Geostatistické metody prostorové interpolace

Geostatistické metody interpolace

Stochastické metody lokální interpolace,

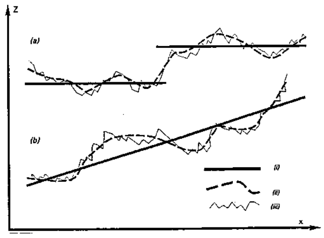
Optimalizují výběr bodů okolí, ze kterých je interpolována nová hodnota.

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

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

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

Základní komponenty spojitého povrchu



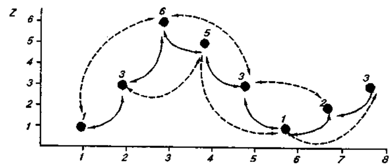
I – trendová složka – drift
ii – regionalizovaná proměnná
iii – náhodná složka

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

Regionalizovaná proměnná se odhaduje v tzv. **strukturní analýze** pomocí hodnot strukturní funkce - tzv. **semivariance** γ(h)

Strukturní analýza

Příklad výpočtu měr prostorové variability pro 1D - řadu hodnot



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

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

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

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

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

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

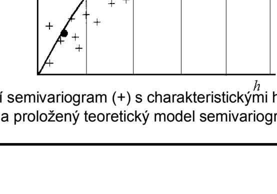
Experimentální semivariogram

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

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

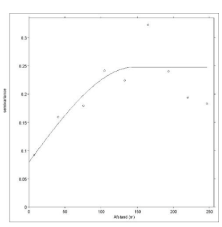
n - počet dvojic bodů pozorování proměnné s atributem *z* vzdálených o hodnotu *h*

h - tzv. lag - vzdálenost dané dvojice bodů.

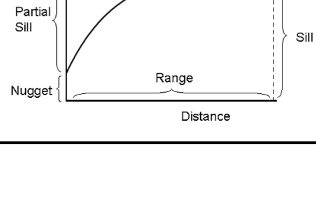


Experimentální semivariogram (+) s charakteristickými hodnotami pro vzdálenosti h (•) a proložený teoretický model semivariogramu (plná čára)

Strukturní analýza



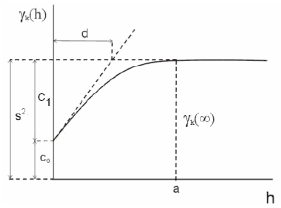
Experimentální semivariogram



Teoretický semivariogram

Strukturní analýza

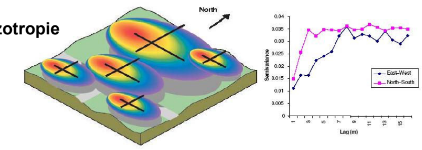
Prvky (parametry) semivariogramu



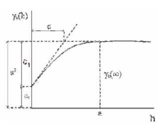
a - dosah (range), d – rozpětí, c0 - zbytkový rozptyl (nugget), c=c0 + c1 - práh (sill), h – lag (krok vzdálenosti)

Strukturní analýza

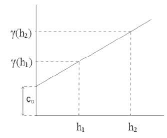
Efekt anizotropie



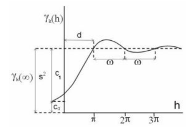
Příklady teoretických semivariogramů:



sférický



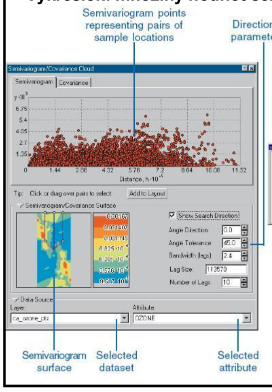
lineární

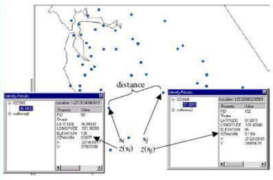


sinový

Základní nástroje průzkumové analýzy prostorových dat

Vykreslení množiny hodnot semivariance či kovariance





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

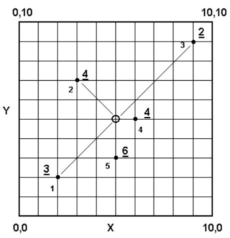
KRIGING jako metoda interpolace

Odhad hodnoty v bodě x_0

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

a odhad jeho rozptylu

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



Je konstruován tak, aby byl minimální

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

váhy, které musíme vypočítat

Hodnoty semivariancí poskytuje strukturní analýza - z vhodného teoretického modelu semivariogramu

Postup řešení:

1. Na základě předem provedené strukturní analýzy použijeme vhodnou strukturní funkci (např. sférický semivariogram) s příslušnými hodnotami parametrů
2. Řešíme soustavu rovnic, kde jednotlivé členy mají následující význam:

$$A \cdot \lambda = b$$

A – matice semivariancí mezi všemi dvojicemi bodů
b – vektor semivariancí mezi všemi body a bodem predikovaným
λ – vektor vah jednotlivých bodů
Φ – tzv. Lagrangeův člen

3. Soustavu rovnic řešíme pro váhy λ, tak, aby byla splněna podmínka

$$\sum \lambda = 1$$ (proto je v soustavě použit člen *Φ)*

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

Postup řešení:

4. K určení hodnot semivariancí je zapotřebí vytvořit matici vzdáleností mezi datovými body a vektor vzdáleností mezi měřenými body a bodem predikovaným
5. Řešením soustavy rovnic získáme hodnoty vah λ a hodnotu *Φ*
6. Vypočteme predikovanou hodnotu v bodě (i=0):

$$\hat{z}(x_0) = \sum_{i=1}^{n} \lambda_i \cdot z(x_i)$$ *$Z(x_{i=0})$* ***=4,560***

7. Vypočteme rozptyl odhadu:

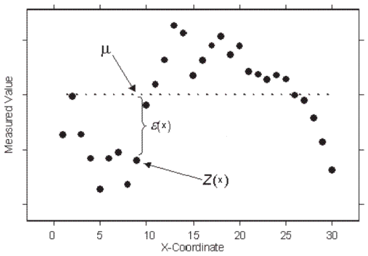
$$\hat{\sigma}_e^2 = \sum_{i=1}^{n} \lambda_i \gamma(x_i, x_0) + \phi$$ *σ_e^2* ***= 4,008***

Typy krigování

Základní krigování (ordinary kriging)

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

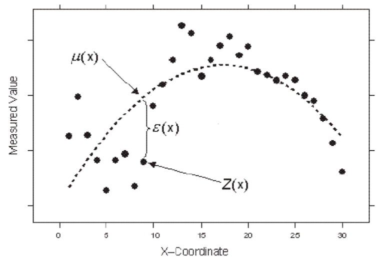
kde μ je neznámá hodnota trendu.



Univerzální krigování (Universal kriging)

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

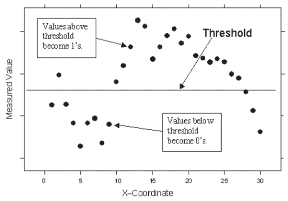
kde $\mu(x)$ je deterministická funkce

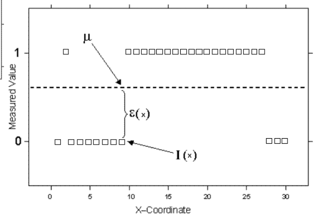


Indikátorové krigování

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

kde μ neznámá konstanta, *ε(x)* autokorelovaná náhodná proměnná a *I(x)* je binární proměnná, která nabývá hodnot 0 nebo 1.





Hodnocení a verifikace modelů

Krigování jako interpolační metoda umožňuje pro každý interpolovaný bod odhadnout potenciální velikost chyby odhadu.

Mapy druhé odmocniny σ_e^2

tzv. **směrodatné chyby** (odchylky) **krigingu (Standard error map)**,

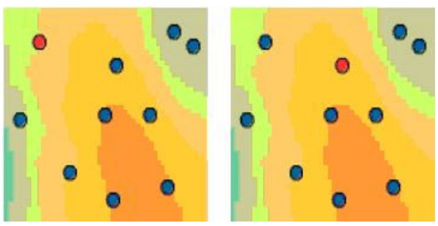
Mají-li chyby predikce normální rozdělení, potom 95% **interval spolehlivosti predikovaných hodnot** lze určit z následujícího vztahu:

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

kde *Z(x₀)* je odhad hodnoty proměnné *z* v bodě x_0 a σ_e^2 je rozptyl odhadu.

Při opakovaném použití stejného modelu padne 95 % odhadovaných hodnot do uvedeného intervalu

Křížová validace modelu



Jednotlivé body měření (červené) jsou po jednom postupně vynechány ze vstupní množiny dat

Ze zbývajících (modrých) je vypočtena hodnota v místě vynechaného bodu.

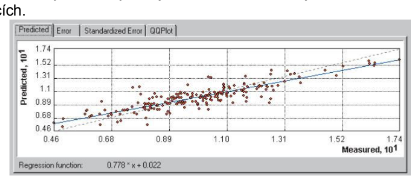
Následně jsou porovnány pozorované a vypočtené hodnoty

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.



Korelační pole měřených a predikovaných hodnot

Obecnou vlastností krigingu jako interpolační metody je podhodnocení vysokých hodnot a naopak nadhodnocení hodnot nízkých.

Sumární statistika rozdílů pozorovaných a vypočtených hodnot

Rozdíly pozorovaných a vypočtených hodnot lze hodnotit dále uvedenými měrami:

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 od teoretických. Slouží k porovnání několika různých modelů. Čím menší je RMSPE, tím vhodnější je model (tím bližší jsou vypočtené hodnoty hodnotám měřeným).

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

Základní kriging