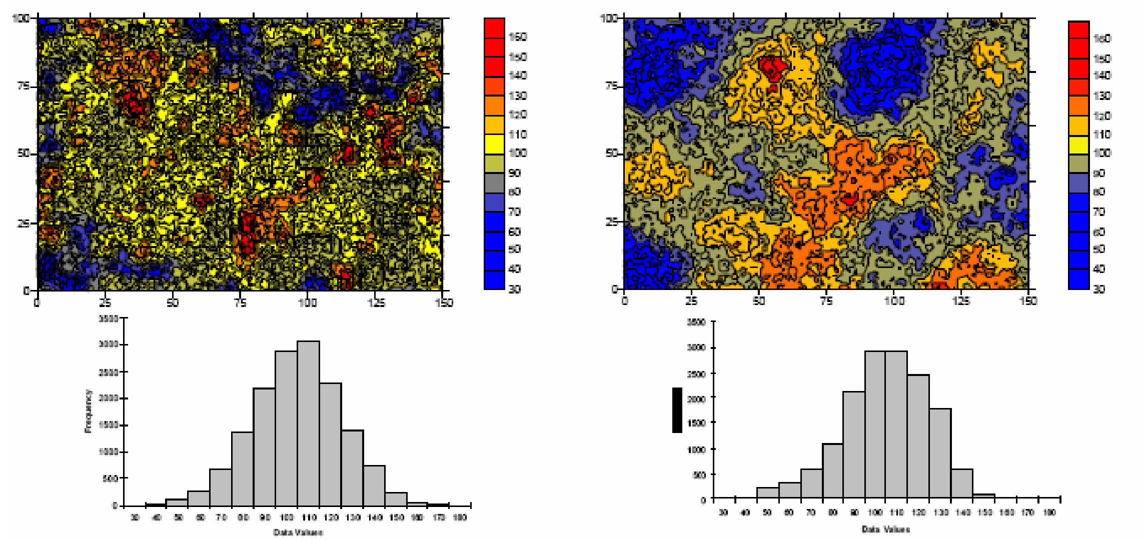
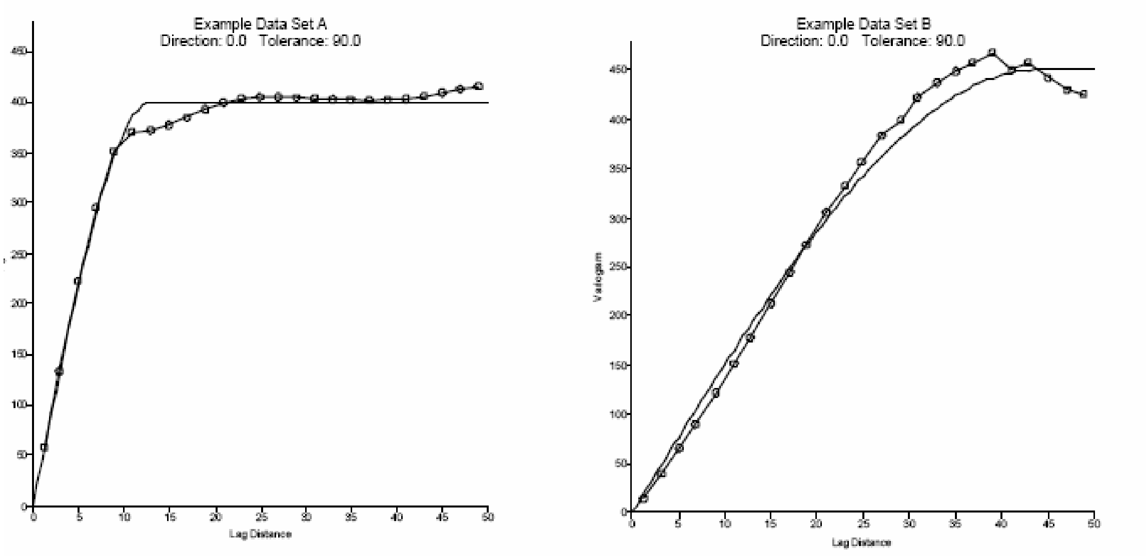
Geostatistika – vymezení pojmu

Geostatistika v užším slova smyslu – skupina interpolačních algoritmů založených na metodě krigingu. V širším slova smyslu – statistická analýza prostorově lokalizovaných dat.

Pomocí „klasických“ statistických metod lze vhodně analyzovat především atributová data – jejich kvantitativní či kvalitativní vlastnosti. Velmi omezeně však jimi lze charakterizovat prostorové vlastnosti objektů a jevů. Tyto prostorové vlastnosti jako např. spojitost jevů, prostorovou autokorelaci, prostorové uspořádání (strukturu) lze charakterizovat právě pomocí geostatistických metod.



	A	B
Count	15251	15251
Average	100.00	100.00
Standard Deviation	20.00	20.00
Median	100.35	100.92
10 Percentile	73.89	73.95
90 Percentile	125.61	124.72



Obr. 1.1. Prezentace prostorového rozšíření spojitého jevu metodami popisné statistiky a pomocí tzv. semivariogramu.

Na obrázku jsou znázorněny dva příklady zcela rozdílného prostorového rozšíření jistého spojitého jevu – např. koncentrace znečištění území jistou látkou. Z níže uvedené tabulky základních popisných charakteristik i histogramů nelze zjistit žádný podstatný rozdíl v prostorovém uspořádání studovaného jevu v obou porovnávaných mapách. Ten je však patrný pokud prostorové rozšíření charakterizujeme pomocí tzv. semivariogramu, který patří k základním nástrojům strukturní analýzy a geostatistických metod.

Geostatistika představuje především:

- Statistický popis prostorově lokalizovaných dat (geografických objektů)
- Statistický popis prostorového uspořádání objektů (bodů, linií, ploch)
- Koncept prostorové autokorelace
- Strukturní analýzu a popis prostorové autokorelace strukturními funkcemi
- Konstrukci spojitých polí metodami krigingu
- Objektivní metody klasifikace jevů

Statistický popis bodů

Body představují nejčastější způsob prezentace geografických jevů. Body jsou zpravidla umísťovány v těžišti objektů. Těžiště se konstruuje např. v místě křížení nejdelší a nejkratší osy objektu (zpravidla plochy). U konvexních objektů se tak může těžiště dostat i mimo vlastní objekt.

To, jaké geografické objekty lze popsat pomocí bodů (tedy stupeň abstrakce) závisí na měřítku, ale také na druhu analýzy (pro modelování optimálního spojení v síti sídel je vhodné je prezentovat centroidem, který tvoří uzel sítě).

Popisná statistika bodových objektů

1. Charakteristiky polohy
2. Charakteristiky rozptylu
3. Charakteristiky asymetrie
4. Charakteristiky špičatosti

Popisují distribuci bodů pomocí základních statistických charakteristik. Používají se k porovnání více bodových vzorků nebo ke sledování jejich vývoje v čase. Jejich výpočet často předchází použití geostatistických metod. Umožňuje totiž ověřit některé vlastnosti studovaných souborů, které jsou pro aplikaci metod geostatistiky nezbytné. Jedná se o ověření takových vlastností jako je normalita rozdělení, stacionarita, linearita vztahu dvou veličin apod.

Charakteristiky polohy

Charakteristiky polohy slouží k určování geografického středu či mediánu.

Průměrný střed (*mean centre*)

Průměrný střed leží na průměru souřadnic X a Y. Má stejné nevýhody jako aritmetický průměr – je to především citlivost na extrémní hodnoty. Například v případě shlukového uspořádání bodů průměrný střed dobře nereprezentuje množinu bodů.

$$(\bar{x}_{mc}, \bar{y}_{mc}) = \left(\frac{\sum_{i=1}^{n} x_i}{n}, \frac{\sum_{i=1}^{n} y_i}{n} \right)$$

kde $\bar{x}_{mc}, \bar{y}_{mc}$ jsou souřadnice průměrného středu, x_i, y_i jsou souřadnice bodu i a n je počet bodů.

Vážený průměrný střed (weighted mean centre)

Používá se v případě výskytu více událostí/objektů na stejném místě. Pak má každý bod váhu přímo úměrnou počtu událostí/objektů na tomto místě. Například při výpočtu prostorového průměru několika měst bude průměrný střed dávat realističtější představu o centrální tendenci jestliže ho budeme vážit počtem obyvatel jednotlivých měst (nebo – koncentrací znečišťující látky v jednotlivých místech či frekvencí výskytu určitého jevu).

$$(\bar{x}_{wmc}, \bar{y}_{wmc}) = \left(\frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}, \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \right)$$

kde w_i jsou váhy jednotlivých bodů.

Mediánový střed (Median Center)

Jedná se o analogii mediánu. Existuje však několik způsobů jeho definování:

A. najdeme medián na ose X a Y a vedeme z nich linie kolmé na směr osy. Takto definovaný „medián ze souřadnic“ ale nemusí odpovídat mediánu souboru bodů, protože distribuce nemusí být mezi kvadranty vyrovnaná.

B. (UK) - Mediánový střed je střed, kterým se studovaná plocha dělí do čtyř kvadrantů, z nichž každý obsahuje stejný počet bodů.

C. (US) - Mediánový střed jako střed vyžadující minimální (nejkratší) cestu. Tj. celková vzdálenost z mediánového středu do každého z bodů je minimální. Jinak řečeno – cesta z jakéhokoliv jiného místa do všech bodů oblasti bude delší než cesta z mediánového středu. Tuto podmínku lze vyjádřit vztahem:

$$\min \sum \sqrt{(x_i - u)^2 + (y_i - v)^2}$$

kde x_i a y_i jsou souřadnice jednotlivých bodů a u, v jsou souřadnice mediánového středu. Analogickým způsobem lze definovat tzv. vážený mediánový střed:

$$\min \sum f_i \sqrt{(x_i - u)^2 + (y_i - v)^2}$$

Váhy f_i pro jednotlivé body mohou být negativní či pozitivní podle toho, zda daný bod přitahuje či naopak odpuzuje polohu mediánového středu. K odvození polohy mediánového středu lze využít iteračního počtu, založeného na následujících krocích:

1. Zjistíme polohu průměrného středu jako iniciační pro hledání polohy mediánového středu. Tedy

$$(u_0, v_0) = (x_{mc}, y_{mc})$$

2. V iteračním kroku t najdeme novou polohu mediánového středu podle vztahů:

$$u_t = \frac{\sum f_i x_i \Big/ \sqrt{(x_i - u_{t-1})^2 + (y_i - v_{t-1})^2}}{\sum f_i \Big/ \sqrt{(x_i - u_{t-1})^2 + (y_i - v_{t-1})^2}}$$

$$v_t = \frac{\sum f_i y_i \Big/ \sqrt{(x_i - u_{t-1})^2 + (y_i - v_{t-1})^2}}{\sum f_i \Big/ \sqrt{(x_i - u_{t-1})^2 + (y_i - v_{t-1})^2}}$$

3. Druhý krok opakujeme do té doby, dokud vzdálenost mezi dvěma posledními polohami mediánového středu (u_t, v_t) a (u_{t-1}, v_{t-1}) je menší než vzdálenost a priori definovaná jako prahová.

Charakteristiky rozptylu

Popisují distribuci hodnot kolem měr polohy

Směrodatná vzdálenost (*standard distance*)

Je mírou rozptylu hodnot v populaci kolem průměrného středu. Na rozdíl od směrodatné odchylky se udává v jednotkách vzdálenosti. Lze ji vyjádřit z následujícího vztahu:

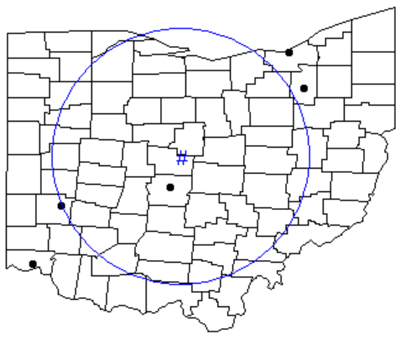
$$SD = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x_{mc})^2 + \sum_{i=1}^{n}(y_i - y_{mc})^2}{n}}$$

Vážená směrodatná vzdálenost (weighted standard distance)

Atributy jednotlivých bodů lze použít jako vah f_i k vyjádření vážené směrodatné vzdálenosti:

$$SD = \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i - x_{mc})^2 + \sum_{i=1}^{n} f_i(y_i - y_{mc})^2}{\sum_{i=1}^{n} f_i}}$$

Směrodatná vzdálenost je nejčastěji používána ve formě kružnice kolem průměrného středu (**Standard distance circle**), jejíž poloměr je právě hodnota směrodatné vzdálenosti. Různé směrodatné vzdálenosti pro různý typ jevů lze zakreslovat do stejného území. Tyto kružnice nám dávají představu o rozptylu hodnot kolem střední hodnoty pro jednotlivé typy jevů. Mohou být použity i pro studium dynamiky jevů (různé kružnice pro jeden jev v různých časových horizontech).



Obr. 1.2. Poloha váženého průměrného středu a kružnice směrodatné vzdálenosti pro pět měst ve státě Ohio. Jako váhy byl použit počte obyvatelstva

Koeficient relativního rozptylu (coefficient of relative dispersion)

Vypočte se jako poměr směrodatné vzdálenosti a poloměru kruhu se stejnou plochou jakou má studovaná oblast. Řeší výše uvedený problém použití absolutní míry směrodatné vzdálenosti. Je-li oblast různě velká (ohraničená), vznikají zavádějící hodnoty. K získání relativní míry při studiu variability obyvatelstva se někdy používá poloměr země nebo státu místo poloměru kruhu se stejnou plochou jakou má studovaná oblast.

Koeficient relativního rozptylu vypočteme:

$$CRD = 100 * \frac{SD}{A_k} = 100 * \frac{SD}{\sqrt{\frac{R}{\pi}}} = 100 * SD * \sqrt{\frac{\pi}{R}}$$

K dalším jednoduchým kritériím popisu uspořádání bodů patří např. hustota bodů v ploše (počet/plocha = n/R). Při výpočtech v relativně malých oblastech používáme euklidovskou geometrii, protože se v nich neprojeví zakřivení Země.