

Geostatistický slovníček pro oblast modelování enviromentálních dat

Janoška, Z., Pechanec, V., Tuček, P.

Katedra geoinformatiky, PřF, Univerzita Palackého v Olomouci <http://www.geoinformatics.upol.cz>

anisotropy (anizotropie) – vlastnost prostorových procesů nebo dat, kdy se prostorová závislost mění jak se vzdáleností, tak se směrem mezi lokacemi (srov. *isotrophy*)

autocorrelation (autokorelace) – statistická korelace mezi prostorově náhodnými proměnnými stejného typu, vlastností apod., kde korelace závisí na vzdálenosti a/nebo směru mezi lokacemi (srov. *crosscorrelation*)

binning – proces, kdy jsou dvojice bodů shlukovány do skupin na základě jejich vzdálenosti od sebe. Každému binu je přiřazena hodnota vzdálenosti a semivariance, které odpovídají průměru všech bodů spadajících do daného binu. Běžně se používají dvě metody binningu – formované buďto čtvercovou, nebo sektorovou metodou (Geostatistical Analyst pracuje s čtvercovou metodou).

cokriging – statistická interpolační metoda, která používá data různých datových typů (různé atributy) k předpovědění hodnoty primárního datového typu (primárního atributu); výsledkem jsou rovněž směrodatné chyby předpovědi – míra toho, jak přesná je předpověď (srov. *kriging*)

correlation (korelace) – kovariance odstupňovaná tak, že hodnota -1 znamená žádnou závislost a hodnota 1 naprostou závislost dvou vstupních datových sad (srov. *covariance*)

covariance (kovariance) – statistická tendence proměnných stejných typů, vlastností apod., kdy hodnoty těchto proměnných vykazují závislost vůči sobě. Pozitivní covariance nastává, když obě z proměnných jsou nad jejich očekávaným průměrem, negativní covariance tehdy, když jedna z proměnných je nad očekávanou, druhá pod očekávanou hodnotou (srov. *crossvariance*)

crosscorrelation – statistická korelace mezi prostorově náhodnými proměnnými různého typu, vlastností apod., kde korelace závisí na vzdálenosti a/nebo směru mezi lokacemi (srov. *autocorrelation*)

crossvariance – statistická tendence proměnných různých typů, vlastností apod., kdy hodnoty těchto proměnných vykazují závislost vůči sobě. Pozitivní Crossvariance nastává, když obě z proměnných jsou nad jejich očekávaným průměrem, negativní crossvariance tehdy, když jedna z proměnných je nad očekávanou, druhá pod očekávanou hodnotou (srov. *covariance*)

cross-validation – metoda, kdy část dat je oddělena od datasetu a zbytek dat je použit k předpovědění hodnot v místě oddělených dat. Výsledky jsou pak porovnány se skutečnými hodnotami. Full Cross-validation znamená oddělit každá jednotlivá data v datasetu a použít zbytek dat k jejich předpovědi.

crossvariogram – je funkcí vzdálenosti a směru dělicí dvě lokace a používá se ke kvantifikaci cross-correlation. Crossvariogram je definován jako rozptyl rozdílu mezi proměnnými různých typů nebo atributů na dvou lokacích. Crossvariogram většinou roste se vzdáleností a je definován parametry: nugget, sill, range.

deterministic (deterministický) – typ nebo část modelu, kde výsledek je kompletně a exaktně odvozen ze vstupních hodnot; nepočítá se s náhodnými procesy

detrending – proces odstraňování trendu. Povrch tvořící trend je odstraněn z originálních dat – výsledné hodnoty po odstranění trendu se nazývají rezidua

empirical (empirický) – ve statistickém pojetí znamená, že kvantita je odvislá od dat, pozorování či experimentů, není to model ani jeho část. Například empirický semivariogram je počítán pouze ze vstupních dat, na rozdíl od teoretického variogramu, který reprezentuje ideální stav.

global polynomial interpolation – jedna z deterministických interpolačních metod, kdy výsledná data neprocházejí přesně vstupními body a metoda neposkytuje hodnocení přesnosti měření

intrinsic stationarity (vnořená stationarita) – předpoklad, že data jsou výsledkem náhodných procesů s konstantním aritmetickým průměrem a že semivariogram je závislý pouze na vzdálenosti a směru mezi dvěma lokacemi

Inverse Distance Weighted interpolation (IDW) – jedna z deterministických interpolačních metod, kdy výsledná data procházejí přesně vstupními body a maximální a minimální hodnoty mohou být dosaženy pouze ve vstupních bodech

isotropy (izotropie) – vlastnost přírodních procesů či dat, kdy prostorová závislost (autokorelace) se mění jen se vzdáleností, směr je nepodstatný (srov. *anisotropy*)

kriging – statistická interpolační metoda, která počítá z dat jednoho typu (hodnoty jednoho atributu) k předpovědi dat stejného typu (hodnoty stejného atributu) na nezměřených lokacích. Výsledkem jsou rovněž směrodatné chyby předpovědi – míra toho, jak přesná je předpověď (srov. *cokriging*)

least-squares fit (metoda nejmenších čtverců) – model (regresní linie, povrch, funkce) prochází daty tak, že je minimalizován součet druhých mocnin odchylek od měřených dat

local polynomial interpolation – jedna z deterministických interpolačních metod, kdy výsledná data neprocházejí přesně vstupními body a metoda neposkytuje hodnocení přesnosti měření

mean stationarity – vlastnost prostorových procesů, kdy všechny prostorově neusprádané proměnné mají stejnou střední hodnotu

nugget – parametr semivariogramu, který reprezentuje náhodnou chybu, chybu měření nebo rozptyl měření v mikroměřítkách, který je příliš malý na to, aby byl zachycen. V grafu je to nespojitost, kdy ve vzdálenosti 0 je prostorová autokorelace nenulová.

outlier – naměřený jev s výrazně odlišnou hodnotou, než jakou mají ostatní měřené body. Globální outlier je hodnota, která výrazně vybočuje z intervalu ostatních měřených hodnot celého datasetu. Lokální outlier je taková hodnota, která sice zapadá do intervalu měřených hodnot, ale vzhledem k nejbližším okolním bodům je nezvykle vysoká či nízká. Outliery mohou být buďto reálnými abnormalitami, a pak jsou často nejdůležitějšími měřenými daty, nebo mohou být výsledkem chybného měření. Pokud jsou výsledkem chybného měření či záznamu, měly by být před jakoukoli další prací s daty odstraněny.

polynomial – funkce, která je složena ze součtu měř jednotlivých proměnných. V prostorovém kontextu, kdy x-souřadnice a y-souřadnice jsou proměnné, polynomní jsou výrazy: 1, x, y, x^2 , y^2 , x^2y , apod., kdy každý z nich je ohodnocen koeficientem b_0 až b_n .

probability map – povrch, který vyjadřuje pravděpodobnost, že měřená hodnota se vyskytuje nad/pod prahem, zadaným uživatelem

QQPlot – druh grafu, kdy kvantily dvou rozdělení jsou vyneseny relativně k sobě – slouží k porovnání rozdělení dvou datasetů

quantile (kvantil) – pro n-tý kvantil (n je mezi 0 a 1) je hodnota, která má poměr n dat nad svou hodnotou (pro n = 0,25 to bude hodnota, pod níž se bude nacházet 75 % dat)

radial basis function – jedna z deterministických interpolačních metod, kdy výsledná data procházejí přesně vstupními body a metoda neposkytuje hodnocení přesnosti měření

regression (regrese) – statistická metoda, kdy proměnná je vytvořena z funkcí jedné nebo více proměnných. Často je to funkce splňující podmínku nejmenších čtverců (tedy je to taková funkce, že celková odchylka od vstupních funkcí je minimální).

residuals (rezidua) – hodnoty získané oddělením trendu od původních vstupních dat

range – parametr semivariogramu, který určuje vzdálenost, ve které přestávají být hodnoty autokorelované (maximální vzdálenost, ve které jsou hodnoty na sobě závislé – ve větší vzdálenosti se už nevyplatí modelovat)

second-order stationarity (stacionarita druhého řádu) – předpoklad, že jsou výsledkem náhodných procesů s konstantním průměrem, a že prostorová kovariance je závislá pouze na vzdálenosti a směru dvou pozorování, nikoli na jejich poloze

semicrossvariogram – semivariogram dělený dvěma

semivariogram – variogram dělený dvěma

sill – parametr semivariogramu, který určuje hodnotu, k níž se blíží model semivariogramu při velkých vzdálenostech. Ve velkých vzdálenostech data přestávají být autokorelována, takže sill je vlastně rozptyl náhodné proměnné. Některé teoretické semivariogramy sill nemají. V Geostatistical Analyst mají všechny modely semivariogramu sill.

spatial dependence (prostorová závislost) – poznatek, že věci, které jsou si prostorově bližší, jsou si podobnější než věci vzdálené

spline interpolation – jedna z deterministických interpolačních metod, speciální případ radial basis functions, kdy výsledná data procházejí přesně vstupními body a metoda neposkytuje hodnocení přesnosti měření

stationarita – předpoklad, že všechny statistické vlastnosti atributů jevu jsou závislé pouze na relativní poloze pozorování

trend – jakýkoli povrch je tvořen dvěma hlavními komponenty – globálním trendem a náhodným rozptylem krátkého dosahu. Trend je povrch tvořený je z fixních parametrů, často polynomů souřadnic x a y . Trend zpravidla modeluje jen změny velkých měřítek, malé odchylky jsou pak výsledkem náhodných chyb. Při geostatistickém modelování bývá často trend odstraněn a modelován jen náhodný rozptyl. Důvodem je zachování předpokladu stacionarity.

variogram – funkce vzdálenosti a směru mezi dvěma lokacemi, používá se ke kvantifikování autokorelace. Je definován jako rozptyl rozdílů hodnot proměnné v párech měřených prvků na jejich prostorové vzdálenosti. Většinou narůstá se vzdáleností a je definován pomocí parametrů: nugget, sill a range.

variography (variografie) – proces odhadování teoretického semivariogramu. Skládá se z průzkumu vstupních dat, výpočtu empirického semivariogramu, binningu, přiřazení modelu k semivariogramu a ohodnocení výsledného modelu.

Text vznikl v rámci bakalářské práce Z. Janošky: Faktorová analýza výskytu vybraných botanických a zoologických taxonů, jež byla řešena v rámci projektu VaV MŽP SP/2d3/54/07 Syntéza poznatků o stavu biodiverzity travních porostů v CHKO Bílé Karpaty s cílem vytvoření metodiky pro zachování biodiverzity tohoto ekosystému.