

## Geostatistické metody interpolace

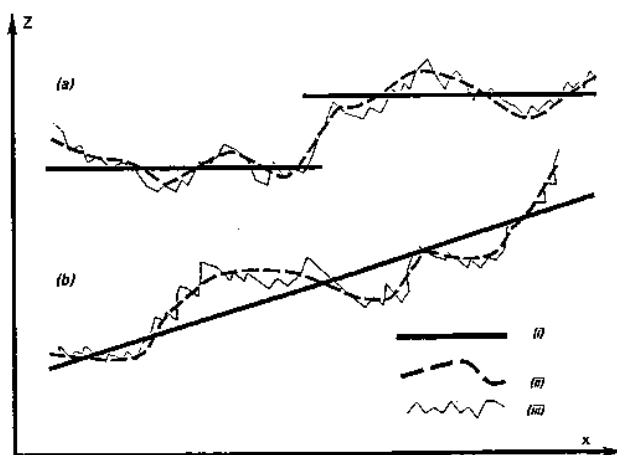
Při použití většiny dosud uvedených interpolačních algoritmů nemáme *a priori* objektivní informaci o tom, zda způsoby definování okolí interpolovaného bodu, body použité k interpolaci a také jejich váhy jsou zvoleny vhodně. V případě kvalitních vstupních dat dává většina interpolačních technik podobné výsledky (Borough, 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. 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 residuum od  $\mu(x)$

$\varepsilon''$  - náhodná, prostorově nezávislá složka, gaussovský šum s nulovým průměrem a s rozptylem  $\sigma^2$ . 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.

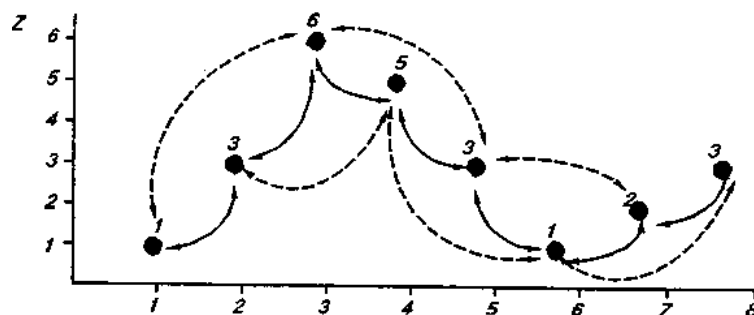
To značí, že jakmile jsme odhadli příspěvek strukturální proměnné  $\mu(x)$ , zbývající kolísání má konstantní rozptyl a difference mezi dvěma místy jsou pouze funkcí jejich vzdálenosti. Výše uvedený vztah lze přepsat:

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

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

### Strukturální analýza

Geostatistická strukturální 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. 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.



Obr. 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 **strukturální funkce**. Strukturální funkce měří sílu korelačního vztahu jako funkci vzdálenosti. Na obr. 2 je pro jednoduchost odvozena strukturální funkce pro řadu pravidelně rozmístěných bodů na linii (tedy pro 1D). Charakteristiky, které popisují prostorovou variabilitu lze odvodit z měř úrovně a variability následovně:

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

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

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

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

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

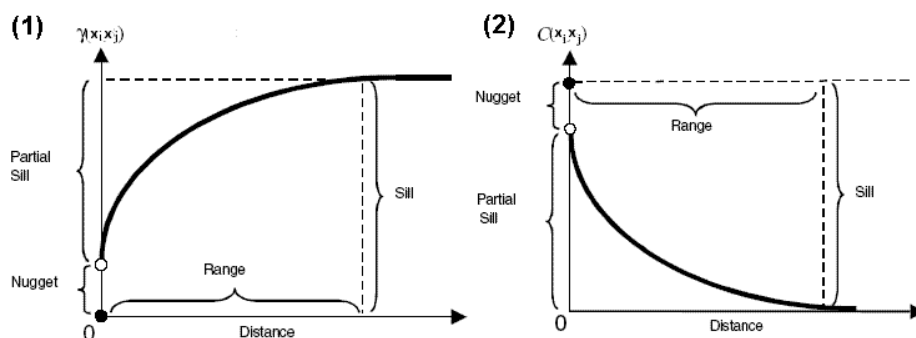
$$\text{semivariance}(3) = [(1-5)^2+(3-3)^2+(6-1)^2+(5-2)^2+(3-3)^2]/5=12,50$$

...

**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 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).

## Strukturní analýza v 2D

Výše uvedený příklad určení semivariance platí pro řadu pravidelně rozmístěných bodů. Strukturní 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.

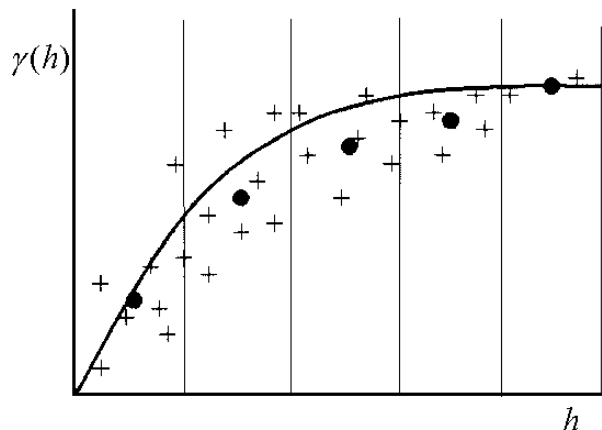
Abychom popsali závislost hodnot studovaného jevu v prostoru, vyneseme hodnoty semivariací pro všechny dvojice bodů do semivariogramu obdobně jako ve výše uvedeném případě. **Semivariogram** je strukturální funkce, která popisuje závislost průměrné kvadratické diference hodnot prostorové proměnné veličiny  $Z$  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 s atributem  $z$  vzdálených o hodnotu  $h$

$h$  - tzv. lag - vzdálenost dané dvojice bodů.



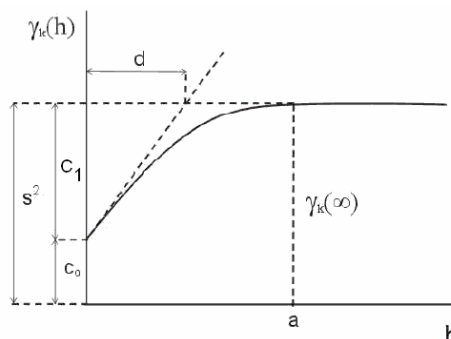
Obr. 4 Experimentální semivariogram (+) s charakteristickými hodnotami pro vzdálenosti  $h$  (●) a proložený teoretický model semivariogramu (plná čára)

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**.

### Prvky semivariogramu

Na obr. 5 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 rozptýl 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.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)

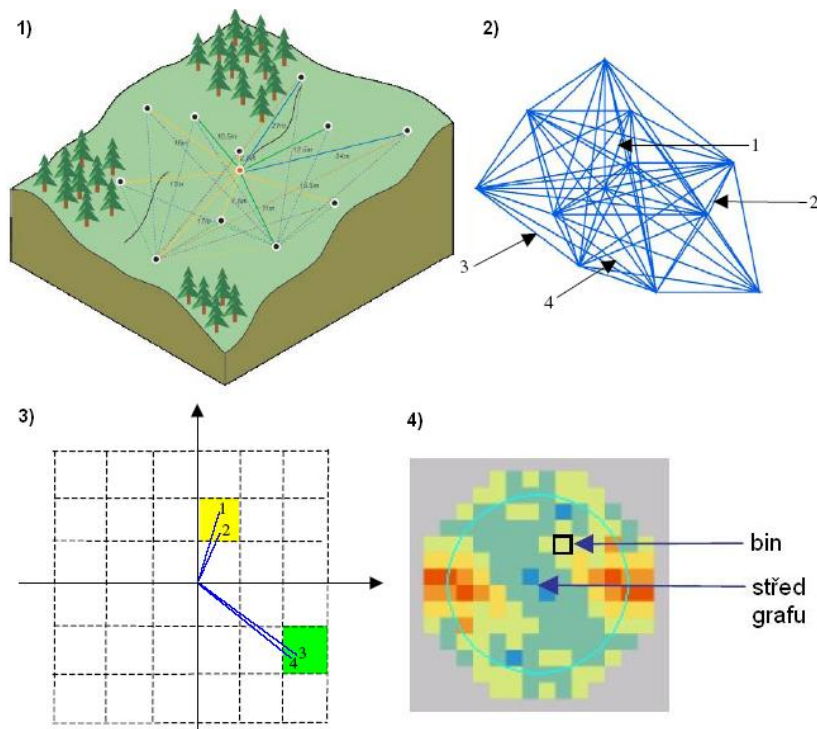
Z výše uvedeného by také mělo platit, že proložená křivka semivariogramu by měla procházet počátkem souřadné soustavy (nulová vzdálenost mezi body znamená zákonitě také nulovou hodnotu rozptylu). Velmi často nenabývají experimentální semivariogramy v počátku nulové hodnoty; protínají osu  $y$  v nenulové 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  $\varepsilon''$  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  $\varepsilon'$ . 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.

### Efekt anizotropie

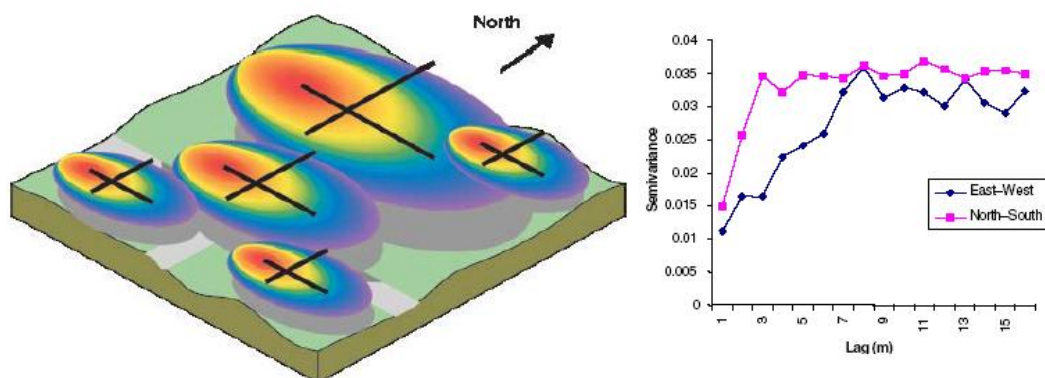
V případě velkého množství nepravidelně rozmístěných bodů je vhodné hodnotu semivariance vyjádřit pro skupiny bodů přibližně stejně vzdálených. V tomto případě se do výpočtu hodnoty semivariance pro dané  $h$  berou všechny body padnoucí do mezikruží určeného **tolerancí délky kroku**. Toleranci délky 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. 6. 2) a 3). Hodnoty semivariací jsou potom vyjádřeny např. barvou.



Obr.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 o 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.

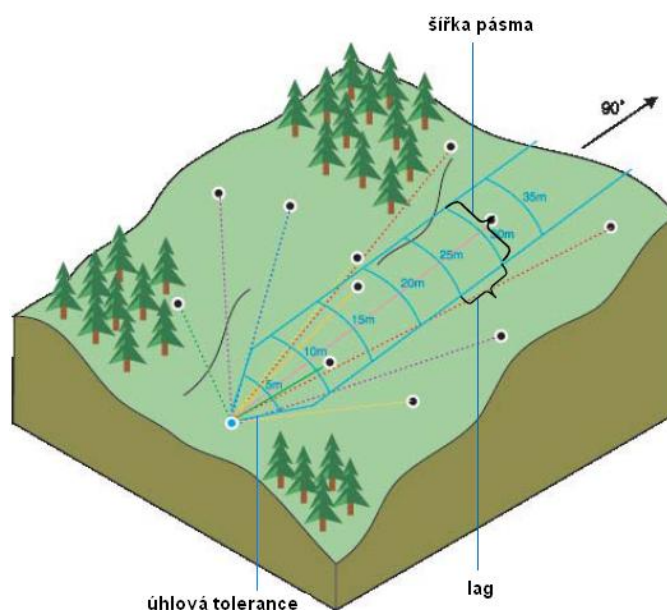


Obr. 7 Povrch vykazující efekt anizotropie a odpovídající empirické semivariogramy

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. 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  $\hat{\gamma}(h)$ .

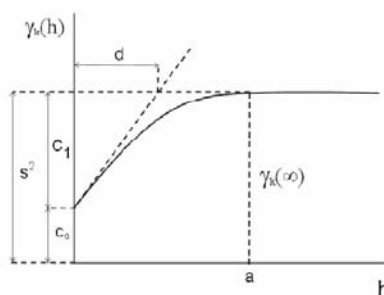
Model nalezený pro danou množinu dat závisí jak na experimentálních, tak teoretických předpokladech. Vlastnosti, které prakticky vedou k určení konkrétního teoretického modelu, jsou:

- 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.

## Přehled teoretických semivariogramů

**1. 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.

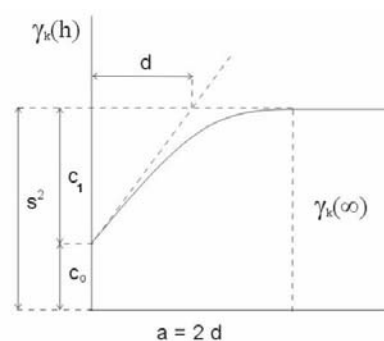


Obr. 9 Sférický model semivariogramu

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

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

## Kvadratický model



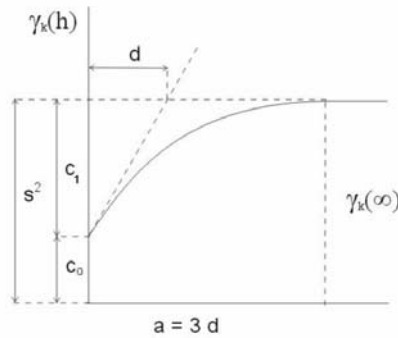
Obr. 10 Kvadratický model semivariogramu



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

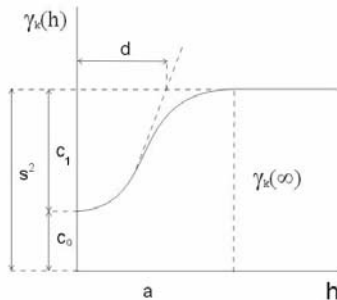
$$\gamma(h) = c_0 + c_1 \dots\dots\dots \text{pro } h > a$$

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



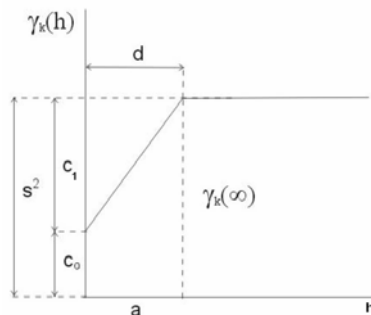
Obr. 11 Exponenciální model semivariogramu  
 $\gamma(h) = c_0 + c_1 * [1 - \exp(-h/d)]$ , kde  $a = 3d$

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

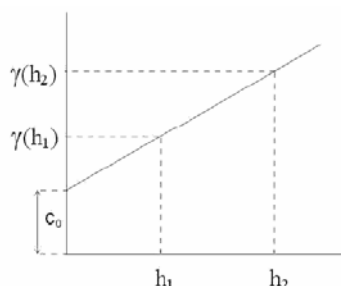
**Lineární model s prahem** - jednoduchý a poměrně často využívaný zvláště programy provádějícími interpolaci pomocí krigování na základě automaticky vypočítaného a vyhodnoceného semivariogramu. Při provádění strukturální analýzy se využívá raději jiných přechodových modelů.



Obr. 13 Lineární model semivariogramus prahem  
 $\gamma(h) = c_0 + bh \dots\dots\dots \text{pro } h \leq a$   
 $\gamma(h) = c_0 + c_1 \dots\dots\dots \text{pro } h > a$

**2. 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.

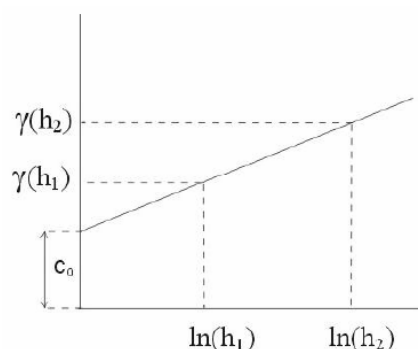
**Lineární model:**



Obr. 14 Lineární model semivariogramu

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

**Logaritmický semivariogram**



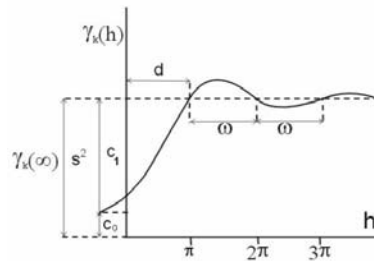
Obr. 15 Logaritmický model semivariogramu

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

**3. 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**

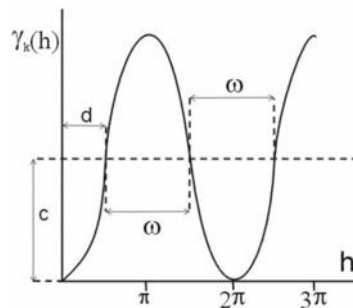


Obr. 16 Sinový model semivariogramu

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

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

**Cosinový model semivariogramu** - Nedochází k postupnému tlumení hodnot oscilací.



Obr. 17 Cosinový model semivariogramu

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

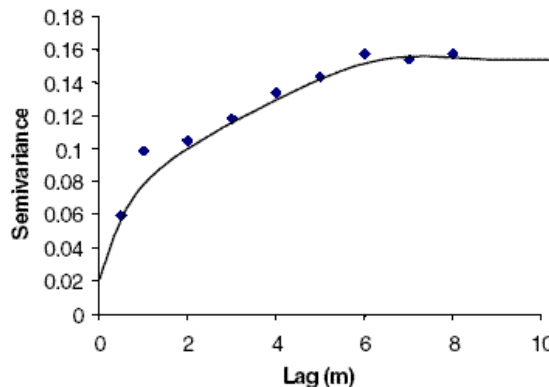
### Čistě náhodný model semivariogramu

$$\gamma(h) = c_0$$

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).

Další druhy semivariogramů

**Složené modely** (komplexní semivariogram)



Obr. 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).

### 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 ( $\epsilon''$ ), potom data obsahují takový šum, že interpolace nemá smysl. Jako nejlepší odhad  $z(x)$  je vhodné použít průměrnou hodnotu.

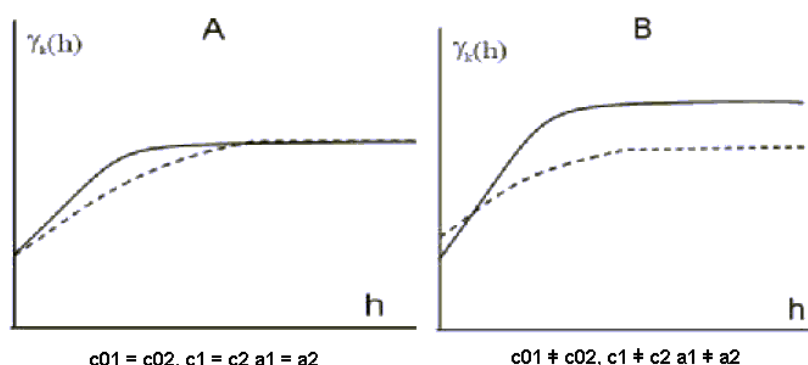
#### Charakteristiky pole popsané strukturní analýzou:

**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 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. 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. 19 Rozdíl mezi geometrickou (A) a zonální (B) anizotropií semivariogramu

## **ESDA – explorační (průzkumová analýza prostorových dat)**

Jedná se o souhrn metod popisné statistiky, které by měly předcházet strukturní analýze. Cílem těchto metod je ověřit některé vlastnosti vstupního datového souboru, které jsou nezbytnými předpoklady pro vlastní geostatistické metody. U analyzovaných vstupních dat zpravidla předpokládáme konstantní hodnotu průměru pro různé části zpracovávaných dat a dále také fakt, že prostorová autokorelace v poli hodnot je konstantní. Tedy 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

V případě, že se rozdělení vstupních dat významně odlišuje od normálního, například v případě značné asymetrie, je nutná jejich transformace. Často se provádí tzv. lognormální transformace. Důležitou charakteristikou je hodnota statistického rozptylu, která se v případě stacionárního pole rovná úrovni prahu. Pokud je v měřených hodnotách obsažen trend, je vhodné tento odstranit a provádět strukturální analýzu a krigování na hodnotách reziduí.