Strukturní analýza

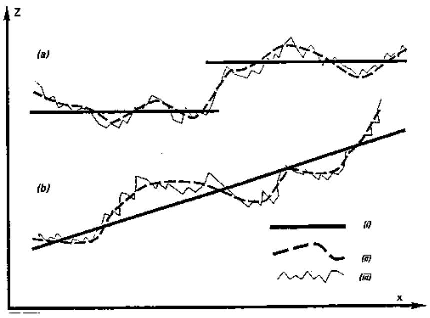
Strukturní analýza a metody krigingu

Žádná z dosud zmíněných interpolačních 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 poskytují geostatistické postupy založené na tzv. **strukturní analýze.** Její výsledky jsou využitelné v interpolačních postupech **krigingu**.

Základní komponenty spojitého povrchu



I – trendová složka – drift
ii – regionalizovaná proměnná
iii – náhodná složka

Základní komponenty spojitého povrchu

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

x **- pozice v 1, 2 či 3 rozměrném prostoru**

Z **- interpolovaná proměnná**

Z(x) **- hodnota proměnné v bodě *x***

μ(x) **- deterministická složka (trend)**

ε'(x) **- stochastická složka (regionalizovaná proměnná) - lokálně proměnné, ale prostorově závislé reziduum od *μ(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*.

Odhad jednotlivých komponent

Nulový trend - *μ(x)* se rovná průměrné hodnotě proměnné *z* v ploše a platí:

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

kde *Z(x)* a *Z(X+h)* jsou odhady hodnot náhodné proměnné z v poloze *x, x+h.*

Předpokládáme, že rozptyl rozdílů závisí pouze na vzdálenosti mezi místy, tedy:

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

γ(h) **- semivariance**

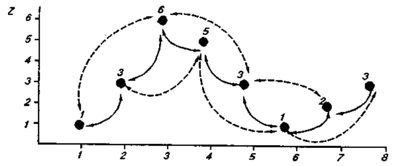
Pokud máme odhad proměnné *μ(x)*, zbývající kolísání má konstantní rozptyl a diference mezi dvěma místy jsou pouze funkcí jejich vzdálenosti:

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

Strukturní analýza - variografie

- Geostatistická strukturní analýza - 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.
- Ke kvantifikaci prostorové autokorelace, která vyjadřuje skutečnost, že objekty blízké si jsou více podobné než objekty vzdálenější slouží **strukturální funkce** -měří sílu korelačního vztahu jako funkci vzdálenosti.
- Strukturální 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.

Příklad výpočtu měr prostorové variability pro 1D - řadu hodnot



průměr = (1+3+6+5+3+1+2+3)/8=3,0

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

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

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

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

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

Semivariance jako strukturální funkce

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

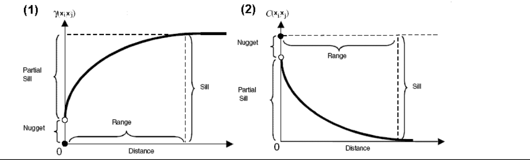
var značí rozptyl

Jsou-li 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**.

Semivariogram (1) je **mírou nepodobnosti**. Jinou strukturální funkcí je kovarianční funkce – ta je mírou podobnosti (2). Obě jsou měrami prostorové autokorelace.



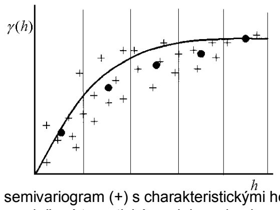
Experimentální semivariogram

Strukturní analýza v 2D – výpočet semivariogramu z naměřených dat:

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

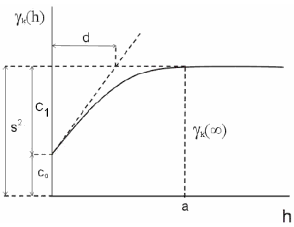
n - počet dvojic bodů pozorování proměnné s atributem *z* vzdálených o hodnotu *h*

h - tzv. lag - vzdálenost dané dvojice bodů.



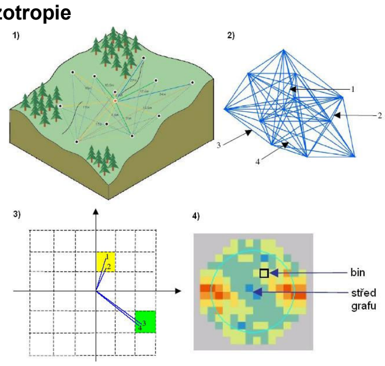
Experimentální semivariogram (+) s charakteristickými hodnotami pro vzdálenosti h (•) a proložený teoretický model semivariogramu (plná čára)

Prvky semivariogramu



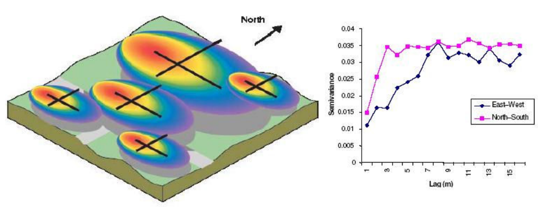
a - dosah (range), d – rozpětí, c_0 - zbytkový rozptyl (nugget), c=c_0 + c_1 - práh (sill), h – lag (krok vzdálenosti)

Efekt anizotropie



Princip grupování hodnot semivariancí na základě podobné vzdálenosti a plošný graf semivariance (4).

Efekt anizotropie



Povrch vykazující efekt anizotropie a odpovídající empirické semivariogramy

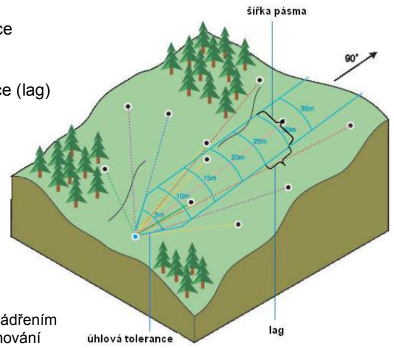
Anizotropní semivariogram se liší především odlišnou hodnotou dosahu pro specifické směry, další charakteristiky semivariogramu (typ, práh, zbytkový rozptyl) se většinou nemění.

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.

Parametry tzv. směrových semivariogramů

- úhlovoá tolerance
- šířka pásma
- délková tolerance (lag)



Efekt anizotropie je vyjádřením náhodného procesu chování studované veličiny. Nelze ho zaměňovat s trendovou složkou.

Teoretický semivariogram

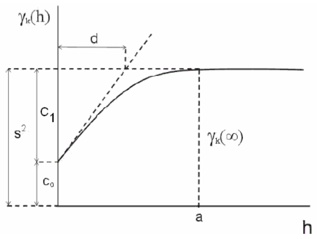
Je to model, který nejlépe aproximuje průběh experimentálního semivariogramu v okolí počátku a prahu.

Právě proces hledání teoretického semivariagramu 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

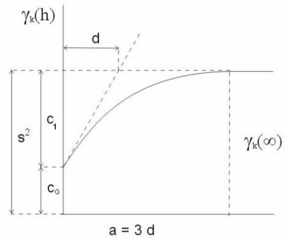
Sférický model



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

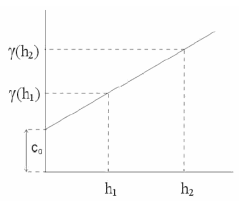
$$\gamma(h) = c_0 + c_1 \quad \text{pro} \quad h > a$$

Exponenciální model



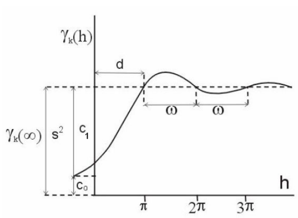
$$\gamma(h) = c_0 + c_1 * [1 - \exp(-h/d)] \quad \text{kde} \quad a = 3d$$

Lineární model



$$\gamma(h) = c_0 + bh$$ kde b je směrnice přímky

Sinový model

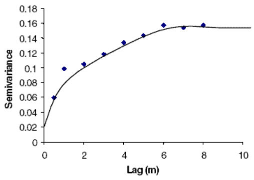


$$\gamma(h) = c_0 + c_1\left[1 - \frac{\sin(gh)}{gh}\right] \quad \text{kde} \quad g = \pi/\omega$$

Náhodný model

$$\gamma(h) = c_0$$

Složené modely



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

Indikátorové modely semivariogramů - konstruují se a využívají při strukturální analýze nominálních (kvalitativních) dat (barva, druh horniny).

Analýza a interpretace strukturálních funkcí I.

Konstrukci semivarioagramu a odvození teoretického modelu by měla předcházet důkladná analýza vstupních dat (ESDA – explorační analýza prostorových dat)

Důležitý je počet bodů uvažovaných pro vyjádření hodnot semivariance pro daný lag (*h*). Značný podíl šumu ve variogramu může být dále způsoben malým rozsahem vzorku použitého k výpočtu

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).

Vzdálenosti mohu být modifikovány efektem anizotropie - potom je nutné měnit tvar okolí. Anizotoropie 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.

Analýza a interpretace strukturálních funkcí II.

Přednost má 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 „rozkolísanosti“ stanovených bodů kolem teoretického modelu.

Úroveň prahu se obvykle doporučuje volit podle hodnoty statistického rozptylu.

Je-li hodnota dosahu použitého teoretického semivariagramu 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 semivarigramů.

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

Analýza a interpretace strukturálních funkcí III.

Pro účely interpolace a metodou krigování je účelné zvolit 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 stejný teoretický model.

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 semivariogamu 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. 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 nepostižitelné 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.

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. Rozlišujeme geometrickou a zonální anizotropii.

ESDA – průzkumová analýza prostorových dat

Souhrn metod popisné statistiky, které předchází strukturní analýze.

Cílem je ověřit některé vlastnosti vstupního datového souboru, které jsou nezbytnými předpoklady pro vlastní geostatistické metody.

Ověřujeme, že podobnost hodnot studovaného jevu závisí pouze na vzdálenosti dvou porovnávaných bodů měření a ne na jejich poloze.

ESDA zahrnuje především:

- výpočet charakteristik úrovně a variability, analýza odlehlých hodnot
- studium histogramu s cílem ověření normality rozdělení
- konstrukci tzv. normálního kvantilového grafu (tzv. Q-Q grafu)
- analýzu trendové složky s cílem ověření stacionarity vstupních dat

Výsledkem ESDA je nutnost provedení transformace vstupních dat