

Geostatistika v R projektu

18. studentská konference o matematice

Martin Dzurov Kristýna Kitzbergerová Lucie Šindelářová

České vysoké učení technické v Praze
Fakulta stavební
Katedra matematiky



16. – 18. červen 2010

Obsah

1 Úvod

- Cíl projektu
- projekt R
- Balíček geoR

2 Vizualizace dat

3 Variogramy

- Algoritmus výpočtu variogramu
- Grafické znázornění variogramů
- Kovarianční funkce
- Modely variogramů
- Odhady parametrů variogramů

4 Ordinary Kriging

- Teoretický úvod do problematiky
- Vlastní zpracování dat

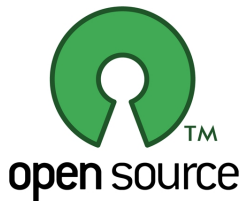
Úvod

Cíl projektu

Cílem projektu je vytvoření učebního textu, který má přiblížit problematiku geostatistiky, především význam interpolační metody kriging.

- modelová data: **nadmořské výšky pro ČR** v měřítku 1 : 500 000
- zpracováno v prostředí *R*
- použit balíček *geoR*

projekt R



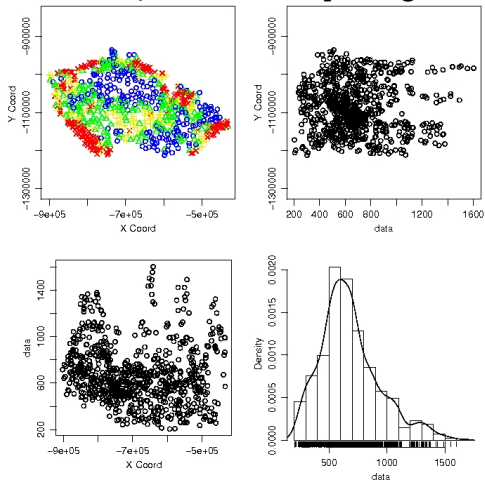
- opensource software dostupný pro platformy UNIX, Windows a MacOS
- efektivní nástroj pro manipulaci a zpracování dat
- poskytuje nástroje pro grafické zobrazení dat
- obsahuje řadu naimplementovaných funkcí
- možnost rozšíření pomocí balíčků

Balíček geoR

- poskytuje funkce pro geostatistickou analýzu dat v prostředí R
- data musí být uložena v objektu typu **geodata**:
 - objekt musí obsahovat alespoň jednu souřadnici a hodnotu atributu
 - objekt se vytvoří pomocí funkce `read.geodata()`, která z textového souboru načte potřebná data
- vytváří grafické interpretace dat (histogramy, korelační diagramy, variogramy, ...)
- poskytuje interpolační metody

Vizualizace dat

Vizualizace dat pomocí funkce `plot.geodata()`



Variogramy

- **Variogram** je základním geostatistickým nástrojem pro vizualizaci, modelování a prostorovou korelaci naměřených dat.
- v balíčku geoR jsou implementovány dva typy odhadů experimentálních variogramů

1 Klasický odhad

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

2 Hawkins-Cressieův robustní odhad

$$\gamma(h) = \frac{\left[\frac{1}{n} \sum_{i=1}^{\infty} [z(x_i) - z(x_i + h)]^2 \right]^{\frac{1}{4}}}{0.914 + \frac{0.998}{n}} \quad (2)$$

- **Variance** poté vyjadřuje míru rozptylu při změně vzdálenosti h z bodu x do bodu $x + h$

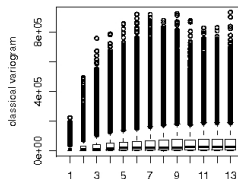
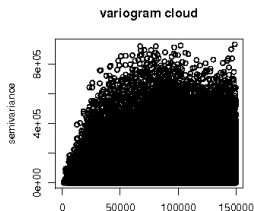
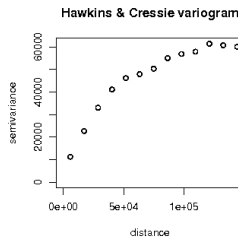
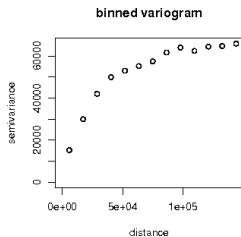
Algoritmus výpočtu variogramu

Algoritmus pro výpočet variogramu funguje zjednodušeně následovně:

- 1 nejdříve vytvoří tzv. **mračna bodů** (variogram cloud)
- 2 body v mračnu uspořádá do **tříd** (bins) podle vzdáleností
- 3 v každé této třídě vypočítá **průměr**
- 4 posledním krokem je **grafické vykreslení** závislosti hodnoty variogramu γ na vzdálenosti mezi body

Grafické znázornění variogramů

- pomocí funkce `variog()` vypočteme parametry variogramu, které poté graficky zobrazíme:



Kovarianční funkce

- kovarianční funkce je definována následovně:

1 $E[z(x)] = m, \quad \text{pro všechna } x \in \mathcal{D}$

2 $E[z(x) \cdot z(x+h)] - m^2 = C(h) \quad \text{pro všechna } x, x+h \in \mathcal{D}$

- vlastnosti kovarianční funkce:

1 $C(0) = \text{Var}(z(x))$

2 $C(h) \leq C(0)$

3 $C(h) = C(-h)$

- vztahy mezi variogramem a kovarianční funkcí je následující:

$$\gamma(h) = C(0) - C(h) \quad (3)$$

Modely variogramů

- experimentální variogramy jsou nahrazeny teoretickými
- 4 základní modely

1 Nugget-effect model – modeluje nespojitost na počátku

2 Exponenciální model – klesá exponenciálně s rostoucí vzdáleností bodů a je definován:

$$C_{exp}(h) = \sigma^2 \exp\left(-\frac{|h|}{R}\right), \quad \sigma, R > 0 \quad (4)$$

σ je tzv. *sill* (neboli výška variogramu)

R je tzv. *range* a vyjadřuje v jak velkém okolí se naměřené hodnoty ještě ovlivňují

3 Sférický model

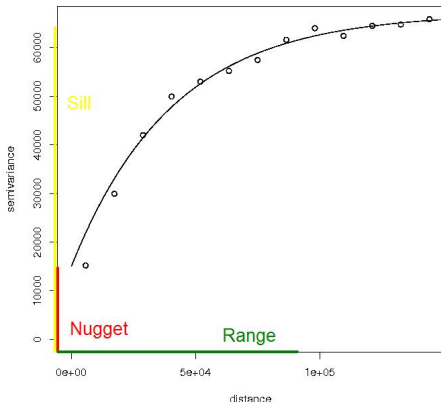
$$C_{sph}(h) = \begin{cases} \sigma^2 \left(1 - \frac{3|h|}{2R} + \frac{|h|^3}{2R^3}\right), & 0 \leq |h| \leq R \\ 0, & |h| > R \end{cases} \quad (5)$$

4 Gaussův model

$$C_{gauss}(h) = \sigma^2 \exp\left(-\frac{3|h|^2}{R^2}\right) \quad (6)$$

Odhady parametrů variogramů

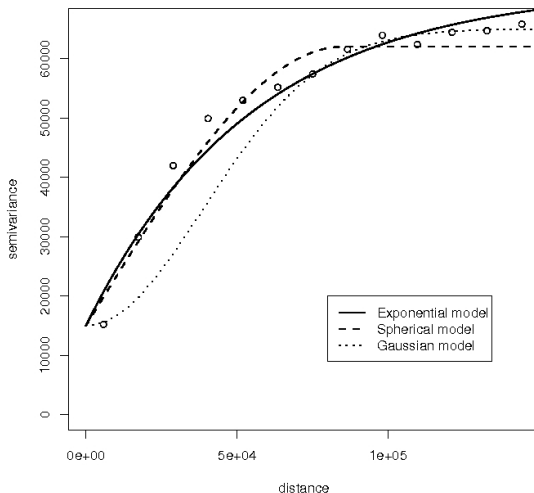
Nugget, Range and Sill



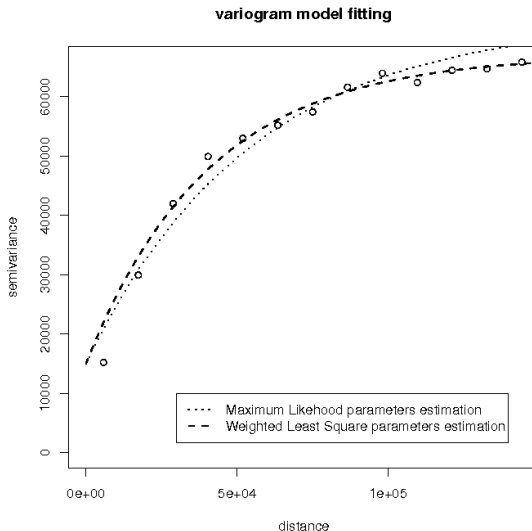
- **okem** – vyzkoušení různých modelů okem odhadnutými hodnotami parametrů *nugget*, *sill* a *range*
- **metodou nejmenších čtverců** – odhad parametrů pomocí implementované funkce `variofit()` s volbami WLS (weighted least squares) a zvoleného modelu variogramu
- **metodou maximální věrohodnosti** – pomocí funkce `likfit()` s volbou ML

Parametry odhadnuté okem

pro exponenciální, sférický a gaussův model variogramu



Parametry odhadnuté metodami WLS a ML pro exponenciální model variogramu



Interpolační metoda Ordinary Kriging

- základní geostatistická metoda pro určování lokálního odhadu
- BLUP – nejlepší lineární nestranný prediktor
 - $\hat{h}_0 = \sum_{i=1}^n w_i h_i + w_0$ je lineární prediktor
 - $\hat{h}_0$ je nestranný prediktor, pokud platí $E \{ \hat{h}_0 - h_0 \} = 0$
 - $\hat{h}_0$ je nejlepší, pokud platí $E \{ \| \hat{h}_0 - h_0 \|^2 \}$ je minimální
- **Nastínění problematiky:**
 - chceme odhadnout hodnotu nadmořské výšky $\hat{z}$ v místě x_0 , kde nemáme provedeno měření
 - použijeme lineární kombinaci hodnot ze sousedních bodů a vah, které jsou těmto bodům přiděleny

$$\hat{z}(x_0) = \sum_{i=1}^n w_i z(x_i), \quad \sum_{i=1}^n w_i = 1 \quad (7)$$

■ odhad variance:

$$E(\|\hat{z}_0 - z_0\|^2) = \sigma^2 + \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij} - 2 \sum_{i=1}^n w_i C_{i0} \quad (8)$$

$$kde C_{ij} = cov(z_i, z_j) = cov(\|x_i, x_j\|) \quad (9)$$

■ minimalizací odhadu variance za podmínky $\sum_{i=1}^n w_i - 1 = 0$ získáme systém rovnic pro Ordinary Kriging

$$\underbrace{\begin{pmatrix} C_{11} & \cdots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \cdots & C_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}}_{C_n} \underbrace{\begin{pmatrix} w_1 \\ \vdots \\ w_n \\ \lambda \end{pmatrix}}_{w_n} = \underbrace{\begin{pmatrix} C_{10} \\ \vdots \\ C_{n0} \\ 1 \end{pmatrix}}_{d_n} \quad (10)$$

■ řešením rovnic získáme neznámou

$$w_n = C_n^{-1} d_n \quad (11)$$

■ odhad variance interpolační metody Ordinary Kriging

$$Var(\hat{z}_0 - z_0) = C_{00} - \sum_{i=1}^n w_i C_{i0} - \lambda \quad (12)$$

Vlastnosti

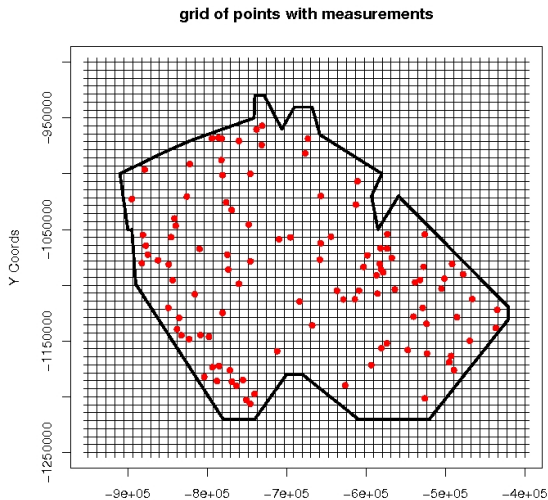
- **přesný interpolátor** – pokud je hodnota pro výpočet odhadu x_0 identická s místem, kde bylo provedeno měření, bude vyinterpolovaná hodnota totožná s hodnotou naměřenou

$$\hat{z}(x_0) = z(x_\alpha), \quad x_0 = x_\alpha \quad (13)$$

- používá se pro interpolaci dat na pravidelné mřížce pomocí nerovnoměrně rozmístěných dat měření
- balíček geoR obsahuje naimplementovanou funkci `krige.conv()`

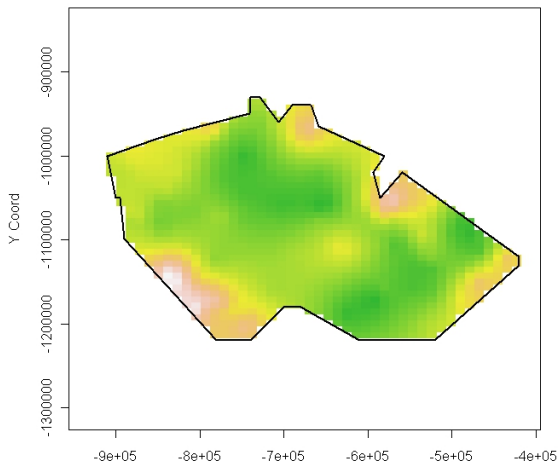
Vlastní zpracování dat

- pravidelná mřížka, na které bude proveden výpočet



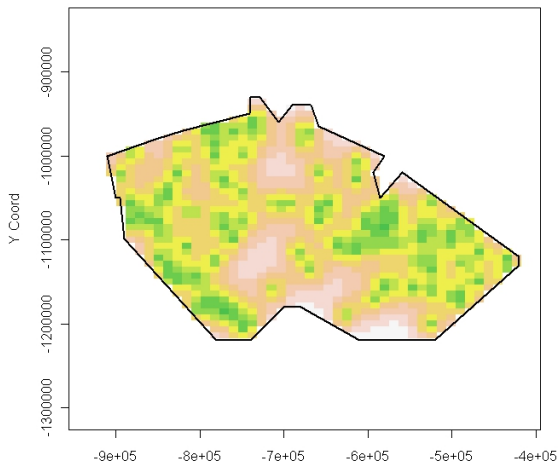
Výsledky

- predikce nadmořských výšek metodou OK



Výsledky

- směrodatná odchylka odhadnutých nadmořských výšek



Děkujeme za pozornost!

Tato práce byla podpořena grantem Studentské grantové soutěže ČVUT
č.OHK1-065/10.