný počet bodů pro výpočet hodnot variogramu. Může dávat vhodnější odhady, je však nutné řešit oblasti na styku subregionů. Např. obsah znečišťujících látek podle oblastí zaplavovaných podél vodního roku s různou frekvencí.

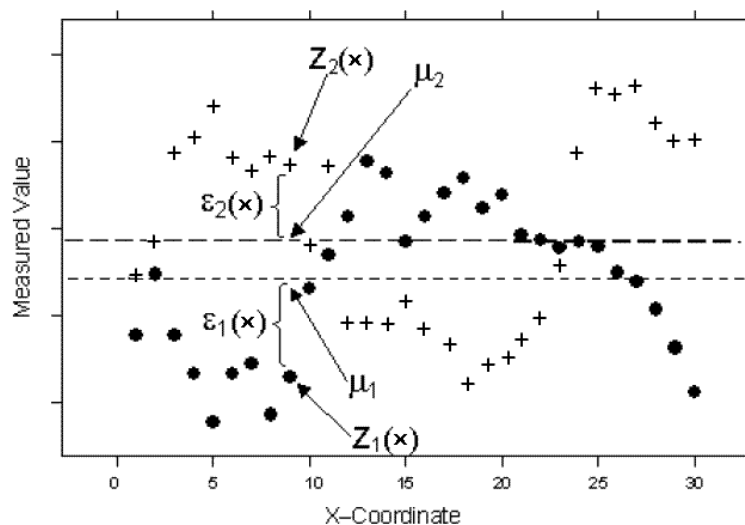
4.1.8 Co-kriging

Máme dvě proměnné z_1 a z_2 , které vykazují prostorovou korelaci. Pak lze využít hodnot proměnné z_2 , k interpolaci hodnot proměnné z_1 . Tento koncept je vhodný zvláště v případech, kdy je proměnná z_2 snáze získatelná a rozšiřitelná i na více než dvě proměnné. Přitom pro přesnější odhady se používá jak autokorelace jednotlivých proměnných, tak vzájemné (cross) korelace všech použitých proměnných. Základní co-kriging využívá následujících modelů:

$$Z_1(x) = \mu_1 + \varepsilon_1(x)$$

$$Z_2(x) = \mu_2 + \varepsilon_2(x)$$

kde μ_1 a μ_2 jsou neznámé konstanty. Dále dostáváme dvě náhodné chyby $\varepsilon_1(x)$ a $\varepsilon_2(x)$. Základní co-kriging odhaduje hodnotu proměnné $Z_1(x_0)$ stejně jako základní krigování, ovšem navíc využívá kovariance s hodnotou $Z_2(x)$.



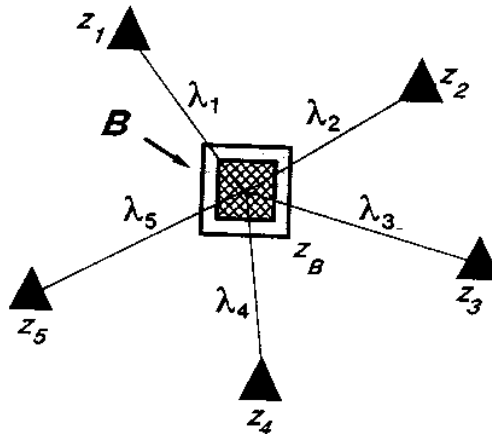
Obr. 4.8 Princip co-kringu

Z obrázku je patrné, že data Z_1 a Z_2 se jeví jako nekorelovaná. Dále pokud Z_1 je pod průměrem μ_1 , potom Z_2 je často nad průměrem μ_2 a naopak. Tedy Z_1 a Z_2 vykazují negativní cross korelaci. Vedle základního co-kringu jsou dalšími variantami např. jednoduchý, univerzální, indikátorový či pravděpodobnostní co-kriging.

4.1.9 Blokový odhad při základním krigování (Block kriging)

Lokální (bodový) odhad metodou krigingu lze určitým způsobem vztáhnout k ploše či objemu v prostoru interpolovaných dat. Mnoho přírodních jevů vykazuje značnou variabilitu a výsledkem bodového odhadu může být mapa obsahující značný počet ostrých vrcholů a depresí. Tento efekt lze potlačit tak, že modifikujeme výše uvedené rovnice a odhadneme

průměrnou hodnotu $z(B)$ proměnné z pro jistou plochu či objem B (viz. obr). Tato modifikace je vhodná, pokud výsledkem interpolací být struktura pravidelných buněk (grid).



Obr. 4.9 Princip blokového krigování

Průměrná hodnota z pro blok B

$$z(B) = \int_B \frac{z(x)dx}{\text{plocha}_B}$$

bude odhadnuta z výrazu:

$$\hat{z}(B) = \sum_{i=1}^n \lambda_i \cdot z(x_i)$$

Kde stejně jako u bodového odhadu je suma všech vah λ_i rovna jedné.

Minimální rozptyl nyní bude:

$$\hat{\sigma}^2(B) = \sum_{i=1}^n \lambda_i \bar{\gamma}(x_i, B) + \phi - \bar{\gamma}(B, B)$$

a získáme ho, když

$$\sum_{i=1}^n \lambda_i \gamma(x_i, x_j) + \phi = \bar{\gamma}(x_j, B) \text{ pro všechna } j.$$

Rozptyly odhadů pro blokový kriging jsou daleko menší než pro bodový kriging. Výsledný interpolovaný povrch je obecně více shladený a neobsahuje takové množství lokálních extrémů. Blokové krigování je aproximující metodou.

4.2 Hodnocení a verifikace modelů

Krigování jako interpolační metoda umožňuje pro každý interpolovaný bod odhadnout potenciální velikost chyby odhadu. Vedle map predikovaných hodnot tak lze především konstruovat mapy hodnot σ_e^2 (rozptyl krigingu), které vypovídají o spolehlivosti interpolovaných hodnot. Tyto hodnoty se obvykle prezentují v podobě map druhé mocniny

σ_e^2 - tzv. **směrodatné chyby** (odchylky) **krigingu (Standard error map)**, protože tyto mají stejné jednotky jako predikované hodnoty. V některých případech se stanovuje také tzv. přesnost (relativní chyba) odhadu:

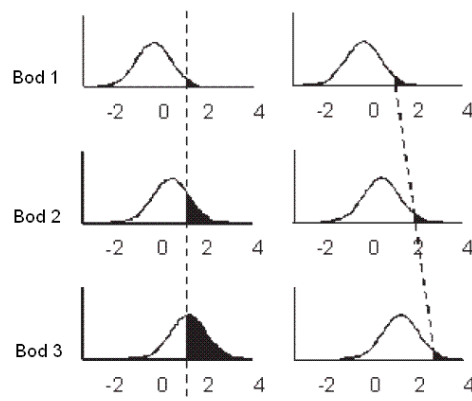
$$\delta_k = 100 \frac{\sigma}{\hat{z}}$$

Vyjdeme-li z výše uvedeného příkladu, kdy rozptyl odhadu je $\sigma_e^2 = 4,008$. Potom směrodatná chyba krigingu bude $\sigma_e = 2,002$. Budeme-li předpokládat, že chyby predikce mají normální rozdělení, potom 95% interval spolehlivosti predikovaných hodnot lze určit z následujícího vztahu:

$$Z(x_0) \pm 1,96 \cdot \sqrt{\sigma_e^2}$$

kde $Z(x_0)$ je odhad hodnoty proměnné z v bodě x_0 a σ_e^2 je rozptyl odhadu. V našem případě tedy při opakovaném použití stejného modelu padne 95 % odhadovaných hodnot do intervalu $(4,560 \pm 1,96 \cdot 2,002)$ tj. $(0,64; 8,48)$

Konstrukce dalších dvou typů map, které nabízí např. ArcGIS a kterými lze zhodnotit kvalitu interpolace vychází následujícího obrázku.



Obr. 4.10 Princip konstrukce Probability map a Quantile map (vysvětlivky viz. text)

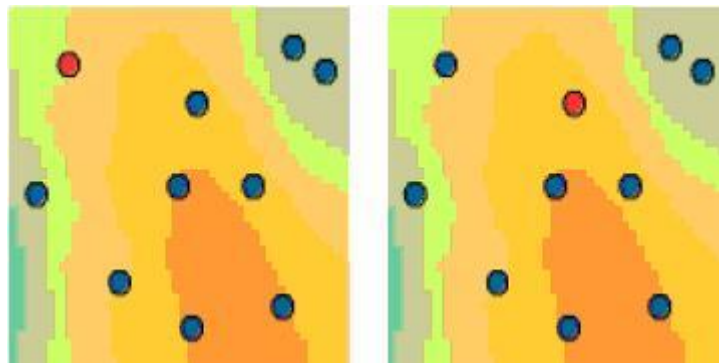
Předpokládáme, že krigováním predikované hodnoty mají ve třech různých bodech normální rozdělení a nacházejí se ve středu každé křivky rozdělení. Chceme-li určit pravděpodobnost, že predikovaná hodnota bude větší než prahová hodnota - např. 1, potom na obrázku vlevo tuto pravděpodobnost představuje na jednotlivých křivkách část plochy vpravo od prahové hodnoty (černé plochy). Při konstantní prahové hodnotě se její pravděpodobnost výskytu pro jednotlivé body mění – tedy lze z ní vytvořit mapu pravděpodobností (**probability map**).

Na obrázku vpravo je schematicky znázorněno, jakým způsobem určit kvantil s např. 5 procentní pravděpodobností výskytu. Tuto pravděpodobnost v tomto případě opět značí černá plocha vpravo od prahové hodnoty a hodnotu kvantitu odečteme na ose x. Při konstantní pravděpodobnosti se budou měnit hodnoty kvantilů a lze je opět prezentovat ve formě kvantilové mapy (**quantile map**).

4.2.1 Validace a křížová validace predikovaných hodnot metodou krigingu

Hodnocení přesnosti interpolace lze provádět také pomocí dále popsaných grafických nástrojů

Křížová validace modelu - k vytvoření spojitého povrchu jsou použita všechna vstupní data v měřených bodech. Poté jsou jednotlivé body měření (červené) po jednom postupně vynechány ze vstupní množiny dat a ze zbývajících (modrých) je vypočtena hodnota v místě vynechaného bodu.



Obr. 4.11 Princip křížové validace modelu

4.2.2 Statistické zhodnocení

Procesem křížové validace obdržíme veličiny, které mají následující význam:

- $\hat{Z}(x_i)$ je predikovaná hodnota pro daný bod x_i , kterou obdržíme v procesu křížové validace
- $\hat{\sigma}_e(x_i)$ je směrodatná chyba predikce, tedy druhá odmocnina z výrazu pro rozptyl krigování:

$$\hat{\sigma}_e^2 = \sum_{i=1}^n \lambda_i \gamma(x_i, x_0) + \phi$$

Pozorované a vypočtené hodnoty jsou následně porovnány dále uvedenými měřeními:

- **MPE – mean prediction error** - průměr rozdílů měřených a predikovaných hodnot - hodnoty chyb odhadů by měly být nestranné – tedy jejich průměr by se měl rovnat nule.

$$MPE = \frac{\sum_{i=1}^n (\hat{Z}(x_i) - z(x_i))}{n}$$

- **RMSPE (root mean square prediction error)** – druhá odmocnina průměrného čtverce vzdálenosti vypočtených hodnot (červené body) od teoretických (zelená přímka v grafech). Tato hodnota slouží k porovnání několika různých modelů. Čím

menší je RMSPE, tím vhodnější je model (tím bližší jsou vypočtené hodnoty hodnotám měřeným).

$$RMSPE = \sqrt{\frac{\sum_{i=1}^n (\hat{Z}(x_i) - z(x_i))^2}{n}}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Z}(x_i) - z(x_i))^2}{n}}$$

- **ASE (average standard error)** - průměrná směrodatná chyba

$$ASE = \sqrt{\frac{\sum_{i=1}^n \hat{\sigma}(x_i)}{n}}$$

Výše uvedené nástroje umožňují posoudit vhodnost modelu a také porovnat více modelů vzájemně mezi sebou.

- **MSPE (mean standardized prediction error)** - průměrná standardizovaná chyba predikce

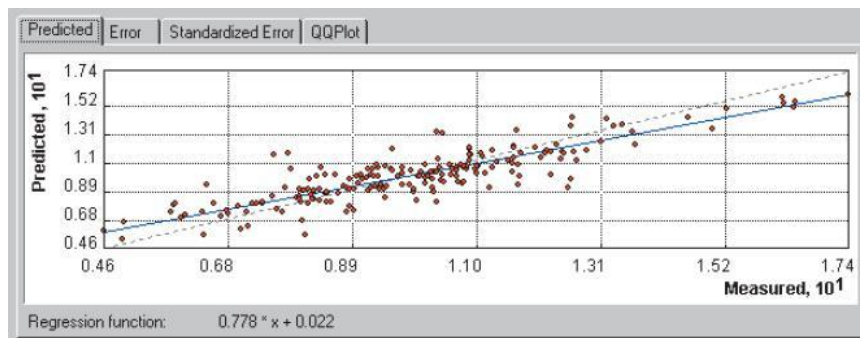
$$MSPE = \frac{\sum_{i=1}^n (\hat{Z}(x_i) - z(x_i)) / \hat{\sigma}(x_i)}{n}$$

- **RMSSPE (root mean square standardized prediction error)**

$$RMSSPE = \sqrt{\frac{\sum_{i=1}^n [(\hat{Z}(x_i) - z(x_i)) / \hat{\sigma}(x_i)]^2}{n}}$$

Validace modelu - vstupní soubor měřených hodnot rozdělí na dvě části – data trénovací a testovací. Trénovací množina dat se použije pro odhad trendu a autokorelačního modelu. Pokud sestavený model vyhovuje trénovacím datům, je ověřen na datech testovacích.

Pro oba zmíněné způsoby ověření vhodnosti modelu se využívá sady grafických nástrojů. Nejběžnějším je graf korelačního pole měřených a predikovaných hodnot. Obecnou vlastností krigingu jako interpolační metody je podhodnocení vysokých hodnot a naopak nadhodnocení hodnot nízkých. Tato vlastnost se projeví menší hodnotou směrnice přímky proložené korelačním polem.



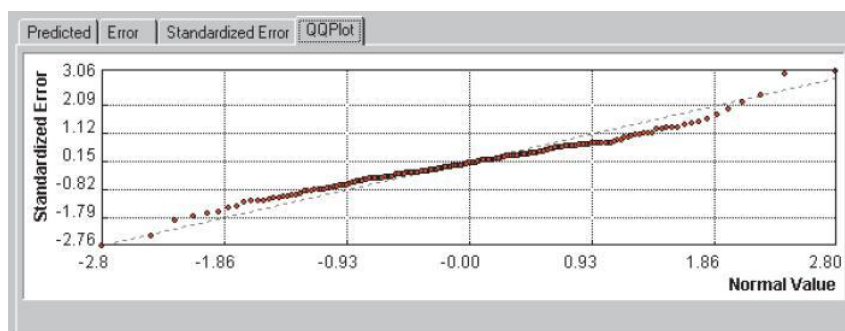
Obr. 4.12 Korelační pole měřených a predikovaných hodnot

Chybový graf (Error plot) – stejný jako předchozí, jsou však vynášeny hodnoty rozdílů mezi měřeními a predikovanými hodnotami

Standardizovaný chybový graf (Standardized Error) – hodnoty rozdílů mezi měřeními a predikovanými hodnotami jsou děleny odhadnutou směrodatnou chybou krigování.

V případě nulové autokorelace budou všechny predikované hodnoty stejné – budou odpovídat průměru a proložená přímka bude mít horizontální průběh. V případě prostorové autokorelace a vhodného modelu krigingu bude proložená přímka totožná s diagonálou a navíc body korelačního pole budou vykazovat malé odchylky od diagonálního směru.

Q-Q graf –znázorňuje graf kvantilů rozdílů mezi měřeními a predikovanými hodnotami dělenými odhadnutou směrodatnou chybou krigování a odpovídajících kvantilů normovaného normálního rozdělení. V případě, že odchylky měřených a odhadnutých hodnot mají normální rozdělení, potom se body v korelačním poli přimykají k přímce (viz. obr.)



Obr. 4.13 Příklad Q-Q grafu

4.3 Interpretace statistických charakteristik k hodnocení vhodnosti modelu:

- **Požadavek nestrannosti odhadu – unbiased** - průměrná chyba odhadu a standardizovaná průměrná chyba odhadu by se měly blížit k nule:
 - $MPE \rightarrow 0$
 - $MSPE \rightarrow 0$
- **Požadavek minimálních chyb** – aby predikované hodnoty byly co nejbliže hodnotám měřeným. Čím menší bude hodnota RMSPE, tím lepší model – tedy tuto podmínku lze použít k porovnání vhodnosti více modelů.
 - $RMSPE \rightarrow \min.$
- **Požadavek vhodné variability predikovaných dat** – variabilita predikovaných hodnot je určována z hodnot měřených. Je tedy důležité, aby i variabilita interpolací vypočtených hodnot byla vhodná:
 - $ASE \approx RMSPE$ – vhodný model (vhodná variabilita predikovaných hodnot)
 - $ASE > RMSPE$ – máš model nadhodnocuje variabilitu odhadnutých hodnot
 - $ASE < RMSPE$ – máš model podhodnocuje variabilitu odhadnutých hodnot

V případě značného podílu šumové složky (např. v důsledku chyb v měření) či v případě značně komplexního povrchu nedává kriging lepší výsledky než jiné interpolátory. Na rozdíl o jiných metod kriging nabízí objektivní, a priori metodu odhadu vhodného okolí pro vlastní interpolaci. Řeší tedy otázku počtu bodů v okolí daného bodu, otázku velikosti a tvaru tohoto okolí. V případě existence bariér (náhlých skoků v hodnotách interpolovaného povrchu nedává kriging dobré výsledky a je nutné jej rozdělit na elementární části neobsahující bariéry.

5. Modelování prostorového uspořádání bodů

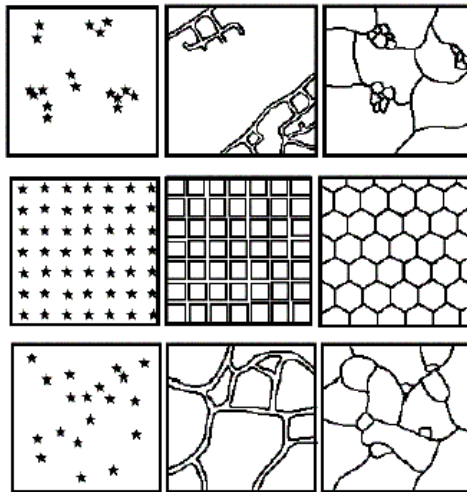
Deskripce bodů pomocí měř úrovně a variability je jen prvním krokem analýzy. V případě prostorové analýzy nás v druhém kroku zajímají body s ohledem na jejich prostorové rozmístění (strukturu - pattern).

Rozmístění bodů je výsledkem určitých procesů a podmínek – např. lokace měst je výsledkem působení faktorů jako je reliéf, přírodní zdroje, komunikace, obdobně výskyt rostlinných druhů, atd.

Cílem studia prostorového rozmístění je zjistit, jak daleko má konkrétní rozmístění objektů k rozmístění teoretickému. (např. teorie centrálních míst – teoretický vzorec – šestiúhelníky). To nám umožňuje jednak porovnávat rozmístění objektů pro různé prostorové jednotky (kategorie landuse, půdní typy, okresy, státy, atd.), jednak studovat dynamiku změn v rámci jedné jednotky (studium dynamiky).

Statisticky prokázaný výskyt určitého prostorového uspořádání (shlukového či pravidelného vzorku) může být základem pro zjišťování příčin, které vedly k pozorovanému uspořádání

5.1.1 Statistická deskripce prostorových vzorů bodových prvků



Obr. 5.1 Základní typy prostorového uspořádání bodů (1. sloupec), linií (2. sloupec) a ploch (3. sloupec). Typy uspořádání: shlukové (1. řádek), pravidelné (2. řádek), náhodné (3. řádek)

Rozlišujeme tři základní typy prostorového uspořádání bodů:

- Shlukové (Clustered)
- Pravidelné (Regular)
- Náhodné (Random)

5.1.2 Základní metody statistického popisu prostorového uspořádání bodů:

- **Analýza kvadrátů** – testujeme, zda rozmístění bodů v ploše je náhodné či nikoliv.
- **Metoda nejbližšího souseda** – porovnává průměrnou vzdálenost mezi nejbližšími sousedy pole bodů vzhledem k teoretickému rozmístění.
- **Metody prostorové autokorelace** – měří, jak podobné či nepodobné jsou hodnoty atributů sousedních bodů.

5.1.3 Problém měřítka, rozsahu studované oblasti a kartografické projekce

Měřítka – je nutné vhodně zvolit tak, aby studovaný jev mohl být prezentován body v prostoru.

Rozsah studované oblasti – v závislosti na zvolené oblasti (často vymezené administrativními hranicemi) se mění jak vzdálenosti mezi jednotlivými body, tak také charakteristiky jejich prostorového uspořádání (Obr. 5.2).



Obr. 5.2 Vliv velikosti studované oblasti na prostorové uspořádání bodů

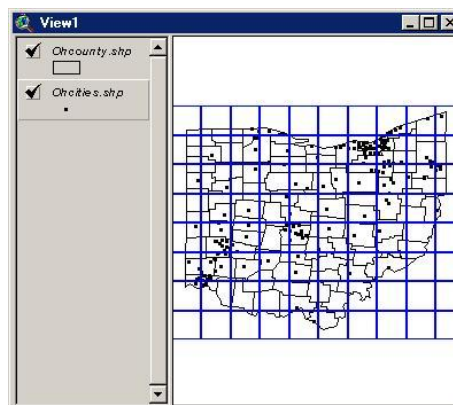
Kartografickou projekci je nutno vhodně zvolit podle účelu (viz. Analýza kvadrátů). Projekcí se mění tvar, vzdálenosti, vzájemná poloha objektů (viz. Obr. 5.3). Čím větší studovaná oblast, tím větší bude role zvolené projekce.



Obr. 5.3 Vliv kartografické projekce na tvar studované oblasti

5.2 Analýza kvadrátů (QUADRAT ANALYSIS)

Metoda pro detekci prostorového uspořádání bodů. Je založena na hodnocení změn hustoty bodů v prostoru. Je porovnáváno, zda rozmístění bodů v prostoru je náhodné, či má blíže k uspořádání shlukovému či pravidelnému.



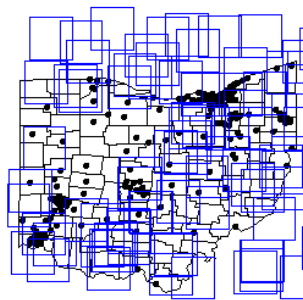
Obr. 5.4 Analýza kvadrátů – pravidelné rozmístění buněk

Postup analýzy spočívá v rozdělení studované plochy pravidelnou sítí na buňky a je zjištěn počet bodů v každé buňce. Následně je analyzováno rozdělení četností buněk s určitým počtem bodů. Toto rozdělení je porovnáváno s náhodným rozdělením četností. Buňky se označují jako **kvadráty** a nemusí jít o čtverce, ale např. i o kruhy či šestiúhelníky. Tvar buněk většinou vychází z empirie. V rámci jedné analýzy však tvar a velikost buněk musí být konstantní.

Extrémně shlukové uspořádání – většina bodů v jedné či několika málo buňkách

Extrémně pravidelné – ve všech buňkách přibližně stejně

Uvedenou metodu lze využít také tak, že se buňky stejné velikosti náhodně rozmístí po studované ploše.



Obr. 5.5 Analýza kvadrátů - náhodné rozmístění buněk

Citlivou stránkou metody je volba velikosti kvadrátů. **Optimální velikost kvadrátů (QS)** lze získat z následujícího vztahu:

$$QS = \frac{2 \cdot A}{n}$$

kde A je plocha studované oblasti a n počet analyzovaných bodů. Velikost strany vhodného kvadrátu je potom:

$$\sqrt{2A/n}$$

Získané rozložení četností bodů v kvadrátech (empirické) je porovnáváno s náhodným rozložením (teoretickým). Vhodným testem je např. K-S test. Testem můžeme kvantifikovat rozdíl empirického a teoretického (shlukové, pravidelné, náhodné) rozdělení bodů v ploše.

5.2.1 Praktický postup testování výsledků analýzy kvadrátů:

Formulujeme nulovou hypotézu - neexistuje statisticky významný rozdíl (je-li rozdíl malý, může být výsledkem náhody, čím je větší, s tím větší pravděpodobností náhodný není, ale je statisticky významný).

Zvolíme hladinu významnosti $\alpha = 0,05$

Vypočítáme kumulované četnosti

Vypočteme testovací kritérium:

$$D = \max |O_i - E_i|$$

Vypočteme kritickou hodnotu

$$D_\alpha = \frac{1,36}{\sqrt{m}}$$

kde m je počet kvadrátů. V případě porovnávání dvou výběrů o různém počtu členů m_1 a m_2 se kritická hodnota vypočte následovně:

$$D_\alpha = 1,36 \sqrt{\frac{m_1 + m_2}{m_1 \cdot m_2}}$$

Je-li vypočtená hodnota D větší než kritická hodnota D_α , potom rozdíl mezi oběma uspořádáními je statisticky významný.

Pozorované rozložení bodů můžeme také porovnávat s rozložením náhodně generovaným (např. podle určitého teoretického rozdělení). Často se využívá rozdělení Poissonovo (Poisson random process)

Poissonovo rozdělení je určeno především průměrnou frekvencí výskytu (λ) v jednotlivých jednotkách (kvadrátech), kde $\lambda = n/m$ při m kvadrátech a n bodech v prostoru. Je-li x počet bodů v kvadrátu, potom pravděpodobnost výskytu x bodů v kvadrátu podle Poissonova rozdělení je definována:

$$p(x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Z uvedeného vztahu můžeme pro různá x vypočítat pravděpodobnost rozložení bodů, které budou mít Poissonovo (náhodné) rozdělení.

Hodnoty pravděpodobnosti lze zjistit i zkráceným výpočtem. Je-li $x=0$, potom $p(0) = e^{-\lambda}$ a pravděpodobnosti pro následná x můžeme určit z $p(0)$, obecně:

$$p(x) = p(x-1) * \frac{\lambda}{x}$$

Je-li $x=1$, potom $p(x-1) = p(0)$ atd.

Vedle K-S testu můžeme k hodnocení rozdělení bodů v kvadrátech použít také vlastnosti Poissonova rozdělení – především hodnoty průměru a rozptylu Poissonova rozdělení, pro které platí, že se rovnají hodnotě (λ). Jinými slovy bude-li distribuce bodů v prostoru

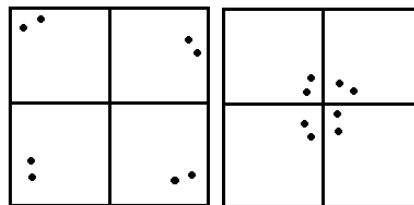
generována náhodným procesem, potom toto rozdělení má stejný průměr a rozptyl. Tedy jejich poměr se bude blížit jedné.

Postup: Vypočteme hodnoty průměru a rozptylu pro četnosti bodů v kvadrátech a hodnoty dáme do poměru. Hodnotu porovnáme s 1. Rozdíl lze dále standardizovat (vyjádřit v násobcích směrodatné odchylky). Vyjde-li hodnota větší než 1,96, potom je rozdíl statisticky významný na hladině $\alpha = 0,05$.

Test založený na poměru průměru a rozptylu je silnější než K-S test, lze ho však použít pouze v případě, že předpokládáme Poissonovo rozdělení studované množiny bodů.

Pozorované rozdělení bodů lze porovnávat i vůči jiným teoretickým rozdělením (např. negativní gamma či negativní binomické).

Omezení analýzy kvadrátů:



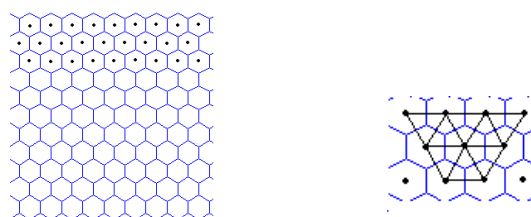
Obr. 5.6 Analýza kvadrátů neřeší otázku rozložení bodů uvnitř kvadrátů

5.3 Analýza nejbližšího souseda (NEAREST NEIGHBOUR ANALYSIS)

Metoda analýzy kvadrátů je založena na konceptu **hustoty** (počet bodů v ploše). Metoda analýzy nejbližšího souseda je naopak založena na konceptu **vzdálenosti** (spacing – plocha připadající na bod).

Metoda analýzy nejbližšího souseda je založena na porovnání pozorované průměrné vzdálenosti mezi nejbližšími sousedy a této průměrné vzdálenosti u známého vzorku (pattern).

Pozorovaná průměrná vzdálenost mezi nejbližšími sousedy může být větší či menší než vzdálenost při náhodném rozmístění bodů.



Obr. 5.7 Analýza nejbližšího souseda – pravidelné uspořádání bodů

Homogenní oblast – nejvíce uniformní vzorek – body v ploše tvoří středy pravidelných šestiúhelníků. Body tvoří trojúhelníkovou mřížku.

Za této konfigurace bude vzdálenost mezi body rovna výrazu

$$1,075\sqrt{A/n}$$

kde A je plocha a n počet bodů v ploše. V reálné situaci tvoří geografické rozložení bodů výjimečně pravidelný vzorek. K testování, zda má určité rozložení bodů v ploše jistý vzorek lze využít **R statistiku (R - randomness)**. Určí se jako poměr mezi pozorovanou a očekávanou průměrnou vzdáleností nejbližších sousedů v určité oblasti:

$$R = \frac{r_{obs}}{r_{exp}}$$

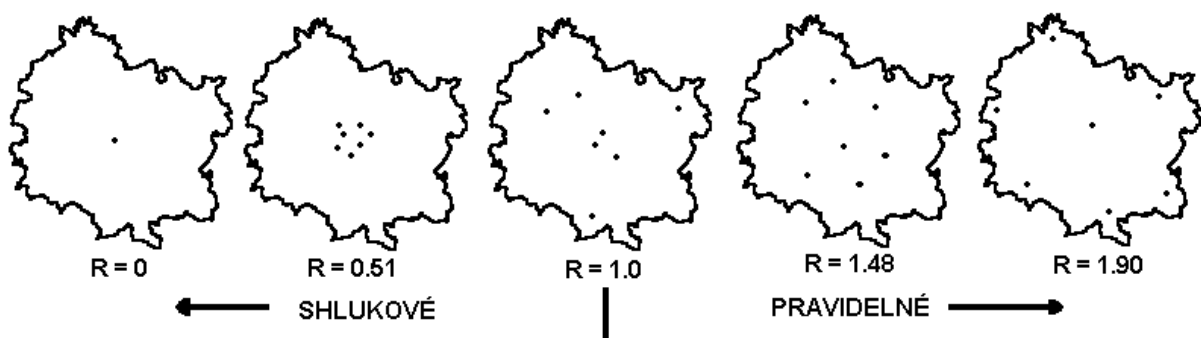
Hodnotu r_{obs} zjistíme tak, že určíme vzdálenost mezi daným bodem a všemi jeho sousedy. Dále najdeme nejkratší vzdálenost – tedy nejbližšího souseda. Tento proces se opakuje pro všechny body. Ze všech nejkratších vzdáleností se vypočte průměr.

Pro teoretické – náhodné - rozložení se průměrná vzdálenost nejbližšího souseda vypočte podle vzorce:

$$r_{exp} = \frac{1}{2\sqrt{n/A}}$$

Čím je hodnota $R < 1$, tím více se prostorové rozložení bodů blíží rozložení shlukovému ($r_{obs} < r_{exp}$).

Čím je hodnota $R > 1$, tím více se prostorové rozložení bodů blíží rozložení pravidelnému ($r_{obs} > r_{exp}$).



Obr. 5.8 Škála hodnot R statistiky

- $R = 0$ zcela shlukové
- $R = 1$ náhodné
- $R = 2,149$ zcela pravidelné

Je-li $R=0$, vzdálenosti jsou 0, všechny body mají stejnou polohu.

Jinou z možností, jak porovnat rozdíl mezi pozorovanou a očekávanou vzdáleností nejbližšího souseda je porovnat tuto diferenci s tzv. **směrodatnou chybou** (Standard Error – SE_r)

Směrodatná chyba popisuje pravděpodobnost, že jakýkoliv rozdíl dvou hodnot je výsledkem náhodných vlivů. Je-li tedy zjištěná diference malá ve srovnání s SE , potom rozdíl není statisticky významný a naopak.

Použití směrodatné chyby SE vychází z vlastností normálního rozdělení, pro které platí následující:

Je-li mezi pozorovanými populacemi rozdíl a jeho velikost náleží do intervalu $(-1SE_r; +1SE_r)$, potom existuje 68 % šance, že tento rozdíl je náhodný – tedy nevýznamný:

$$\text{Pravděpodobnost (<68\%)} = (-1SE_r; +1SE_r)$$

Za statisticky významný považujeme rozdíl, který můžeme obdržet v 5 případech ze sta – tedy s pravděpodobností 5 %, $\alpha=0,05$. Vyjádřeno v násobcích směrodatné chyby - rozdíl mezi dvěma populacemi považujeme za statisticky významný, jestliže je menší než $-1,96SE_r$ a nebo větší než $+1,96SE_r$:

$$\text{Pravděpodobnost (<95\%)} = (-1,96SE_r; +1,96SE_r)$$

Výpočet směrodatné chyby pro pozorované vzdálenosti bodů:

$$SE_r = \frac{0,26136}{\sqrt{n^2/A}}$$

Pomocí směrodatné chyby lze vypočítat standardizovanou hodnotu (Z-score):

$$Z_R = \frac{r_{obs} - r_{exp}}{SE_r}$$

Je-li tedy $Z_R < -1,96$ či $Z_R > 1,96$ potom vypočtený rozdíl mezi pozorovaným a náhodným uspořádáním je statisticky významný – tedy není náhodný a naopak.

Nelze spoléhat na vizuální srovnání prostorového rozložení ani na vypočtenou hodnotu R . Ta by měla být doplněna hodnotou Z_R pro ověření statistické významnosti pozorovaného rozdílu.

Metoda analýzy nejbližšího souseda může být rozšířena na analýzu nejbližších sousedů druhého, třetího a vyšších řádů. Například u obr. 2.6 dokumentujícího nevýhody kvadrantové analýzy by až analýza nejbližšího souseda druhého řádu odhalila, že se obě uspořádání výrazně liší. Na obrázku vlevo je R -statistika druhého řádu velká, na obrázku vpravo naopak malá.

Použití analýzy nejbližšího souseda rozdílných řádů může odhalit heterogenity v uspořádání bodů na rozdílných prostorových úrovních.

Problémy spojené s metodou analýzy nejbližšího souseda: výsledky jsou vysoce citlivé k měřítku (lokální vs. regionální) a vymezení zpracovávané oblasti. V závislosti na studovaném jevu by měla být věnována pozornost také vymezení studované plochy (administrativní či přirozené hranice).

5.4 Prostorová autokorelace (SPATIAL AUTOCORRELATION)

Jak analýza kvadrátů, tak analýza vzdálenosti nejbližšího souseda pracují pouze s polohou bodů. Nerozlišují body podle hodnot jejich atributů.

Oba parametry (polohu i atributy) hodnotí **prostorová autokorelace** – je tedy metodou vhodnější.

Východiska prostorové autokorelace: Většina jevů se v prostoru mění spojitě. Blízké body budou mít i podobné hodnoty studovaného jevu a naopak. (First law of geography - Tobler, 1970)

Koeficient prostorové autokorelace – uvažuje polohu bodů (vzájemnou vzdálenost) a hodnotí rozdílnost hodnot atributů bodů v prostoru. Mezi nejpoužívanější koeficienty prostorové autokorelace náleží **Gearyho poměr C (Geary's Ratio)** a **Moranův index I (Moran's I)**. Lze jich využít pro intervalová a poměrová data.

Dále používaná notace:

- c_{ij} – podobnost atributu v bodě i a j
- w_{ij} – vzdálenost bodu i a j . $w_{ii} = 0$ pro všechny body
- x_i – hodnota studovaného atributu v bodě i
- n – počet bodů ve vyšetřovaném vzorku

Obě míry prostorové autokorelace kombinují v jednom výrazu **míry podobnosti atributů** i **míry podobnosti polohy** – tento výraz je potom východiskem pro definování dalších vztahů:

$$\sum_{i=1}^n \sum_{j=1}^n c_{ij} w_{ij}$$

Koeficient prostorové autokorelace SAC (spatial autocorrelation coefficient) je úměrný vážené míře podobnosti atributů bodů – obecně:

$$SAC \approx \frac{\sum_{i=1}^n \sum_{j=1}^n c_{ij} \cdot w_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}$$

V případě Gearyho poměru se podobnost hodnot atributu mezi dvěma body vypočte podle následujícího vztahu:

$$c_{ij} = (x_i - x_j)^2$$

Gearyho poměr C se tedy vyjádří jako:

$$C = \frac{\sum_{i=1}^n \sum_{j=1}^n c_{ij} \cdot w_{ij}}{2 \sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot \sigma^2} = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot (x_i - x_j)^2}{2 \sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot \sigma^2}$$

kde σ^2 je rozptyl hodnot atributu x s průměrem \bar{x}

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}$$

V případě hodnoty **Moranova indexu I** se podobnost hodnot atributu v bodech i a j vyjádří následovně:

$$c_{ij} = (x_i - \bar{x}) \cdot (x_j - \bar{x})$$

Moranův index I je potom určen:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n c_{ij} \cdot w_{ij}}{s^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}} = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \cdot (x_i - \bar{x}) \cdot (x_j - \bar{x})}{s^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}}$$

kde s^2 je v tomto případě výběrový rozptyl:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Ve výše uvedených vzorcích lze všechny neznámé přímo určit z hodnot atributů bodů. Jedinou doposud nedefinovanou neznámou zůstává míra podobnosti (blízkosti) polohy bodů i a j , tedy hodnota w_{ij} .

Ta se běžně uvažuje jako inverzní hodnota vzdálenosti těchto bodů. Tedy podle výše uvedených předpokladů dáváme malou váhu hodně vzdáleným bodům a velkou váhu hodně vzdáleným bodům, tedy:

$$5.4.1.1 \quad w_{ij} = 1/d_{ij}$$

Rozdíly mezi oběma indexy jsou dány způsobem výpočtu rozdílů mezi hodnotami atributu. Obor hodnot, kterých mohou oba indexy nabývat se tedy také liší, jak uvádí následující tabulka:

Prostorové uspořádání	Gearyho poměr C	Moranův index I
-----------------------	-----------------	-----------------

Shlukové uspořádání, sousední body vykazují podobné hodnoty	$0 < C < 1$	$I > E(I)$
Náhodné uspořádání, body nevykazují znaky podobnosti	$C \sim 1$	$I \cong E(I)$
Pravidelné uspořádání, sousední body vykazují rozdílné charakteristiky	$1 < C < 2$	$I < E(I)$

kde $E(I) = (-1)/(n-1)$

5.4.2 Předpoklad náhodnosti a předpoklad normality

Při studiu prostorového uspořádání, můžeme předpokládat dva základní způsoby, kterými jsou atributy přiřazeny jednotlivým bodům.

Předpoklad náhodnosti (randomization, nonfree sampling) – předpokládáme, že hodnoty atributů v bodech představují pouze jednu z možných variant uspořádání při použití stejné množiny hodnot.

Alternativně můžeme předpokládat, že hodnoty atributů v množině studovaných bodů jsou pouze jednou z nekonečného množství možností. Každá hodnota je nezávislá na hodnotách jiných v množině bodů – **předpoklad normality (normality, free sampling)**. Předpoklad normality dovoluje nahrazení hodnot pozorování na rozdíl od předpokladu náhodnosti.

5.4.3 Určení odhadů očekávaných hodnot

Výše uvedené předpoklady náhodnosti (R) a normality (N) ovlivňují způsob výpočtu očekávaných (e – expected) hodnot i hodnot rozptylu. Očekávané hodnoty indexů a hodnoty rozptylů potřebujeme pro testování, zda se vypočtené hodnoty indexů C a I statisticky významně liší od náhodného uspořádání.

Odhad očekávaných hodnot pro náhodné uspořádání (random pattern) a rozptyly pro Gearyho poměr C:

$$E_N(C) = 1$$

$$E_R(C) = 1$$

$$VAR_N(C) = \frac{[(2S_1 + S_2)(n-1) - 4W^2]}{2(n+1)W^2}$$

$$VAR_R(C) = \frac{(n-1)S_1[n^2 - 3n + 3 - (n-1)k]}{n(n-2)(n-3)W^2} - \frac{(n-1)S_2[n^2 + 3n - 6 - (n^2 - n + 2)k]}{4n(n-2)(n-3)W^2} + \frac{W^2[n^2 - 3 - (n-1)^2k]}{n(n-2)(n-3)W^2}$$

kde

$$W = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$$

$$S_1 = \frac{\sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2}{2}$$

$$S_2 = \sum_{i=1}^n (w_{i.} + w_{.i})^2$$

$$k = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{\left(\sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

Očekávané hodnoty Moranova indexu I a hodnoty rozptylu se pro náhodné uspořádání vypočtou obdobně:

$$E_N(I) = E_R(I) = \frac{-1}{n-1}$$

$$VAR_N(I) = \frac{(n^2 S_1 - n S_2 + 3W^2)}{W^2(n^2 - 1)} - [E_N(I)]^2$$

$$VAR_R(I) = \frac{n[(n^2 - 3n + 3)S_1 - nS_2 + 3W^2]}{(n-1)(n-2)(n-3)W^2} - \frac{k[(n^2 - n)S_1 - nS_2 + 3W^2]}{(n-1)(n-2)(n-3)W^2} - [E_R(I)]^2$$

Máme-li vypočteny očekávané hodnoty indexů a jejich rozptyly, můžeme vyjádřit standardizované hodnoty (Z-skore)

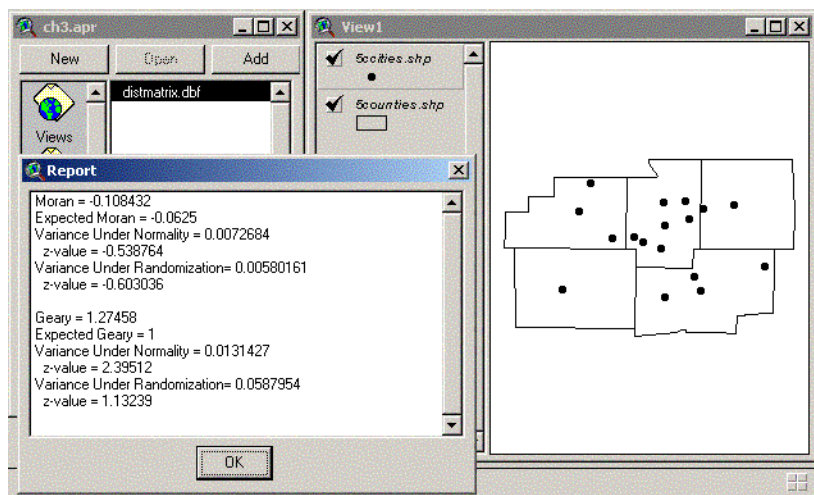
$$Z = \frac{I - E(I)}{VAR(I)}$$

nebo

$$Z = \frac{C - E(C)}{VAR(C)}$$

Pro hodnoty Z pak mohou být použity stejné kritické hodnoty, tedy na hladině významnosti $\alpha=0,05$:

$$-1,96 < Z < +1,96$$



Obr. 5.9 Příklad výpočtu měr prostorové autokorelace

Interpretace hodnot koeficientů prostorové autokorelace: Pokud zjištěné hodnoty z-skóre padnou vně intervalu $(-1,96 ; +1,96)$, potom se prostorové uspořádání bodů statisticky významně liší (na hladině 5 %) od uspořádání náhodného.

5.4.4 Alternativy výpočtu:

V uvedených vztazích lze modifikovat výrazy pro vyjádření podobnosti polohy. Například hodnoty w_{ij} mohou nabývat binárních hodnot 0, 1 podle toho, zda jde o body sousední či nikoliv (viz. např. teorie nodálních regionů, kde jako sousední body považujeme centroidy regionů, které obklopují daný region).

Modifikovat lze také váhy vzdálenosti bodů výrazem:

$$w_{ij} = \frac{1}{d_{ij}^b}$$

kde koeficient b může nabývat různých hodnot v závislosti na povaze studovaného problému (vzdálenost měřená dosažitelností autem a letadlem je jiná). Hodnota b je často rovna 2.

Uvedených koeficientů prostorové autokorelace lze využít pro výpočet podobnosti mezi polygony (viz. dále).

6. Statistická analýza liniových prvků

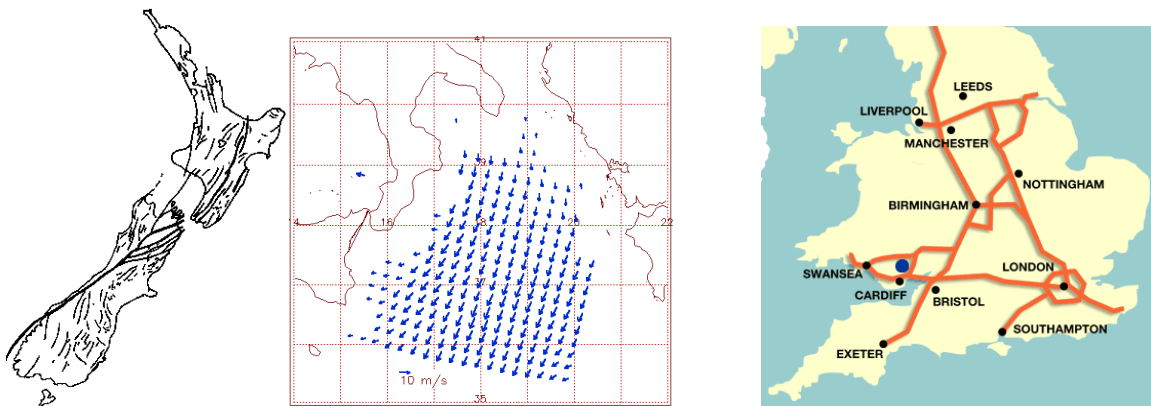
Linie mohou na mapách reprezentovat dva příbuzné objekty:

- **Vlastní linie** - reprezentují a lokalizují skutečně lineární geografické fenomény (řeky, silnice, potrubí)
- **Hrany** - rozdělují plochy a povrchy (hraniční linie, lomové linie). Hrany nemají šířku.

Problémy prezentace „přirozených linií“ v prostředí GIS jsou spojeny především s procesy generalizace a zjednodušení průběhu. Linie je prezentována jako spojnice posloupnosti lomových bodů, mezi lomovými body je rovná.

Problém měření vzdáleností - Někdy se místo měření vzdálenosti v délkových jednotkách používá cestovní čas a dopravní náklady.

Pro analýzu linií jsou vedle délky významné také atributy jako orientace, směr či spojení. Existence spojení mezi soustavou bodů, které tvoří linii, znamená, že lokace (body) na sobě nejsou nezávislé, ale jsou spojené v určitém směru. Body spojené v určitém pořadí musí zachovávat tuto posloupnost.



Obr. 6.1 Liniové prvky na digitální mapě - prosté linie, trajektorie, síť

Linie mohou v GIS vystupovat **na třech úrovních**, které představují jistou hierarchii (Obr. 6.1):

1. „**Prosté**“ linie – např. zlomy – lze určit jen délku a orientaci. Může existovat jako jednoduchá spojnice dvou bodů či jako „řetězec“
2. „**Trajektorie**“ – vektor pole větru – lze určit velikost (délku), orientaci a směr
3. **Sítě** - dopravní síť, říční síť – lze určit prostorové uspořádání – topologické vztahy, konektivitu, dostupnost, ...

Geometrické charakteristiky – linie může být prezentována jako:

- **Jednoduchá** spojnice – pouze dvou bodů (koncový a počáteční – délka je Euklidovská vzdálenost
- Posloupnost několika liniových segmentů - **řetězec**

Příklady analýzy prostorových vazeb liniových prvků:

- analýza převládající orientace, průměrné délky spoje,
- charakterizování liniových vzorků – „uspořádání sítí“
- dopravní dostupnost
- gravitační modely
- hledání optimální trasy

6.1 Prostorové atributy liniových prvků

Délka linie může být definována jako:

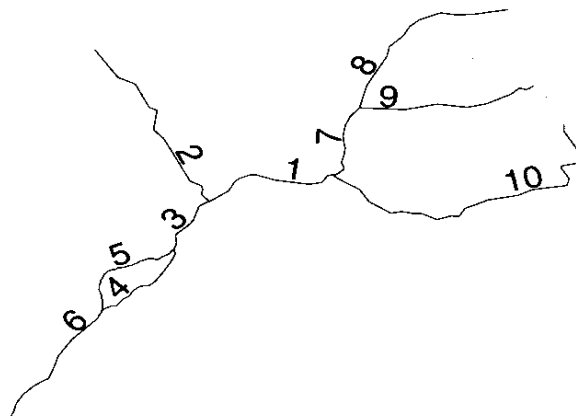
- přímá vzdálenost (vypočtená z Pythagorovy věty)
- „skutečná“ vzdálenost (součet přímých vzdáleností jednotlivých segmentů)

Orientace linie - orientace neurčuje směr (např. JV = SZ) – orientace zlomů, ulic. Nemá smysl otázka odkud - kam?

Směr linie - **typicky – vektor pole větru**

6.1.1 Topologie (sítě)

Výše uvedené atributy linií lze vyjádřit i pro jednotlivé segmenty sítě či pro celou síť jako celek (průměrná délka sítě, převládající orientace či směr segmentů sítě). Vedle toho jsou pro charakterizování sítě důležité atributy popisující jejich strukturu a uspořádání jako celek a dále popisují vztahy segmentů uvnitř sítě (topologii).



Obr. 6.2 Příklad sítě

Tabulka 6.1 Matice konektivity

ID	1	2	3	4	5	6	7	8	9	10
1	0	1	1	0	0	0	1	0	0	1
2	1	0	1	0	0	0	0	0	0	0
3	1	1	0	1	1	0	0	0	0	0
4	0	0	1	0	1	1	0	0	0	0
5	0	0	1	1	0	1	0	0	0	0
6	0	0	0	1	1	0	0	0	0	0

7	1	0	0	0	0	0	0	1	1	1
8	0	0	0	0	0	0	1	0	1	0
9	0	0	0	0	0	0	1	1	0	0
10	1	0	0	0	0	0	1	0	0	0

Základním topologickým aspektem sítě je způsob propojení jednotlivých segmentů – tedy její **konektivita**. Tradičním nástrojem používaným k charakterizování konektivity je **matice konektivity**. Je to matice čtvercová, binární, symetrická o n řádcích (sloupcích), kde n je počet segmentů sítě. Jednička v matici značí, že dva příslušné segmenty jsou bezprostředně spojeny. Na hlavní diagonále matice jsou nuly.

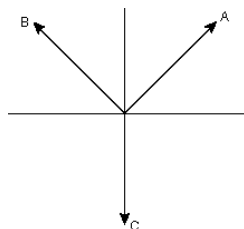
6.2 Směrová statistika (*Directional statistics*)

Topologii sítě lze charakterizovat jednoduchými mírami. Takovou je např. poměr mezi skutečnou délkou linie a spojnicí počátečního a koncového bodu. Tato charakteristika se určuje křivost linie (**sinuosity**). Čím větší číslo, tím větší křivost.

Směr linie – vizuální hodnocení směru linií lze provést přidáním šipek. Např. u pole větru je možné odhalit strukturu proudění v celé oblasti.

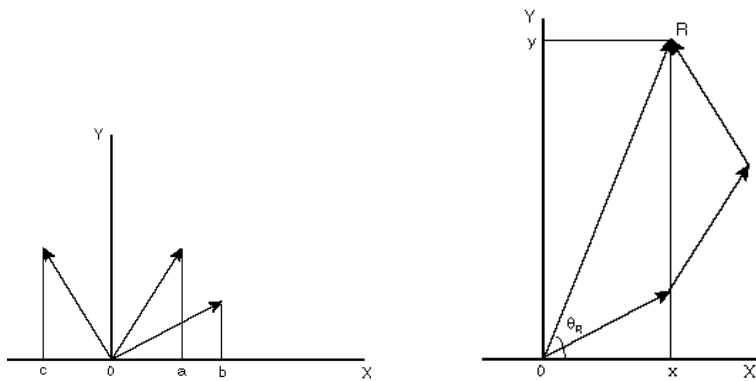
6.3 Směrový průměr (*directional mean*).

Využití klasických měr popisné statistiky pro charakterizování směru a orientace linií je nevhodné (viz. obr. 3.3).



Obr. 6.3 Problém popisné statistiky při určování charakteristik směru linie

Jak je patrné z obrázku, aritmetický průměr dvou vektorů s úhly 45 a 315 stupňů dává 180 (jižní směr), avšak měl by být 0 stupňů (severní směr). Průměrný směr je však nutné určit vektorovým součtem či tzv. **směrovým průměrem** (*directional mean*). Protože pracuje se směrem (úhlem) a ne s délkou, je možné ho prezentovat na základě jednotkových vektorů. Vektorovým součtem – přidáním počátku druhého vektoru na konec prvního dostaneme směrový průměr.



Obr. 6.4 Koncept směrového průměru

Směr výsledného vektoru lze získat také z následujícího vztahu:

$$\tan \theta_R = \frac{oy}{ox}$$

kde oy je suma délek vektorů ve směru osy y a ox suma délek vektorů ve směru osy x . Protože všechny vektory jsou jednotkové, délka ve směru osy y je v podstatě \sin úhlu a délka na ose x je \cos inus úhlu. Potom, jsou-li vektory označeny a, b, c a odpovídající úhly $\theta_a, \theta_b, \theta_c$, potom:

$$\tan \theta_R = \frac{\sin \theta_a + \sin \theta_b + \sin \theta_c}{\cos \theta_a + \cos \theta_b + \cos \theta_c}$$

Obecně, máme-li n vektorů v a úhel vektoru v od osy x je θ_v , výsledný vektor OR má úhel θ_R , měřený proti směru hodinových ručiček od osy x :

$$\tan \theta_R = \frac{\sum \sin \theta_v}{\sum \cos \theta_v}$$

což je tedy tangenta úhlu výsledného vektoru. Směrový průměr je potom \arctan z výše uvedeného výrazu.

Výsledná hodnota směrového průměru musí zohledňovat specifika jednotlivých kvadrantů, jak uvádí následující pravidla:

1. čítec i jmenovatel v $\tan \theta_R$ jsou oba kladné – není nutná žádná úprava (vektor leží v 1. kvadrantu)
2. čítec je kladný jmenovatel záporný – směrový průměr bude $180 - \theta_R$, (vektor leží v 2. kvadrantu)
3. čítec i jmenovatel v $\tan \theta_R$ jsou oba záporné – směrový průměr bude $180 + \theta_R$, (vektor leží v 3. kvadrantu)
4. čítec je záporný, jmenovatel kladný – směrový průměr bude $360 - \theta_R$, (vektor leží v 4. kvadrantu)

Praktický výpočet spočívá v určení \sin a \cos úhlů všech vektorů. Určí se jejich sumy a vytvoří poměr, který je tangentou výsledného úhlu. Směrový průměr je potom \arctan .

6.4 Směrový rozptyl (Circular variance)

Stejně jako v případě klasické popisné statistiky je charakterizování souboru prvků pouze měrou úrovně, kterou je výše uvedený směrový průměr, je často nedostatečné a může být i zavádějící. Např. pokud dva vektory budou svírat úhel 180 stupňů. Proto je nutné použít i měř variability (rozptylu).

Pokud dáme dohromady vektory podobného směru, výsledný vektor bude relativně dlouhý. Jeho délka se bude blížit n , pokud bude n jednotkových vektorů. Naproti tomu, pokud dáme dohromady vektory opačného či značně rozdílného směru, výsledný vektor bude významně menší než n . Tedy délku výsledného vektoru můžeme použít jako statistiku, která reflektuje variabilitu ve směru jednotlivých vektorů. Na základě výše uvedeného tedy platí:

$$OR = \sqrt{(\sum \sin \theta_v)^2 + (\sum \cos \theta_v)^2}$$

Směrový rozptyl (circular variance) S_v se potom vypočte:

$$S_v = 1 - OR/n$$

kde n je počet vektorů. S_v může nabývat hodnot 0 až 1. Je-li $S_v=0$, potom $OR=n$ a všechny vektory mají stejný směr. Je-li $S_v=1$, potom $OR=0$, všechny vektory mají opačný směr a výsledný vektor je bod.

6.5 Úvod do statistického popisu sítí

Nebude probírána síťová analýza – ta vyžaduje speciální prostředí a nástroje (maticový počet) i speciálně upravená vstupní data.

Základní pojmy používané v síťové analýze: **nódy** a **hrany** (spoje), jejich počet také charakterizuje síť. Ke křížení dvou a více hran dochází pouze ve vrcholu (planar graph topology)

Deskriptory sítě lze rozdělit do dvou skupin:

1. Deskriptory sítě jako celku
2. Deskriptory relací jednotlivých segmentů sítě.

6.6 Konektivita a matice konektivity

Matice konektivity (tab. 3.1) shrnuje informaci o tom, které segmenty sítě spolu souvisí (jsou bezprostředně spojeny). Lze však charakterizovat i úroveň konektivity sítě jako celku. Pro fixní počet vrcholů má síť s větším počtem spojů lepší konektivitu. Dále existuje minimální počet spojů, který zajišťuje spojení všech vrcholů.

Bude-li v – počet vrcholů sítě, e - počet hran sítě potom:

$$e_{\min} = v - 1$$

Minimálně propojená síť (Minimally conneted network - MCN) – odstraníme-li jakoukoliv jednu hranu, síť se rozpadne na dva subsystemy.

Podobně lze pro daný počet vrcholů vytvořit **maximální počet hran**, které spojují všechny vrcholy. Tedy maximální počet hran v síti o v vrcholech:

$$e_{\max} = 3(v - 2)$$

Jednoduchou charakteristikou konektivity sítě je **Gamma index** (γ) – je definován jako poměr aktuálního a maximálního počtu vrcholů sítě.

$$\gamma = \frac{e}{e_{\max}}$$

Další jednoduchou charakteristikou konektivity sítě je **počet okruhů**. Výskyt okruhů v síti značí možnost dostat se z jednoho místa do jiného alternativními cestami. Síť s minimální konektivitou nemá žádný okruh.

Počet okruhů lze zjistit tak, že od aktuálního počtu hran v síti odečteme počet hran potřebný pro minimálně propojenou síť (MCN), tedy $e - (v - 1)$ nebo $e - v + 1$.

Obdobně pro daný počet vrcholů je **maximální počet okruhů** roven $2v - 5$. S oběma uvedenými počty okruhů lze vytvořit poměr aktuálního počtu k počtu maximálnímu – tedy tzv. **alfa index**:

$$\alpha = \frac{e - v + 1}{2v - 5}$$

Pomocí alfa indexu můžeme snadno porovnat dvě sítě.

6.7 Dostupnost sítě (Accessibility)

Jedná se o charakteristiku jednotlivých vrcholů či hran sítě. Popisuje jejich dostupnost v rámci sítě. Další text se týká dostupnosti hran sítě, obdobné vztahy lze definovat i pro vrcholy.

Jednoduchým ukazatelem dostupnosti hrany v rámci sítě je, s kolika jinými hranami daná linie přímo souvisí. Tuto informaci lze vyčíst z binární matice konektivity, pokud tuto doplníme např. řádkovým součtem.

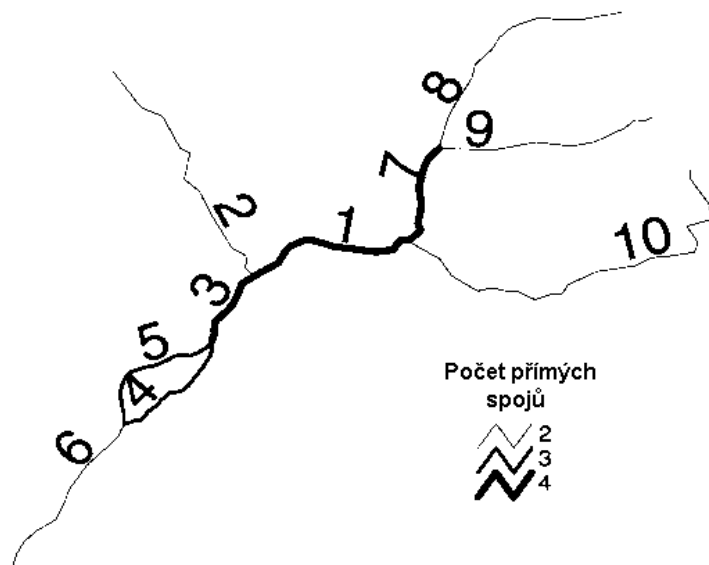
Tabulka 6.2 Matice konektivity a dostupnost hran v rámci sítě

ID	1	2	3	4	5	6	7	8	9	10	SUMA
1	0	1	1	0	0	0	1	0	0	1	4
2	1	0	1	0	0	0	0	0	0	0	2
3	1	1	0	1	1	0	0	0	0	0	4

4	0	0	1	0	1	1	0	0	0	0	3
5	0	0	1	1	0	1	0	0	0	0	3
6	0	0	0	1	1	0	0	0	0	0	2
7	1	0	0	0	0	0	0	1	1	1	4
8	0	0	0	0	0	0	1	0	1	0	2
9	0	0	0	0	0	0	1	1	0	0	2
10	1	0	0	0	0	0	1	0	0	0	2

Tabulka 6.3 Charakteristiky dostupnosti sítě (viz. obr 3.5, 3.6, 3.7)

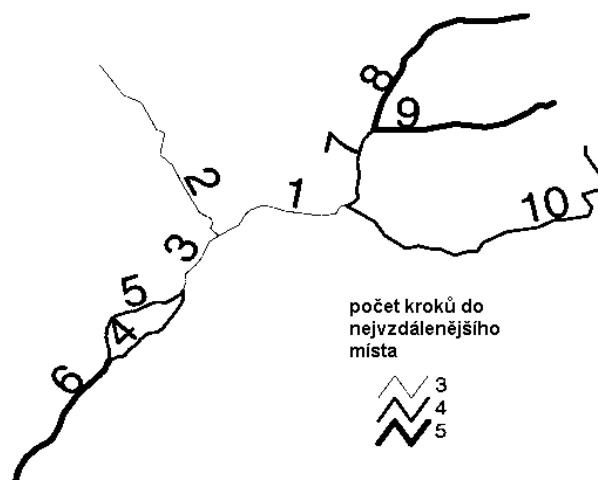
ID	počet přímých spojů	počet kroků k dosažení nejvzdálenějšího místa	celkový počet přímých a nepřímých spojů
1	4	3	15
2	2	3	19
3	4	3	16
4	3	4	21
5	3	4	21
6	2	5	28
7	4	4	18
8	2	5	25
9	2	5	25
10	2	4	20



Obr. 6.5 Dostupnost jednotlivých segmentů sítě charakterizovaná počtem přímých spojů

Uvedená charakteristika však může být zavádějící, protože nebere v úvahu relativní (topologickou) polohu hrany v rámci sítě. Hrana může mít i pouze jeden či dva spoje, přesto může být snadno dostupná, protože se nachází uprostřed sítě (a naopak).

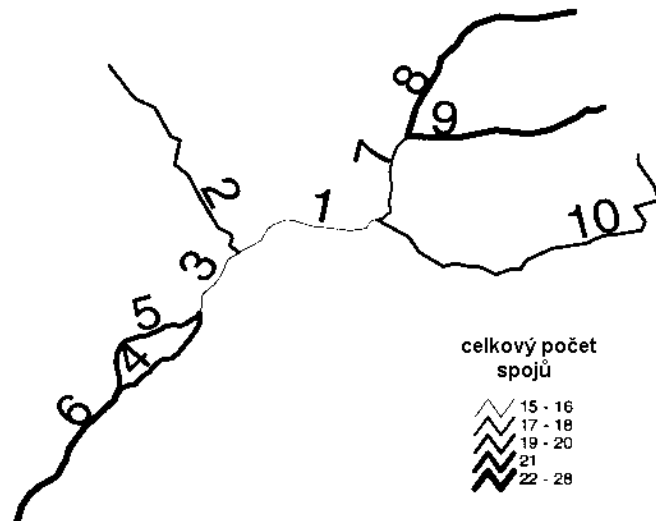
Relativní pozici každé hrany v rámci sítě lze zjistit např. pomocí počtu hran, kterými se lze z daného spoje dostat do nejvzdálenějšího místa sítě.



Obr. 6.6 Dostupnost jednotlivých segmentů sítě charakterizovaná počtem kroků nutných k dosažení nejvzdálenějšího místa sítě.

Diametr (poloměr) sítě – je to jedna (1) plus největší počet hran nutných k dosažení nejvzdálenějšího místa v síti.

Kvalitu spojení dvou hran (vrcholů) definuje počet hran mezi nimi. Spojení mohou být přímá a nepřímá. Tedy počet přímých a nepřímých spojů, které jsou třeba, aby byla daná hrana spojena se všemi hranami ostatními. Nepřímé spoje lze vážit počtem kroků. Zřejmě platí, že čím větší je celkový potřebný počet spojů, tím hůře dostupná je daná hrana. Celkový počet spojů (přímých i nepřímých) je mírou dostupnosti.



Obr. 6.7 Dostupnost jednotlivých segmentů sítě charakterizovaná počtem přímých a nepřímých spojů nutných k dosažení jakéhokoliv místa v síti

7. Prostorové uspořádání ploch

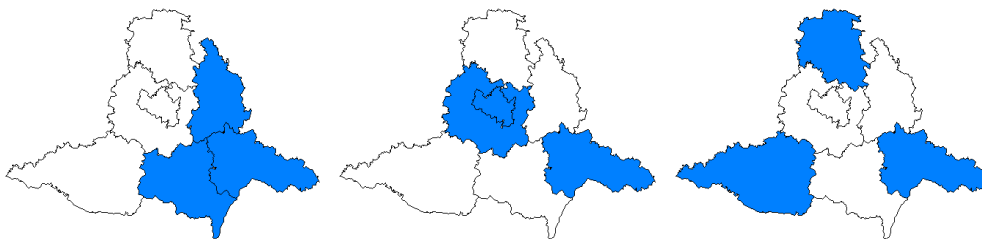
Využití prostorové statistiky k popisu měř úrovně a variability geografických jevů spojených s plochami (polygony) má v řadě geografických disciplín dlouhou tradici (demografie, krajinná ekologie apod.). Studium prostorových vztahů může být zaměřeno na následující typy úloh:

1. porovnání prostorového uspořádání studovaného jevu s **uspořádáním teoretickým** (shlukovým, pravidelným či náhodným)
2. **typologie** prostorového uspořádání jevů (bez územní souvislosti)
3. **regionalizace** - seskupování jednotek (polygonů) do vyšších územně souvisejících celků
4. **interpolace** a vyhlazování areálových dat

7.1 Míry prostorového uspořádání ploch

Prostorová autokorelace – hodnoty atributů ploch spolu korelují v závislosti na jejich vzájemné poloze. To je v důsledku podobných přirozených (přírodních) podmínek (např. produkce zemědělských podniků) či v důsledku přirozené spojitosti jevů.

U prostorově autokorelovaných dat nejsou hodnoty atributů v prostoru náhodné, ale prostorově závislé. Tato vazba (autokorelace) může být **pozitivní** (shlukové uspořádání - sousední objekty mají podobné hodnoty) či **negativní** (u pravidelného uspořádání). V případě náhodného uspořádání – slabá či žádná prostorová autokorelace. Také v případě prostorové autokorelace lze měřit její sílu.



Obr. 7.1 Příklad pozitivní prostorové autokorelace (shlukové uspořádání - vlevo) a negativní prostorové autokorelace (disperzní uspořádání – vpravo)

Prostorová autokorelace je významným ukazatelem k hodnocení dynamiky a časových změn v prostorovém uspořádání objektů a pro predikce.

Další význam prostorové autokorelace spočívá ve skutečnosti, že řada statistických ukazatelů (např. regresní modely) požaduje splnění předpokladu náhodnosti výběru objektů a jejich vzájemné nezávislosti. Míry prostorové autokorelace tak mohou potvrdit či vyvrátit splnění uvedených předpokladů.

7.2 Matice prostorových vah (*Spatial weights matrices*)

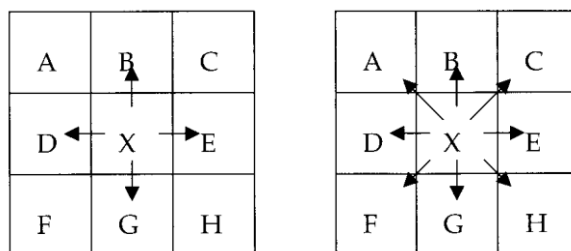
Prostorová autokorelace měří stupeň podobnosti atributů mezi danou plochou a plochami sousedními. Nejprve proto musí být **vztahy sousedství** jistým způsobem kvantifikovány.

Máme plochu s n prostorovými jednotkami. Potom můžeme definovat $n \times n$ párů sousedství – maticí typu $n \times n$. Každá prostorová jednotka je prezentována jedním řádkem a sloupcem. Každá hodnota v matici prezentuje prostorový vztah mezi jednotkami prezentovanými daným řádkem a sloupcem v matici. Buňky matice mohou nabývat různých hodnot v závislosti na způsobu definování sousedství (např. binární matice s 0 a 1 podle toho, zda jednotky spolu přímo sousedí či nikoliv, nebo – buňky nesou vzdálenost mezi centroidy obou jednotek. Protože hodnoty v buňkách představují váhy při výpočtu prostorové autokorelace, potom se sestavené matice označují jako matice prostorových vah).

7.2.1 Způsoby definování sousedství

Označují se podle pohybu šachových figur (Rook's case – věž, Queen's case – Dáma) – viz.

Obr. 7.2 Bezprostřední sousedé (se společnou hranicí, i jedním bodem v případě Queens case) jsou sousedé prvního řádu. Analogicky lze definovat sousedy vyšších řádů.



Obr. 7.2 Způsoby definování sousedství

Vedle **sousedství** je další běžně užívanou mírou prostorové relace objektů jejich **vzdálenost**. Intenzita vztahu dvou vzdálených jednotek bude obecně menší než intenzita vztahu jednotek blízkých. Tato vzdálenost může být arbitrárně určena (na základě zkušenosti či povahy studovaného problému: např. k danému domu jsou sousedé definováni jako domy do vzdálenosti 1 km, výsledek potom ze vyjádřit v binární podobě).

7.3 Binární matice konektivity (BCM – binary connectivity matrix)

Analogicky jako v případě linií – binární, čtvercová symetrická matice C s prvky c_{ij} , 1 – sousedí, 0 - ne)

<i>Id</i>	<i>Brno_venko</i>	<i>Blansko</i>	<i>Vyškov</i>	<i>Brno_město</i>	<i>Hodonín</i>	<i>Znojmo</i>	<i>Břeclav</i>
Brno-venkov	0.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000
Blansko	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000
Vyškov	1.0000	1.0000	0.0000	0.0000	1.0000	0.0000	1.0000
Brno-město	1.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Hodonín	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000
Znojmo	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000
Břeclav	1.0000	0.0000	1.0000	0.0000	1.0000	1.0000	0.0000

7.3.1 Binární matice sousedství

Vlastnosti BCM:

- Prvky na hlavní diagonále mají hodnoty 0
- Matice je symetrická – redundance uložené informace
- Suma v řádku nese informaci o počtu sousedů dané jednotky
- Pro větší počet prostorových jednotek obsahuje velké množství nul a je tedy paměťově náročná

Vhodnější způsob zaznamenání vztahů sousedství je uchovávání ID či názvu sousedů pro každou plochu, tedy např.:

Polygon	Soused1	Soused2	...		
Brno-město	Brno-venkov	Blansko			
Blansko	Brno-venkov	Vyškov	Brno-město		
.....			

7.3.2 Stochastická matice či matice se standardizovanými řádkovými vahami (RSWM)

Zaznamenání sousedství v binární podobě není v řadě případů výhodné – váhy jsou stejné bez ohledu na počet sousedů. Vhodnějším způsobem je nahrazení jedniček vahou w_{ij} , vypočtenou jako poměr mezi hodnotu c_{ij} a sumou v řádku – tj. počtem sousedů. Tedy má-li jednotka 4 sousedy, bude její váha rovna 0,25 – tak dostaneme z matice C matici W,

označovanou jako matici se standardizovanými řádkovými vahami. Stejně jako matice C má i W na hlavní diagonále nuly, není vak již symetrická.

<i>Id</i>	<i>Brno_venkov</i>	<i>Blansko</i>	<i>Vyškov</i>	<i>Brno_město</i>	<i>Hodonín</i>	<i>Znojmo</i>	<i>Břeclav</i>
Brno-venkov	0.0000	0.2000	0.2000	0.2000	0.0000	0.2000	0.2000
Blansko	0.3333	0.0000	0.3333	0.3333	0.0000	0.0000	0.0000
Vyškov	0.2500	0.2500	0.0000	0.0000	0.2500	0.0000	0.2500
Brno-město	0.5000	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000
Hodonín	0.0000	0.0000	0.5000	0.0000	0.0000	0.0000	0.5000
Znojmo	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5000
Břeclav	0.2500	0.0000	0.2500	0.0000	0.2500	0.2500	0.0000

Obr. 7.3 Matice se standardizovanými řádkovými vahami

7.4 Vzdálenosti centroidů

Vztahy prostorové závislosti lze charakterizovat také vzdáleností jednotek (viz. první zákon geografie – Tobler, 1970: Všechny objekty spolu souvisí, ale blízké objekty spolu souvisejí více). Tedy vzdálenost je vhodnou vahou pro definování prostorových vztahů.

Existuje několik způsobů definování vzdálenosti dvou polygonů, např. **vzdálenost centroidů**. Existuje několik způsobů určení centroidu pro daný polygon. V závislosti na tvaru polygonu nemusí jeho centroid ležet uvnitř něho.

Jsou-li jako váhy použity vzdálenosti (zde vzdálenosti centroidů), matice se označuje D s prvky d_{ij} . Váhy jsou potom definovány jako převrácená hodnota vzdálenosti:

$$w_{ij} = \frac{1}{d_{ij}}$$

V řadě případů síla vztahu mezi dvěma jednotkami klesá rychleji než se zvětšuje jejich vzdálenost, proto se váhy definují jako:

$$w_{ij} = \frac{1}{d_{ij}^2}$$

7.4.1 Nejbližší vzdálenosti

Na místo vzdáleností centroidů jsou použity vzdálenosti dvou nejbližších částí dvou polygonů. Takto definované váhy jsou výhodné pro charakterizování prostorových kontaktů či difuze. U takto sestavené matice buňky s nulami mimo hlavní diagonálu (sousedé) odpovídají buňkám s jedničkami v binární matici sousedství.

<i>Id</i>	<i>Brno_venko</i>	<i>Blansko</i>	<i>Vyškov</i>	<i>Brno_město</i>	<i>Hodonín</i>	<i>Znojmo</i>	<i>Břeclav</i>
Brno-venkov	0.0000	0.0000	0.0000	0.0000	6.3679	0.0000	0.0000
Blansko	0.0000	0.0000	0.0000	0.0000	23.0282	29.5297	24.4276
Vyškov	0.0000	0.0000	0.0000	3.7893	0.0000	23.7376	0.0000
Brno-město	0.0000	0.0000	3.7893	0.0000	15.7463	14.2933	8.6112
Hodonín	6.3679	23.0282	0.0000	15.7463	0.0000	30.5051	0.0000
Znojmo	0.0000	29.5297	23.7376	14.2933	30.5051	0.0000	0.0000
Břeclav	0.0000	24.4276	0.0000	8.6112	0.0000	0.0000	0.0000

Obr. 7.4 Matice vzdáleností mezi nejbližšími částmi polygonů

7.5 Míry prostorové autokorelace

Výše uvedené matice slouží k definování měř prostorové autokorelace (SA). Míry SA mohou být vztaheny k poli bodů (viz. výše) či ploch. V případě ploch lze zpracovávat **data nominální** (JCS - joint count statistics – Statistika charakteru sousedství), **intervalová i poměrová** (Moranův index I, Gearyho poměr C, G-statistika)

Uvedené míry lze označit jako **globální** míry prostorové autokorelace (asociace). Tedy jedna hodnota je vypočtena pro celou studovanou oblast. Avšak také prostorová autokorelace se může měnit v rámci studované oblasti – k deskripci prostorové heterogenity prostorové autokorelace lze využít **lokálních měř** – Local Indicator of Saptial Association (LISA) a lokální verze G-statistiky (local G-statistics).

Ke grafickým prostředkům hodnotícím prostorovou autokorelaci patří **Moranův scatterplot diagram**.

Základní notace používaná v následujícím popisu indexů prostorové autokorelace

w_{ij} – obecně buňka matice vah W pro řádek i a sloupec j . (nejen matice stochastické – viz. výše)

Sumace vah daného řádku i přes všechny sloupce (řádková suma):

$$w_{i.} = \sum_j w_{ij}$$

Sumace vah daného sloupce j přes všechny řádky (sloupcová suma):

$$w_{.j} = \sum_i w_{ij}$$

Sumace všech buněk matice vah:

$$W = \sum_i \sum_j w_{ij}$$

Pro testování významnosti indexů prostorové autokorelace lze váhy v jednotlivých výrazech sumarizovat do následujících výrazů:

$$SUM_1 = \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2$$

a

$$SUM_2 = \sum_i \left(\sum_j w_{ij} + \sum_j w_{ji} \right)^2$$

SUM_1 – suma přes váhy. Jsou-li váhy binární a matice symetrická, potom $(w_{ij} + w_{ji})^2 = 4$

SUM_1 je tedy čtyřnásobek celkového počtu spojů (společných hranic) v celé studované ploše.

Hodnota SUM_2 je založena na sumování vah každé plošné jednotky v obou směrech (w_{ij} i w_{ji}). Výsledná hodnota je potom získána jejich součtem, umocněním a sumací pro všechny jednotky studované oblasti.

Nechť n je počet plošných jednotek ve studované oblasti. Existují-li dvě skupiny jednotek definovaných atributy s hodnotami x a y , potom výrazy n_x a n_y značí počet jednotek v jednotlivých skupinách.

Podobně:

$$n^{(x)} = n * (n - 1) * (n - 2) * (n - 3) * \dots * (n - x + 1)$$

kde $n > x$

Například, bude-li $n=5$, potom $n^{(3)} = n(n-1)(n-2) = 5 \times 4 \times 3$ a $n^{(1)} = n$

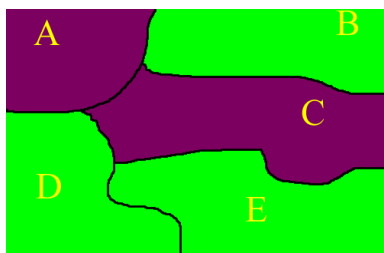
Jestliže x_i je hodnota atributu pro plochu i , můžeme definovat nový parametr m_j , založený na hodnotách x_i :

$$m_j = \sum_{i=1} x_i^j$$

kde $j = 1, 2, 3, 4$. Potom, jestliže $j=1$, m_j je suma x_i pro všechna i . Jestliže $j=2$, m_j bude suma všech čtverců x_i .

7.5.1 Statistika charakteru sousedství - Joint count statistics (JCS)

Touto metodou lze zjistit, zda uspořádání ploch, které mohou nabývat **binárních** hodnot vykazuje prvky náhodnosti. Tedy zda existuje pozitivní (clustered pattern) či negativní (random pattern) prostorová autokorelace.

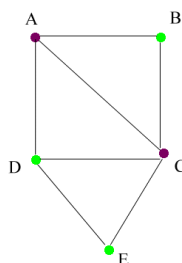


Obr. 7.5 Statistika četnosti spojů (JCS)

Podstata metody – jednoduchý příklad:

Máme mapu se dvěma kategoriemi landuse: U – zástavba, R – volná krajina. Potom mohou existovat čtyři typy sousedských vztahů: UU, RR, UR, RU. V případě čistě náhodného uspořádání se bude každá kombinace vyskytovat v 25% případů. Dvojice ploch s odlišným atributem se budou vyskytovat v 50 % případů. Pokud $UR + RU < 50\%$, potom výskyt dvojic ploch se stejným atributem UU a RR bude vyšší než 50% - což je případ pozitivní prostorové autokorelace. V případě 50 na 50 – uspořádání je náhodné a pokud $UR + RU > 50\%$, pak se jedná o negativní SA, kdy dominují hranice nepodobných ploch.

Mapu (obr. 1) s pěti plochami můžeme prezentovat také grafem s vrcholy a spoji, zaznamenávajícími druh povrchu a také bezprostřední sousedství jednotlivých ploch s plochami jinými, jak je patrné z obr. 4.4



Obr. 7.6 Grafická prezentace druhů spojů

Sestavíme matici sousedství pro jednotlivé plochy. V této matici nula značí, že obě plochy spolu bezprostředně nesousedí, 1 naopak. Zároveň je barvou buňky v matici naznačeno, o jaký typ spoje se jedná (Obr. 7.7).

	A	B	C	D	E
A	0	1	1	1	0
B	1	0	1	0	0
C	1	1	0	1	1
D	1	0	1	0	1
E	0	0	1	1	0

Obr. 7.7 Binární matice sousedství pro nominální data

Pořadí řádků a sloupců v uvedené matici je určeno abecedním pořadím identifikátorů ploch. Nic nebrání sestavit matici v jiném pořadí řádků a sloupců – například podle typu povrchu – (viz Obr. 7.8).

	A	C	B	D	E
A	0	1	1	1	0
C	1	0	1	1	1
B	1	1	0	0	0
D	1	1	0	0	1
E	0	1	0	1	0

Obr. 7.8 Binární matice sousedství uspořádaná podle hodnot atributů

Obě matice jsou symetrické, ve druhém případě navíc je možné jednoduše popsat prostorovou autokorelaci pomocí čtyř sub-matic. Z matice lze zjistit, že 14 buněk obsahuje jedničku, která značí výskyt hrany (14 párů sousedství). Dále platí, že jednotlivé typy sousedství se na mapě vyskytují s těmito četnostmi:

- UU=2
- UR=5
- RU=5
- RR=2

Z toho plyne, že $RU + UR > 14/2$, tedy naše mapa vykazuje negativní autokorelaci, nepodobné plochy (s odlišným typem povrchu) se shlukují.

Uvedený koncept lze dále rozšířit využitím počtu pravděpodobnosti a statistických testů. Ty nám umožní testovat statistickou významnost prostorového uspořádání ploch v mapě. V dalším výkladu jsou používány dvě hodnoty atributů B – black, černá, W – white, bílá. Tedy bude-li prostorové uspořádání indikovat uspořádání do shluků, potom můžeme předpokládat více hranic typu BB či WW než BW nebo WB – tedy pozitivní prostorovou autokorelaci.

JCS tedy nejprve určuje počet jednotlivých druhů spojů s cílem testovat četnost jejich výskytu. Pro plochu s malým počtem polygonů lze počty jednotlivých spojů zjistit manuálně, pro velký počet ploch je nutné využití metod matematické statistiky. Obecné kroky výpočtu jsou následující:

Nechť $x_i=1$ jestliže polygon i je černý a $x_i=0$ jestliže polygon i je bílý.

Potom pro BB spoje bude: $O_{BB} = \frac{1}{2} \sum_i \sum_j (w_{ij} x_i x_j)$

Pro WW spoje bude platit: $O_{WW} = \frac{1}{2} \sum_i \sum_j [w_{ij} (1 - x_i)(1 - x_j)]$

Pro BW nebo WB spoje bude platit: $O_{BW} = \frac{1}{2} \sum \sum [w_{ij}(x_i - x_j)^2]$

Uvedené vzorce představují výrazy pro pozorované (O – observed) počty spojů popisující dané **uspořádání**.

Vysoké hodnoty O_{BB} či O_{WW} či obou indikují pozitivní prostorovou autokorelaci (slukování). Pozorované počty spojů však musíme porovnat s náhodným uspořádáním a musíme testovat, zda eventuelní zvýšené počty O_{BB} či O_{WW} nejsou výsledkem pouhé náhody, zda jsou či nejsou statisticky významné. Budeme tedy pracovat s počtem pravděpodobnosti.

Způsob určení pravděpodobnosti výskytu B a W polygonů však může významně ovlivnit výsledek analýzy. Hodnoty atributů mohou být jednotlivým polygonům přiřazeny na základě předpokladu normality či náhodnosti (viz. prostorová analýza bodů)

Předpoklad normality: (NORMALITY - FREE - SAMPLING) – pravděpodobnost, že se jedná o polygon B či W je založena na teorii či na trendu hodnot atributů odvozeném z větší oblasti. Pravděpodobnost, že polygon má B či W není ovlivněna celkovým počtem B či W polygonů v oblasti.

Předpoklad náhodnosti: (RANDOMIZATION – NONFREE – SAMPLING) – pravděpodobnost, že polygon bude mít B či W je omezena či závisí na celkovém počtu B či W polygonů.

Příklad: Plocha obsahující sedm polygonů:

Předpoklad náhodnosti – může existovat různá konfigurace 4 „černých“ a 3 „bílých“ ploch.

Předpoklad normality - může existovat různá konfigurace jakéhokoliv (0 až 7) počtu „černých“ a „bílých“ ploch.

U metody JCS bychom neměli pracovat s předpokladem normality v případě, že informace získané z teorie, zkušenosti či z trendové funkce z širšího okolí jsou nespolehlivé. Náhodné vzorkování totiž vyžaduje méně rigorózní podmínky použití.

7.6 Normální vzorkování

V obou výše komentovaných případech je nutné vedle pozorovaných (O) počtů jednotlivých typů spojů či hranic (joint) zjistit počty očekávané (E) a také jejich směrodatné odchylky. Očekávané počty odrážejí efekt náhodnosti či nevýznamné prostorové autokorelace jakéhokoliv typu (pozitivní či negativní). Tedy zjistí se difference mezi pozorovanými a očekávanými četnostmi spojů. Tyto difference jsou následně standardizovány hodnotami příslušných směrodatných odchylek a získáme tak standardizovaná skóre. Z hodnot těchto skóre můžeme rozhodnout, zda je ve studované oblasti významná pozitivní či negativní prostorová autokorelace v uspořádání polygonů podle hodnot atributu. Jinými slovy, je nutné provést tři typy porovnání. Dále je prezentován případ pouze pro testování negativní prostorové autokorelace.

Pro případ normálního vzorkování jsou vztahy pro očekávané četnosti jednotlivých druhů spojů (joint) (E_{BB} , E_{WW} , E_{BW}) následující:

$$E_{BB} = \frac{1}{2}Wp^2 \quad E_{ww} = \frac{1}{2}Wq^2 \quad E_{WB} = Wpq$$

p – pravděpodobnost, že plocha bude B (černá)

q – pravděpodobnost, že plocha bude W (bílá)

Pravděpodobnosti p , q musí dávat 100% nebo ($p + q = 1$). Pokud není k dispozici jiná informace, potom $p = n_B/n$, jsou však i jiné způsoby určení p . Pokud je použita prostorová matice vah binární, lze výrazy pro očekávané počty typů spojů zjednodušit:

$$E_{BB} = Jp^2 \quad E_{ww} = Jq^2 \quad E_{BW} = 2Jpq$$

kde J značí celkový počet spojů ve studované oblasti.

K testování statistické významnosti zjištěného prostorového uspořádání lze využít Z-testu.

K němu je zapotřebí zjistit směrodatné odchylky očekávaných počtů spojů. Směrodatné odchylky se vypočtou v závislosti na použité váhové matici následovně:

Pro stochastickou matici vah:

$$\sigma_{BB} = \sqrt{\frac{1}{4}p^2q[S_1q + S_2p]}$$

$$\sigma_{ww} = \sqrt{\frac{1}{4}q^2p[S_1p + S_2q]}$$

$$\sigma_{BW} = \sqrt{\frac{1}{4}\{4S_1pq + S_2pq * [1 - 4pq]\}}$$

Pro binární matici vah:

$$\sigma_{BB} = \sqrt{p^2J + p^3K - p^4(J + K)}$$

$$\sigma_{ww} = \sqrt{q^2J + q^3K - q^4(J + K)}$$

$$\sigma_{BW} = \sqrt{2pqJ + pqK - 4p^2q^2(J + K)}$$

kde σ je směrodatná odchylka počtu příslušných spojů

S_1, S_2, J, p, q byly definovány výše

$$K = \sum_{i=1}^n L_i(L_i - 1)$$

Hodnota n v tomto výrazu značí celkový počet polygonů a L_i je počet spojů mezi polygonem i a jeho sousedy.

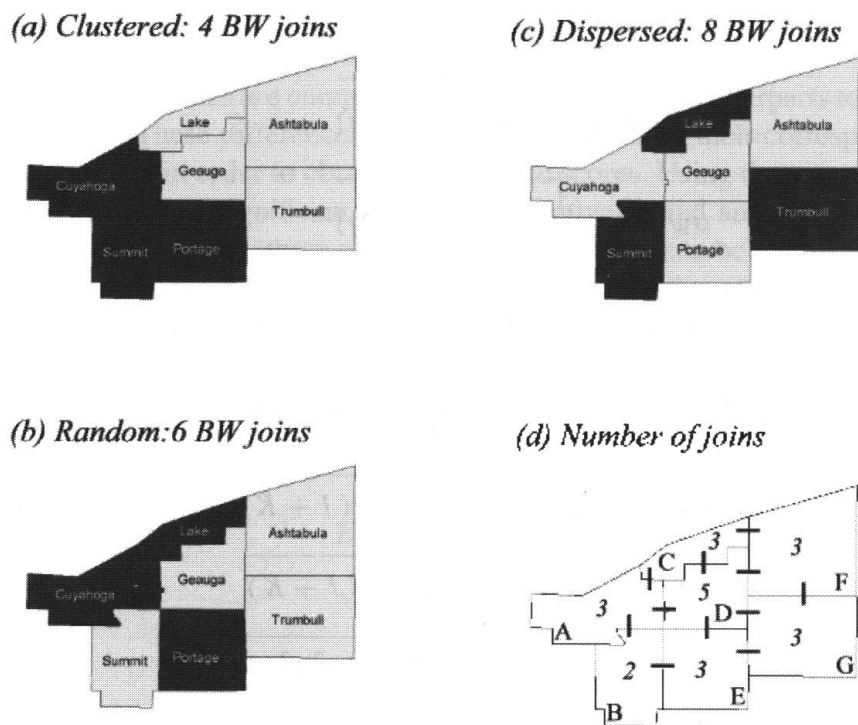
Obecný postup testování (na příkladu negativní prostorové autokorelace (BW spoje) při použití binární matice):

Pro výpočet očekávaných potřebujeme znát hodnoty pravděpodobností p , q . Rozhodneme se pro určité pravidlo definující sousedství (rook, queen). Dále určíme J (počet spojů) – zjistíme sumováním všech členů binární matice vah a dělíme dvěma. Odhad správných hodnot p a q – ze zkušenosti, z teorie (např. mortalita v určitém regionu – použijeme údaje o mortalitě celého státu). Potom určíme hodnotu výrazu $L(L-1)$ pro každý polygon a provedeme sumaci pro celou oblast. Potom určíme hodnoty E_{BW} a σ_{BW} .

Máme-li k dispozici pozorované počty spojů (O_{BW}), potom můžeme vyjádřit hodnotu z-skóre:

$$Z = \frac{O_{BW} - E_{BW}}{\sigma_{BW}}$$

Podle pravděpodobnosti rozdělení hodnot Z-skóre platí, že jakákoliv hodnota Z ležící mimo interval $(-1,96; 1,96)$ má pravděpodobnost výskytu menší než 5 případů ze 100 ($\alpha=0,05$).



Obr. 7.9 Příklady prostorového uspořádání černých a bílých polygonů v rámci studované oblasti (a, b, c) a počty sousedů jednotlivých ploch (d)

PŘÍKLAD:

Na obrázku (Obr. 7.9) je oblast obsahující 7 polygonů. Naším cílem je metodou JCS určit, zda v této oblasti existuje statisticky významná negativní prostorová autokorelace ve výskytu „černých“ (B) a

„bíých“ (W) ploch. Jako vah využijeme prvků binární matice. Podle výše uvedených vzorců musíme vyčíslit hodnoty O_{BW} , E_{BW} , σ_{BW} .

- 1) Spočteme celkový počet všech spojů ve studované oblasti, tedy hodnota $J=11$.
- 2) Určíme způsob definice sousedství – v tomto případě za sousedy považujeme pouze polygony, které spolu sousedí hranou (rook's case).
- 3) Určíme hodnoty pravděpodobností p, q výskytu „černé“ či „bílé“ plochy. V tomto případě předpokládáme, že $p=0,3$ a $q=0,7$.
- 4) Z obr. d určíme pomocí následující tabulky hodnotu $\sum L(L-1)$

Oblast	L	L-1	L(L-1)
A	3	2	6
B	2	1	2
C	3	2	6
D	5	4	20
E	3	2	6
F	3	2	6
G	3	2	6
Σ	22		52

- 5) Vyčíslíme hodnoty , E_{BW} , σ_{BW} :

$$E_{BW} = 2Jpq = 2 * 11 * 0,3 * 0,7 = 4,62$$

$$\sigma_{BW} = 2,1$$

- 6) Pro jednotlivé varianty na obrázku a, b, c jsou hodnoty pozorovaných počtů spojů (O_{BW}) $O_{BW} = 4, 6$ resp 8

- 7) Pro konfigurace „černých“ a „bíých“ poch uvedené na obrázku vyjádříme hodnotu z-skóre:

$$a) Z = \frac{4 - 4,62}{2,1} = -0,29$$

$$b) Z = \frac{6 - 4,62}{2,1} = 0,65$$

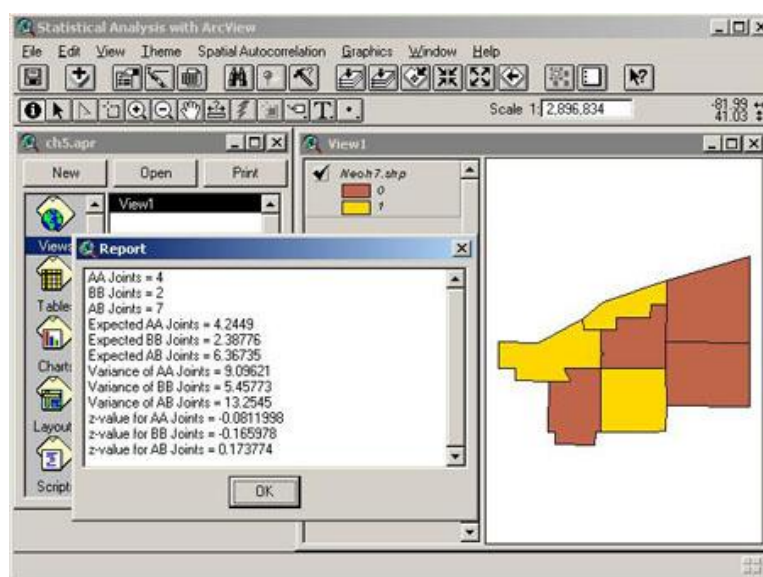
$$c) Z = \frac{8 - 4,62}{2,1} = 1,61$$

- 8) Interpretace: Žádná z hodnot Z-skóre nepřesahuje prahovou hodnotu $\pm 1,96$ a tedy uvedená uspořádání nevykazují statisticky významnou negativní prostorovou autokorelaci na hladině významnosti $\alpha=0,05$.

7.7 Náhodné vzorkování

V tomto případě závisí pravděpodobnost, zda je polygon bílý nebo černý, na celkovém počtu černých polygonů a počtu bílých polygonů ve studovaném území. Obrázek 4.7. uvádí tři typy prostorového uspořádání sedmi polygonů ve studované oblasti. Protože ve všech třech případech jsou počty B a W polygonů stejné (jsou jen jinak uspořádané) hodnoty pravděpodobnosti budou: $p=3/7$ a $q=4/7$.

Dále se vypočtou hodnoty očekávaných počtů spojů a jejich směrodatné odchylky. Výpočetní vzorce jsou jiné než v případě normálního vzorkování (viz. Lee, Wong, 2000, str. 154 – 155). Postup výpočtu je však analogický výše uvedenému příkladu.



Obr. 7.10 Příklad výstupu z metody JSC v programu ArcView

7.8 MORAN a GEARY indexy pro hodnocení prostorové autokorelace plošných jevů

Metoda Joint count statistics (JCS) má značná omezení z hlediska typu dat. Pro intervalová a poměrová data jsou stejně jak v případě jevů vztažených k bodům nejvyužívanějšími měřeními prostorové autokorelace plošných jevů indexy Moranův (I) a Gearyho (C)

Oba indexy mají některé společné charakteristiky, jejich statistické vlastnosti však jsou rozdílné. Vhodnější vlastnosti vzhledem k rozdělení hodnot má index I. Oba indexy jsou založeny na porovnávání hodnot atributů sousedních ploch. Mají-li tyto sousední plochy v celé studované oblasti podobné hodnoty, potom obě statistiky budou svědčit o silné pozitivní prostorové autokorelaci a naopak. Obě statistiky využívají odlišný přístup k porovnávání hodnot sousedních ploch.

7.8.1 Moranův index I

Index se vypočte podle následujícího vzorce:

$$I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2}$$

kde x_i je hodnota proměnné v ploše i

w_{ij} jsou váhy, W matice vah

Hodnota indexu kolísá od -1 pro negativní prostorovou autokorelaci do +1 pro pozitivní prostorovou autokorelaci. Očekávaná hodnota indexu je v případě nulové prostorové autokorelace je rovna

$$E_I = -\frac{1}{(n-1)}$$

Váhy se v případě tohoto indexu počítají z matic binární či stochastické (viz výše). Je-li použita binární matice, potom W ve jmenovateli je rovno dvojnásobku počtu hranic ve zpracovávané oblasti ($2J$).

Pokud jsou plochy s indexem i a j sousedé bude v čitateli $w_{ij} = 1$, pokud nesousedí bude 0. Pokud sousedí, vyjádří se součin odchylek hodnot i a j od průměru. Tyto součiny se sumují pro všechny sousedy. Jestliže **obě** sousední hodnoty budou nadprůměrné (ale i podprůměrné) dostaneme velké kladné číslo. Obě tyto situace ukazují na pozitivní autokorelaci – tedy podobné hodnoty jsou vedle sebe (sousedí spolu). Naopak, pokud hodnota v jedné ploše bude nadprůměrná a ve druhé podprůměrná – potom to indikuje negativní autokorelaci. **Budou-li ve zpracovávané oblasti převažovat sousedé s obdobnými hodnotami, Moranův index I bude kladný.**

Čítec obsahuje výraz pro kovarianci $(x_i - \bar{x})(x_j - \bar{x})$, která je také základem pro definování Pearsonova korelačního koeficientu r . Na rozdíl od korelačního koeficientu, kovariance v případě Moran's I je kovariancí dvou ploch v prostoru a ve výše uvedeném vztahu pro I je vypočtena pouze pro případy, kdy plochy spolu sousedí. Jmenovatel vzorce je suma čtverců odchylek vážená maticí sousedství W .

Interpretace Moran's I :

Vypočteme hodnoty I a $E(I)$ a následně musíme zjistit, zda rozdíl mezi nimi je statisticky významný. Tento rozdíl je opět nutné vztáhnout k míře rozptylu (např. směrodatné chybě - SE - viz. výklad k bodům) a pomocí ní odvodit standardizovanou hodnotu z-skóre

Odhady rozptylu resp. směrodatné chyby se budou lišit podle způsobu, jakým mohou být hodnoty vyšetřovaného atributu přiřazeny k jednotlivým plochám („sampling assumption“).

Za **předpokladu normality** jsou hodnoty atributu x_i nezávislé a pocházejí ze základního souboru s normálním rozdělením, nejsou nijak omezeny daným prostorovým uspořádáním ve studované oblasti. Z tohoto předpokladu se rozptyl vypočte:

$$\sigma^2(I) = \frac{n^2 S_1 - n S_2 + 3(W)^2}{(W)^2 (n^2 - 1)}$$

Za **předpokladu náhodnosti** je množina hodnot fixní. Konstantní není poloha spojená s určitou hodnotou atributu. Jinými slovy – existuje mnoho způsobů, jak je v prostoru rozmístěna daná množina hodnot. Naše rozmístění je jen jedno z možných.

Určení hodnoty rozptylu:

$$\sigma^2(I) = \frac{n[(n^2 - 3n + 3)S_1 - nS_2 + 3W^2] - \left[\frac{1/n \sum (x_i - \bar{x})^4}{[1/n \sum (x_i - \bar{x})^2]^2} \right] [S_1 - 2nS_2 + 6W^2]}{(n-1)(n-2)(n-3)(W^2)}$$

Získáme-li hodnotu rozptylu, potom můžeme vyčíslit standardizovanou hodnot $Z_n(I)$

$$Z_n = \frac{I - E(I)}{\sigma^2(I)}$$

Pokud je hodnota $Z_n(I)$ menší (resp. větší) než -1,96 (resp. 1,96) je hodnota indexu I statisticky významně negativní (resp. pozitivní) na hladině významnosti $\alpha=0,05$.

7.8.2 Gearyho poměr C (Geary's Ratio, C index)

Tento index je definován obdobně:

$$C = \frac{(n-1) \sum \sum w_{ij} (x_i - x_j)^2}{2W \sum (x_i - \bar{x})^2}$$

Pro výpočet indexu se jako vah využívá jedné z výše uvedených typů matic prostorových vah, nejčastěji matice binární či stochastické. Ve srovnání se vzorcem pro výpočet Moranova indexu je zřejmé, že Gearyho index se liší především v čitateli výrazu. Moranův index porovnává hodnoty atributů sousedních ploch prostřednictvím odchylek od průměru, naproti tomu Gearyho index porovnává hodnoty atributů přímo mezi sebou. Pro hodnotu indexu není rozhodující, která z hodnot x_i a x_j je větší či menší, ale jaký je jejich absolutní rozdíl – jejich nepodobnost (ve výrazu je druhá mocnina jejich rozdílu).

Gearyho index nabývá hodnot v intervalu 0 až 2. Hodnota nula indikuje dokonalou pozitivní autokorelaci (všechny sousední hodnoty atributů jsou stejné). Naopak hodnota 2 indikuje dokonalou negativní prostorovou autokorelaci. Na rozdíl od Moranova indexu, očekávaná hodnota Gearyho indexu nezávisí na počtu posuzovaných ploch n , ale má vždy hodnotu 1. Hodnota 1 znamená nulovou prostorovou autokorelaci.

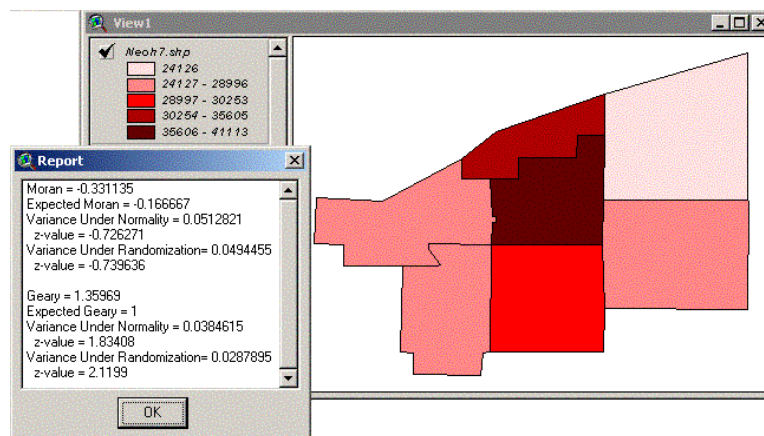
Vypočtené hodnoty indexu C lze porovnat s hodnotou jedna (očekávanou), pro prokázání statisticky významného rozdílu je však stejně jako v předchozích případech nutné vypočítat hodnotu z -skóre. Nejprve je nutné vypočítat rozptyl hodnoty indexu C . Hodnota rozptylu se opět vypočte rozdílně v závislosti na předpokladu normality či náhodnosti.

Například za předpokladu normality:

$$\sigma^2(c) = \frac{(2S_1 + S_2)(n-1) - 4W^2}{2(n+1)W^2}$$

Za předpokladu náhodnosti: (vzorec viz. Lee a Wong, 2000, s. 162)

Hodnoty z -skóre jsou založené na rozdílu pozorovaných a očekávaných hodnot. Jestliže hodnota indexu $C = 0$ značí perfektní pozitivní prostorovou autokorelaci a $C = 1$ nulovou, potom negativní hodnota z -skóre značí pozitivní prostorovou autokorelaci a kladná hodnota z -skóre značí autokorelaci negativní.



Obr. 7.11 Vstupní data a výsledky prostorové autokorelace (I a C indexy) pro průměrný příjem sedmi států v Ohio.

Příklad 1: Na obrázku 5.1 je kartogram průměrného příjmu pro sedm států Ohia. Z hodnot vypočtených indexů vyplývá, že hodnota Moranova indexu indikuje **negativní** prostorovou autokorelaci (státy s vysokou hodnotou studovaného atributu jsou blízko států s nízkými hodnotami). Tato tendence však není statisticky významná na hladině 5 %.

Naopak podle vypočtených hodnot Gearyho indexu existuje statisticky významná negativní prostorová autokorelace v hodnotách průměrného příjmu u sedmi studovaných států celého regionu.

7.9 Obecná G-statistika

Oba výše uvedené indexy I a C mají dobře definované statistické vlastnosti, které popisují prostorovou autokorelaci globálně (jednou hodnotou pro celou zpracovávanou oblast). Nejsou však efektivní k identifikaci rozdílných shluků prostorového uspořádání uvnitř oblasti. Oba indexy jsou sice citlivé k identifikaci oblastí s podobnými hodnotami atributů, nerozlišují však, zda tyto podobné hodnoty nabývají vysokých či nízkých hodnot. Shluky ploch (těž. místa prostorové koncentrace - spatial concentration) vysokých hodnot vyšetřovaného atributu ve studované oblasti se označují jako „**hot spots**“, naopak místa se shluky nízkých hodnot jako „**cold spots**“.

Odlišit oby typy shluků lze pomocí tzv. **obecné G-statistiky (general G-statistics)**. Stejně jako v případě Moranova a Gearyho indexu je i G -statistika založena na míře prostorové asociace, která dává v čitateli výrazu do vztahu hodnoty atributu v ploše (bodě, místě) i a j . Obecná G -statistika je definována takto:

$$G(d) = \frac{\sum_i \sum_j w_{ij}(d) x_i x_j}{\sum_i \sum_j x_i x_j}$$

pro i různá od j . G -statistika je definována vzdáleností d mezi plochou i a plochami sousedními. Váha $w_{ij}(d)$ má hodnotu 1, jestliže se plocha j nachází ve vzdálenosti menší či rovné d od plochy i , jinak má váha hodnotu 0. Matice vah je tedy maticí binární a symetrickou, vztahy sousedství jsou však definovány vzdáleností d . Suma těchto vah matice se rovná:

$$W = \sum_i \sum_j w_{ij}(d)$$

pro i různá od j . V důsledku takového definování vah, páry x_i a x_j nebudou zahrnuty v čitateli, pokud i a j jsou od sebe dále než d . Naproti tomu ve jmenovateli jsou zahrnuty všechny páry x_i a x_j bez ohledu na jejich vzdálenost. Z toho plyne, že jmenovatel bude vždy větší, maximálně však roven (při velkém d) čitateli. Čítec výrazu pro $G(d)$ statistiku, bude mít velkou hodnotu pokud sousední hodnoty budou velké a naopak. Vysoké hodnoty $G(d)$ potom indikují prostorovou asociaci vysokých hodnot (hot spots) zkoumaného atributu, nízké $G(d)$ potom prostorovou asociaci nízkých hodnot (cold spots).

Před výpočtem $G(d)$ je nutné určit vzdálenost d , která definuje plochy, které budou považovány za sousedy plochy posuzované. Musí být vhodně zvolena tak, aby posuzovaná plocha měla alespoň jednoho souseda.

K interpretaci a k hodnocení statistické významnosti $G(d)$ je nutné jako u výše uvedených indexů I a C vyčíslit očekávanou hodnotu $G(d)$, tedy $E(G)$ a následně standardizovanou hodnotu z -skóre a tedy i rozptyl hodnoty $G(d)$. Očekávaná hodnota $G(d)$ bude:

$$E(G) = \frac{W}{n(n-1)}$$

Očekávaná hodnota statistiky odpovídá případu, kdy neexistuje žádná prostorová asociace. Např. je-li vypočtená hodnota $G(d)$ větší než očekávaná, můžeme říci, že pozorované uspořádání vykazuje pozitivní prostorovou asociaci. Statistickou významnost tohoto tvrzení je opět nutné testovat výpočtem hodnoty rozptylu $Var(G)$ (vzorec viz. Lee a Wong, 2000, s. 166) a následně z -skóre. Opět, hodnota z -skóre menší než 1,96 indikuje statisticky nevýznamný výsledek na hladině $\alpha=0,05$.

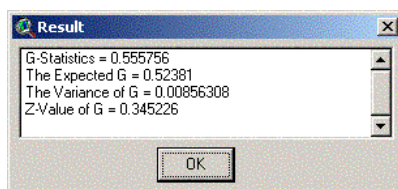
Příklad 2: Jsou použita stejná vstupní data jako v případě I a C indexů. Výchozí matice vzdáleností centroidů (Obr. 7.12) je převedena na matici binární na základě zvolené vzdálenosti d ($d=30$ mil)-Obr. 7.13.

Id	Geauga	Cuyahoga	Trumbull	Summit	Portage	Ashtabula	Lake
Geauga	0.0000	25.1508	26.7057	32.7509	25.0389	26.5899	12.6265
Cuyahoga	25.1508	0.0000	47.8151	23.4834	31.6155	50.8064	28.2214
Trumbull	26.7057	47.8151	0.0000	41.8561	24.4759	29.5633	36.7535
Summit	32.7509	23.4834	41.8561	0.0000	17.8031	58.0869	42.7375
Portage	25.0389	31.6155	24.4759	17.8031	0.0000	45.5341	37.4962
Ashtabula	26.5899	50.8064	29.5633	58.0869	45.5341	0.0000	24.7490
Lake	12.6265	28.2214	36.7535	42.7375	37.4962	24.7490	0.0000

Obr. 7.12 Výchozí matice vzdáleností centroidů

Id	Geauga	Cuyahoga	Trumbull	Summit	Portage	Ashtabula	Lake
Geauga	0.0000	1.0000	1.0000	0.0000	1.0000	1.0000	1.0000
Cuyahoga	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000
Trumbull	1.0000	0.0000	0.0000	0.0000	1.0000	1.0000	0.0000
Summit	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Portage	1.0000	0.0000	1.0000	1.0000	0.0000	0.0000	0.0000
Ashtabula	1.0000	0.0000	1.0000	0.0000	0.0000	0.0000	1.0000
Lake	1.0000	1.0000	0.0000	0.0000	0.0000	1.0000	0.0000

Obr. 7.13 Matice sousedství vypočtená pro d=30 z matice na obr. 5.2



Obr. 7.14 Výsledky výpočtu obecné G- statistiky pro vstupní data na obrázku 5.1 při použití matice vzdáleností centroidů a hodnotě definující vzdálenost d=30 mil.

Vypočtená hodnota $G(d)$ vykazuje mírnou úroveň prostorové asociace, podle hodnoty z-skóre však výsledek není statisticky významný. Jinými slovy – dané uspořádání průměrného příjmu v sedmi státech Ohia je spíše výsledkem náhody než určitého systematického procesu.

7.9.1 Lokální statistiky prostorové autokorelace

Všechny tři uvedené indexy jsou příkladem indexů globálních. Jsou sumární hodnotou prostorové autokorelace pro celou zpracovávanou oblast. Je však pravděpodobné, že hodnoty prostorové autokorelace se budou v různých sub-oblastech měnit. Navíc můžeme očekávat, že pozitivní autokorelaci lze nalézt v jednom sub-regionu a negativní v jiném. Proměnlivost prostorové autokorelace v rámci studované oblasti lze vyšetřovat výše uvedenými indexy modifikovanými pro detekování prostorové autokorelace v lokálním měřítku.

LISA (Local Indicators of Spatial Association)

Jedná se o lokální verze Moranova a Gearyho indexu. Ke zjištění úrovně prostorové autokorelace na lokální úrovni je nutné vypočítat hodnotu indexu pro každou plochu zpracovávaného území. Lokální Moranův index pro jednotku i je definován takto:

$$I_i = z_i \sum_j w_{ij} z_j$$

kde z_i a z_j jsou odchylky od průměru nebo

$$z_i = \frac{(x_i - \bar{x})}{\sigma}$$

kde σ je směrodatná odchylka x_i . Podobně jako v případě globálního Moranova indexu znamenají vysoké hodnoty kumulaci podobných hodnot atributů (vysokých či nízkých) v sousedních plochách, nízké hodnoty potom kumulaci odlišných hodnot atributů. Obecně hodnoty w_{ij} mohou představovat po řadách standardizovanou matici vah, lze použít i jiných matic vah.

Zjištěné hodnoty lokálního Moranova indexu je nutné porovnat s očekávanými hodnotami a testovat statistickou významnost jejich rozdílu pomocí z-skóre.

Očekávané hodnoty při hypotéze náhodnosti:

$$E[I_i] = -w_{i.}/(n-1)$$

a hodnota rozptylu:

$$Var[I_i] = w_{i.}^2 \frac{(n - m_4/m_2^2)}{n-1} + 2w_{i(kh)} \frac{(2m_4/m_2^2 - n)}{(n-1)(n-2)} - \frac{w_{i.}^2}{(n-1)^2}$$

kde

$$w_{i.}^2 = \left(\sum_j w_{ij} \right)^2$$

$$w_{i.}^{(2)} = \sum_j w_{ij}^2 \quad \text{pro } i \neq j$$

a výraz

$$2w_{i(kh)} = \sum_{k \neq i} \sum_{h \neq i} w_{ik} w_{ih}$$

Každá plocha ve zpracovávaném území má svoji I hodnotu a té přísluší hodnota očekávaná a také jistá hodnota rozptylu. Hodnoty I mohou být vynášeny do mapy v podobě kartogramu.

Lokální verze Gearyho poměru je definována následovně:

$$c_i = \sum_j w_{ij} (z_i - z_j)^2$$

Hodnoty rozdělení lokálního Gearyho indexu nemají tak vhodné vlastnosti jako v případě indexu Moranova. Jejich interpretace je však obdobná jako v případě globální verze indexu. Shlukování podobných hodnot atributů vede k nízkým hodnotám tohoto indexu a naopak.

Lokální G-statistika

Měří asociaci hodnot atributů v ploše i a v plochách okolních definovaných vzdáleností d :

$$G_i(d) = \frac{\sum w_{ij}(d)x_j}{\sum_j x_j} \quad \text{pro } i \neq j$$

Obdobně jako v předchozích případech je nutné interpretovat hodnotu indexu pomocí, očekávaných hodnot, hodnot rozptylu a standardizovaných skóre. Očekávané hodnoty se vypočtou následovně:

$$E(G_i) = W_i / (n - 1)$$

kde

$$W_i = \sum_j w_{ij}(d)$$

Definice rozptylu:

$$\text{Var}(G_i) = E(G_i^2) - [E(G_i)]^2$$

a

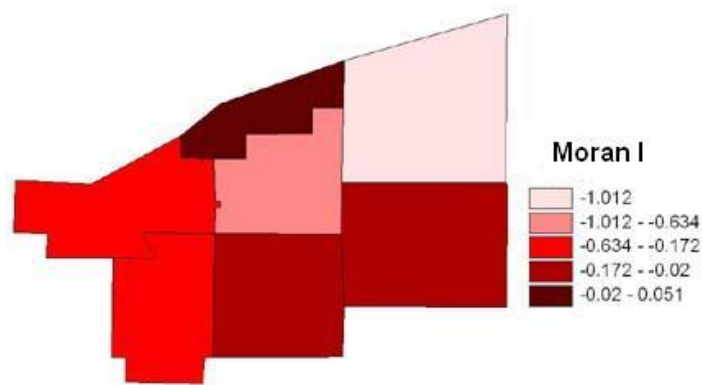
$$E(G_i^2) = \frac{1}{(\sum_j x_j)^2} \left[\frac{W_i(n-1-W_i) \sum_j x_j^2}{(n-1)(n-2)} \right] + \frac{W_i(W_i-1)}{(n-1)(n-2)} \quad \text{pro } i \neq j$$

Vysoká hodnota z -skóre je spojena s výskytem shluků podobných a vysokých hodnot indexu. Jestliže je shluk tvořen nízkými hodnotami, z -skóre bude nabývat velkých záporných hodnot. Hodnoty z -skóre kolem nuly indikují neexistenci zřejmého prostorového uspořádání hodnot atributů v plochách studovaného území.

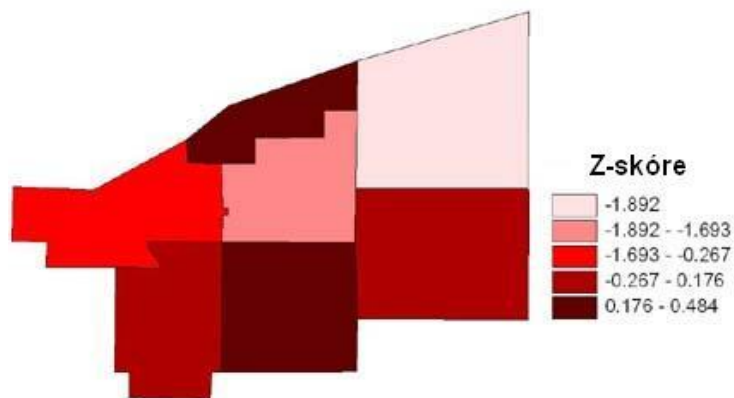
Příklad 3: Pro data z příkladu 1 byly vypočteny hodnoty lokálního Moranova indexu I (pro každý stát). Jako matice vah byla použita matice stochastická (Obr. 7.15). Výsledky jsou prezentovány ve formě kartogramu na následujících obrázcích (Obr. 7.16 a Obr. 7.17).

Id	Geauga	Cuyahoga	Trumbull	Summit	Portage	Ashtabula	Lake
Geauga	0.0000	0.1667	0.1667	0.1667	0.1667	0.1667	0.1667
Cuyahoga	0.2500	0.0000	0.0000	0.2500	0.2500	0.0000	0.2500
Trumbull	0.3333	0.0000	0.0000	0.0000	0.3333	0.3333	0.0000
Summit	0.3333	0.3333	0.0000	0.0000	0.3333	0.0000	0.0000
Portage	0.2500	0.2500	0.2500	0.2500	0.0000	0.0000	0.0000
Ashtabula	0.3333	0.0000	0.3333	0.0000	0.0000	0.0000	0.3333
Lake	0.3333	0.3333	0.0000	0.0000	0.0000	0.3333	0.0000

Obr. 7.15 Stochastická matice vah k definování sousedství pro výpočet lokálního Moranova indexu I



Obr. 7.16 Kartogram hodnot lokálního Moranova indexu I



Obr. 7.17 Kartogram hodnot z-skóre pro lokální Moranův index I

Interpretace: Vysoké hodnoty indexu I mají ty státy, jejichž sousedé mají velmi podobné hodnoty studované charakteristiky. Podle z-skóre žádná z hodnot není statisticky významná a dané uspořádání průměrných příjmů v sedmi státech lze interpretovat jako náhodný proces.

Obdobným způsobem lze vizualizovat a hodnotit výsledky analýzy založené na lokálním indexu C a lokální G -statistice.

Moranovo korelační pole (Moran Scatterplot)

Lokální statistiky vystihují prostorovou heterogenitu v jednotlivých částech studovaného území. Pomocí nich je tedy možné jistým způsobem identifikovat oblasti s neobvyklými hodnotami měř prostorové autokorelace, které lze označit jako oblasti s odlehlými hodnotami (outliers). Efektivním nástrojem pro takovouto diagnostiku území je Moranovo korelační pole založené na regresním počtu.

Předpokládejme, že x značí vektor hodnot x_i s odchylkami od průměru ($x_i - \bar{x}$) a dále W značí po řádcích standardizovanou matici vah. Potom můžeme sestavit regresní závislost hodnot Wx na x . Směrnice této regresní závislosti indikuje vzájemný vztah sousedních hodnot atributů.

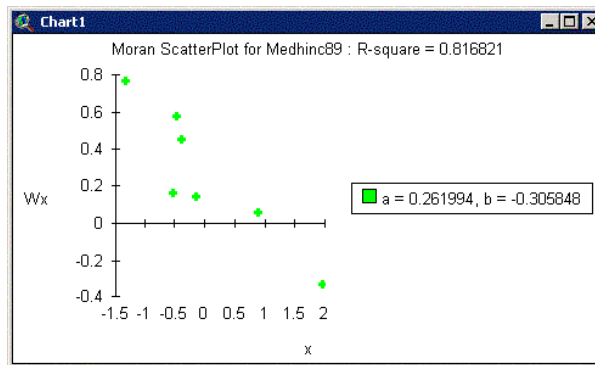
Tedy

$$x = a + IWx$$

kde a značí vektor koeficientů - (intercept). Hodnota I je regresní koeficient reprezentující směrnici a také hodnotou Moranova globálního indexu I . Vynesení regresní závislosti Wx na x umožňuje identifikovat odlehle hodnoty. Pokud budou mít všechna pozorování podobné hodnoty prostorové autokorelace, v korelačním poli budou body blízko regresní čáry. Naopak pokud některá pozorování budou ukazovat lokálně výrazně vysoké či nízké hodnoty prostorové autokorelace ve vztahu k jejich sousedům, tato pozorování budou v grafu tvořit body výrazně nad či pod regresní čarou.

Regresní čára vyjadřuje obecný trend hodnot prostorové autokorelace v celém zpracovávaném území a parametr její směrnice je index I .

Příklad 4: Hodnota Moranova indexu (viz. Příklad 1) indikuje slabou **negativní** prostorovou autokorelaci (státy s vysokou hodnotou studovaného atributu jsou blízko států s nízkými hodnotami).



Obr. 7.18 Výsledek regresní analýzy a Moranovo korelační pole (Moran Scatterplot) pro průměrný příjem sedmi států Ohia (příklad 1). Parametr b představuje hodnotu Moranova indexu I

Z grafu je patrné že příjem (x) je nepřímo úměrný vážené hodnotě příjmu (Wx). Množinou bodů lze proložit přímkou. Body, které se výrazně odchyľují od přímky představují „outliers“ – představují oblasti s výrazně odlišnými hodnotami prostorové autokorelace.

Interpelace s ohledem na polohu bodů v jednotlivých kvadrantech

- high-high, low-low (2. nebo 3. kvadrant) = spatial clusters
- high-low, low-high (1. nebo 4. kvadrant) = spatial outliers