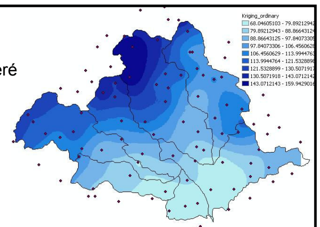
KRIGING
geostatistické metody interpolace

Kriging



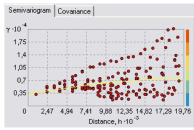
Skupina interpolačních metod, které optimalizují výběr bodů okolí, ze kterých je odhadována nová hodnota podle obecného modelu:

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

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.



Vlastnosti krigování

Metoda krigování se často označuje akronymem **BLUE** (Best Linear Unbiased Estimator – tedy nejlepší lineární nezkreslený odhad).

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

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

Odhad je **nejlepší**, protože je minimalizován rozptyl odhadu

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

Vlastnosti krigování

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 byl nestranný a odhad rozptylu byl menší, než jakákoliv jiná lineární kombinace pozorovaných hodnot (minimální).

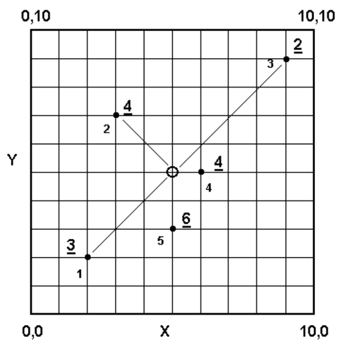
Přitom pro minimální rozptyl hodnot $\hat{z}(x_0)$ platí výraz :

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

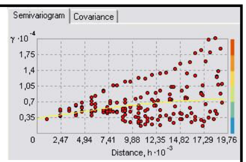
$\gamma(x_i, x_0)$ semivariance proměnné *z* mezi body x_i a x_0

ϕ Lagrangeův multiplikátor, který zajišťuje požadavek minimalizace odchylek a zároveň podmínku, že suma vah je rovna jedné.

PŘÍKLAD: Výpočet neznámé hodnoty v bodě metodou základního krigingu.



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:

$$A \cdot \lambda = b$$

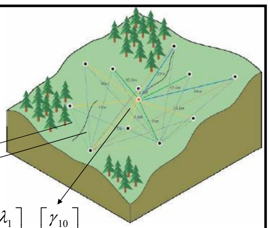
A – matice semivariancí mezi všemi dvojicemi bodů
b – vektor semivariancí mezi všemi body a bodem predikovaným
λ – vektor vah jednotlivých bodů
Φ – tzv. Lagrangeův člen

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:



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

Postup řešení:

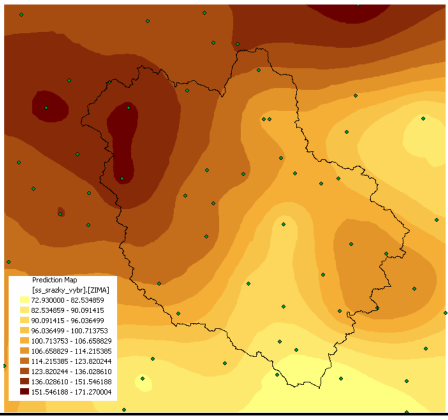
4. K určení hodnot semivariancí je zapotřebí vytvořit matici vzdáleností mezi datovými body a vektor vzdáleností mezi měřenými body a bodem predikovaným
5. Řešením soustavy rovnic získáme hodnoty vah λ a hodnotu Φ
6. Vypočteme predikovanou hodnotu v bodě (i=0):

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

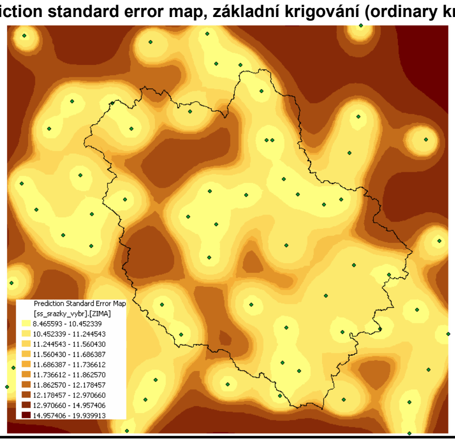
7. Vypočteme rozptyl odhadu:

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

Prediction map, základní krigování (ordinary kriging)



Prediction standard error map, základní krigování (ordinary kriging)



Typy krigování

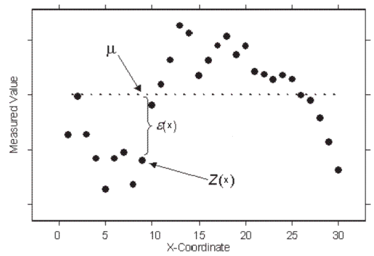
- krigování s bodovým odhadem
- krigování s blokovým odhadem

- základní (ordinary) krigování
- jednoduché (simple) krigování
- univerzální krigování
- indikátorové krigování
- pravděpodobnostní krigování
- co-kriging
- lognormální krigování.

Základní krigování (ordinary kriging)

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

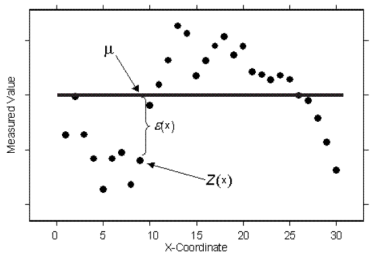
kde μ je neznámá hodnota trendu.



Jednoduché krigování (Simple kriging)

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

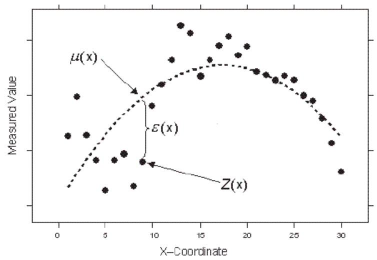
kde μ je známá konstanta.



Univerzální krigování (Universal kriging)

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

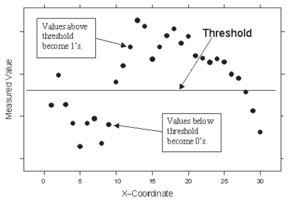
kde $\mu(x)$ je deterministická funkce



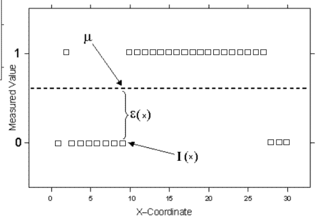
Indikátorové krigování

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

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



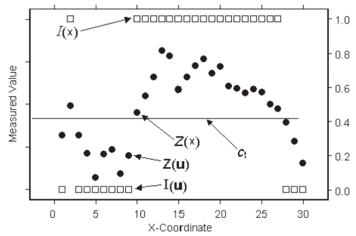
Výsledkem odhadu je pravděpodobnost překročení určité prahové hodnoty



Pravděpodobnostní krigování (probalility kriging)

$$I(x) = I(Z(x) > c_1) = \mu_1 + \varepsilon_1(x) \qquad Z(x) = \mu_2 + \varepsilon_2(x)$$

kde μ_1 a μ_2 jsou neznámé konstanty. *I(x)* je binární proměnná vytvořená indikátorovým prahováním (*I(Z(x) >* c_1) a $\varepsilon_1(x)$ a $\varepsilon_2(x)$ jsou dvě náhodné chyby.

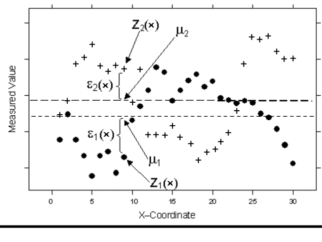


Co-kriging

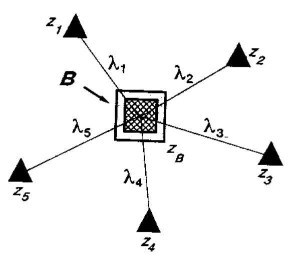
Proměnné z_1 a z_2 vykazují prostorovou korelaci.
Proměnná z_2 je snáze získatelná
Hodnot proměnné z_2 je využito k interpolaci hodnot proměnné z_1.
Základní co-kriging využívá následujících modelů:

$$Z_1(x) = \mu_1 + \varepsilon_1(x)$$
$$Z_2(x) = \mu_2 + \varepsilon_2(x)$$

kde μ_1 a μ_2 jsou neznámé konstanty a $\varepsilon_1(x)$ a $\varepsilon_2(x)$ dvě náhodné chyby.



Blokový odhad při základním krigování



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)**,

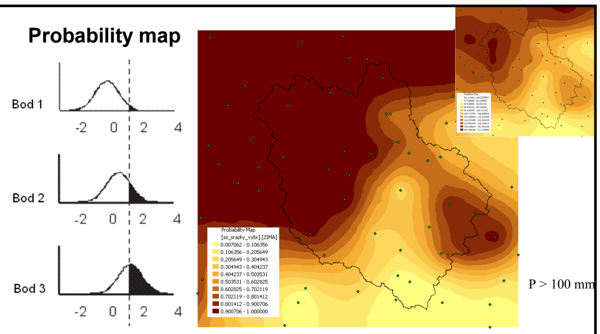
Mají-li chyby predikce normální rozdělení, potom 95% **interval spolehlivosti predikovaných hodnot** lze určit z následujícího vztahu:

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

kde $Z(x_0)$ 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

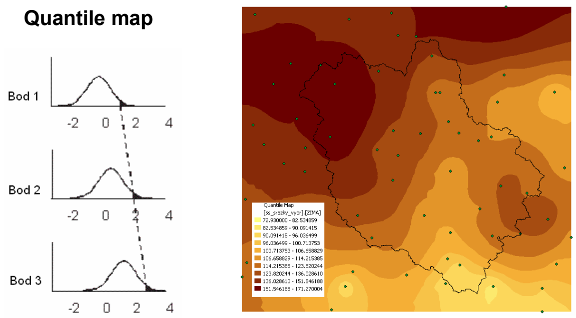
Probability map



Jaká je pravděpodobnost, že predikovaná hodnota bude v daném bodě větší než prahová hodnota (např. 1 na obr. vlevo)?

Při konstantní prahové hodnotě se její pravděpodobnost výskytu pro jednotlivé body mění – tedy lze z ní vytvořit mapu pravděpodobností (**probability map**).

Quantile map

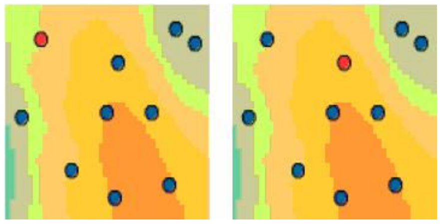


P = 90%

Jaká je v daném bodě hodnota překročená se zadanou pravděpodobností?

Při konstantní pravděpodobnosti se budou měnit hodnoty kvantilů a lze je opět prezentovat ve formě tzv. kvantilové mapy (**quantile map**).

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.

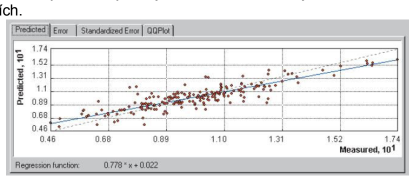
Následně jsou porovnány pozorované a vypočtené hodnoty

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.

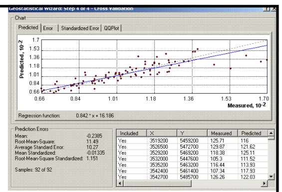
Pokud sestavený model vyhovuje trénovacím datům, je ověřen na datech testovacích.



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 předikovaných 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).

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

Porovnání metod interpolace

a) Thiessenovy polygony

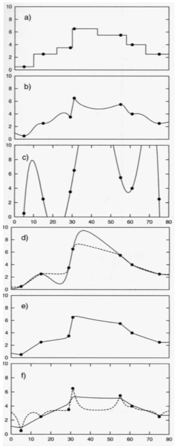
b) IDW

c) Globální polynom 6. řádu)

d) Spline s různými parametry (tension)

e) Kriging (nulový zbytkový rozptyl, „dlouhý“ dosah

f) Kriging velký zbytkový rozptyl, dlouhý dosah resp. nulový zbytkový rozptyl, velmi krátky dosah



Porovnání metod interpolace (Burrough, 1986)

Method	Deterministic/ stochastic	Local/ global	Transition abrupt/ gradual	Exact interpolator	Limitations of the procedure	Best for	Output data structure	Computing load	Assumptions of interpolation model
'Eyeball'	Subjective/ deterministic	Global	Abrupt	No	Non-reproductible, Subjective	Field data, aerial photo interpretation	Polygons	None	Intuitive understanding of spatial processes; homogeneity within boundaries
Edge-finding algorithm	Deterministic	Global	Abrupt	No	Often requires shapes to be defined and stored; better for man-made features than for natural landscapes	Raster images from remote sensors	Raster	Moderate	Homogeneity within boundaries
Proximal (Theissen poly.)	Deterministic	Local	Abrupt	Yes	One data point per cell; no error estimates possible; tesselation pattern depends on data point distribution	Normal data from point patterns	Polygons	Light/ Moderate	'Nearest neighbour' gives best information
Fourier series	Stochastic	Global	Gradual	No	Edge effects, outliers, complex polynomials do not necessarily have meaning; errors are rarely spatially independant	Demonstrating broad features and removing them prior to other methods of interpolation	Points on a raster	Light/ Moderate	Multiple regression-phenomenological explanation of trend surface; independent Gaussian errors
Trend surface	Stochastic	Global	Gradual	No	Not applicable to data; lacking periodicity	Periodic features such as sand dunes, ripple marks or gilgai, or man-made features	Points on a raster	Moderate	Strict periodicity in phenomenon of interest
B-splines	Deterministic	Local	Gradual	Yes	No estimates of errors; masks all uncertainties in surface	Very smooth surfaces	Points on a raster	Light/ Moderate	Absolute smoothness of variation
Moving average	Deterministic	Local	Gradual	No unless constrained	Results depend on configuration of data points and size of window; sample versions assume isotropy; no error estimates unless retrospectively calculated	Quick contour plots of moderately smooth data	Points on a raster	Moderate	Continuous, differentiable surface is appropriate
Optimal interpolation (kriging)	Stochastic	Local	Gradual	Yes	Practical and theoretical problems of non-stationarity in data; large computing costs for mapping	Situations where the most detailed estimates and their errors are required	Points on a raster	Heavy (very heavy for universal kriging	Intrinsic hypothersis (homogeneity of first differences) average local values can be represented by a continuous surface