

GEOSTATISTIKA

Ing. Jan Popelka, Ph.D.

Fakulta životního prostředí

UJEP

(výběr z materiálu)

ZÁKLADNÍ POJMY

ŠIRŠÍ VYMEZENÍ GEOSTATISTIKY

GEOSTATISTIKA je statistická *analýza* prostorově lokalizovaných dat.

GEOSTATISTIKA je odvětví prostorové statistiky s konečným počtem měřených dat s prostorovým rozmístěním.

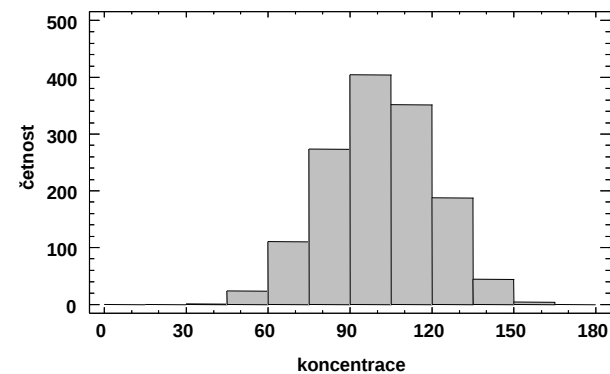
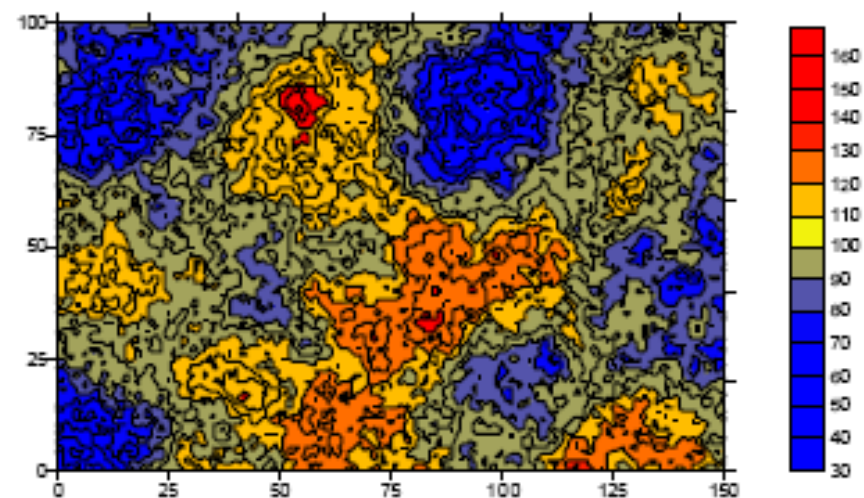
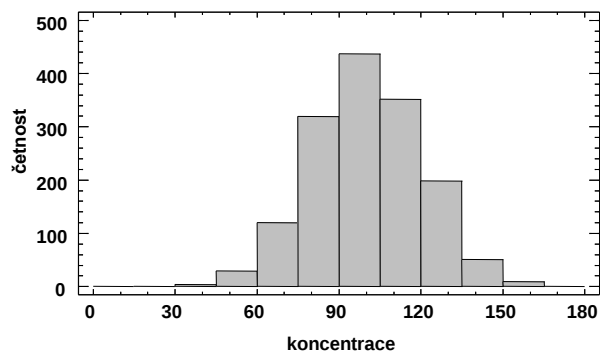
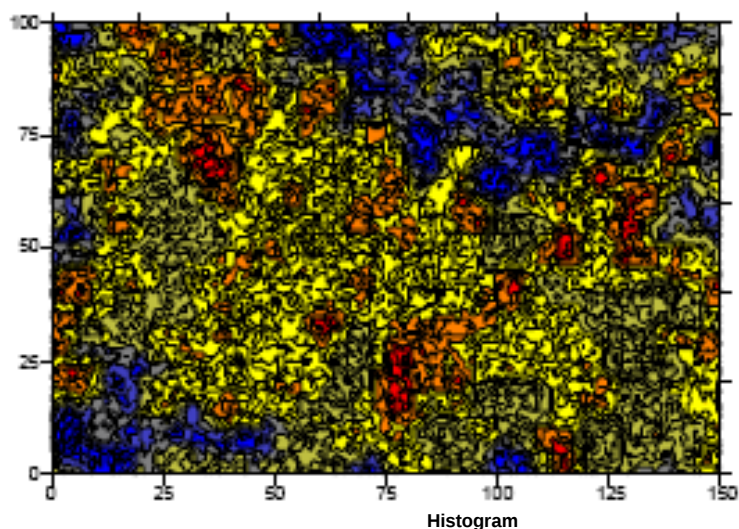
ZÁKLADNÍ POJMY

Jedno- a vícerozměrné „klasické“ statistické metody lze jen omezeně využít k popisu prostorových vlastností objektů a jevů jako jsou:

- spojitost jevů v prostoru,
- prostorová závislost – autokorelace,
- prostorové uspořádání (struktura).

ZÁKLADNÍ POJMY

Příklad: Znečištění půdy na vybraném území.



GEOSTATISTICKÝ POPIS BODŮ

Body představují nejčastější způsob prezentace geografických jevů.

Body jsou zpravidla umísťovány v těžišti objektů.

Např. místo křížení nejdelší a nejkratší osy objektu (zpravidla plochy).

Geostatistika dokáže pracovat i s liniemi (např. vodní toky) a plochami (např. územní celky).

CHARAKTERISTIKY POLOHY

Geostatistické charakteristiky polohy (úrovně)

Slouží k určování geografického středu či mediánu bodových objektů.

- průměry (průměrný střed, vážený průměrný střed, agregovaný průměrný střed, geograficky vážený průměr),
- mediánový střed.

CHARAKTERISTIKY POLOHY

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

x_i a y_i jsou souřadnice i -tého bodu, n je počet bodů.

Leží na průměru souřadnic X a Y.

Má stejné nevýhody jako „klasický“ aritmetický průměr

- citlivost na geograficky odlehlé a extrémní hodnoty,
- v případě shlukového uspořádání dobře nereprezentuje množinu bodů.

CHARAKTERISTIKY POLOHY

Mediánový střed (*median center*) $\min \sum \sqrt{(x_i - u)^2 + (y_i - v)^2}$

x_i a y_i jsou souřadnice i -tého bodu, u a v jsou souřadnice mediánového středu.

Je robustní charakteristikou polohy, není citlivý na geograficky odlehlé nebo extrémní hodnoty.

Minimalizuje celkovou vzdálenost (euklidovskou) z mediánového středu do každého z bodů. Počítá se iterativním postupem - postupné přibližování k mediánovému středu.

CHARAKTERISTIKY POLOHY

Vážený průměrný střed (*weighted mean centre*)

w_i jsou váhy i -tého bodu
(hodnota proměnné)

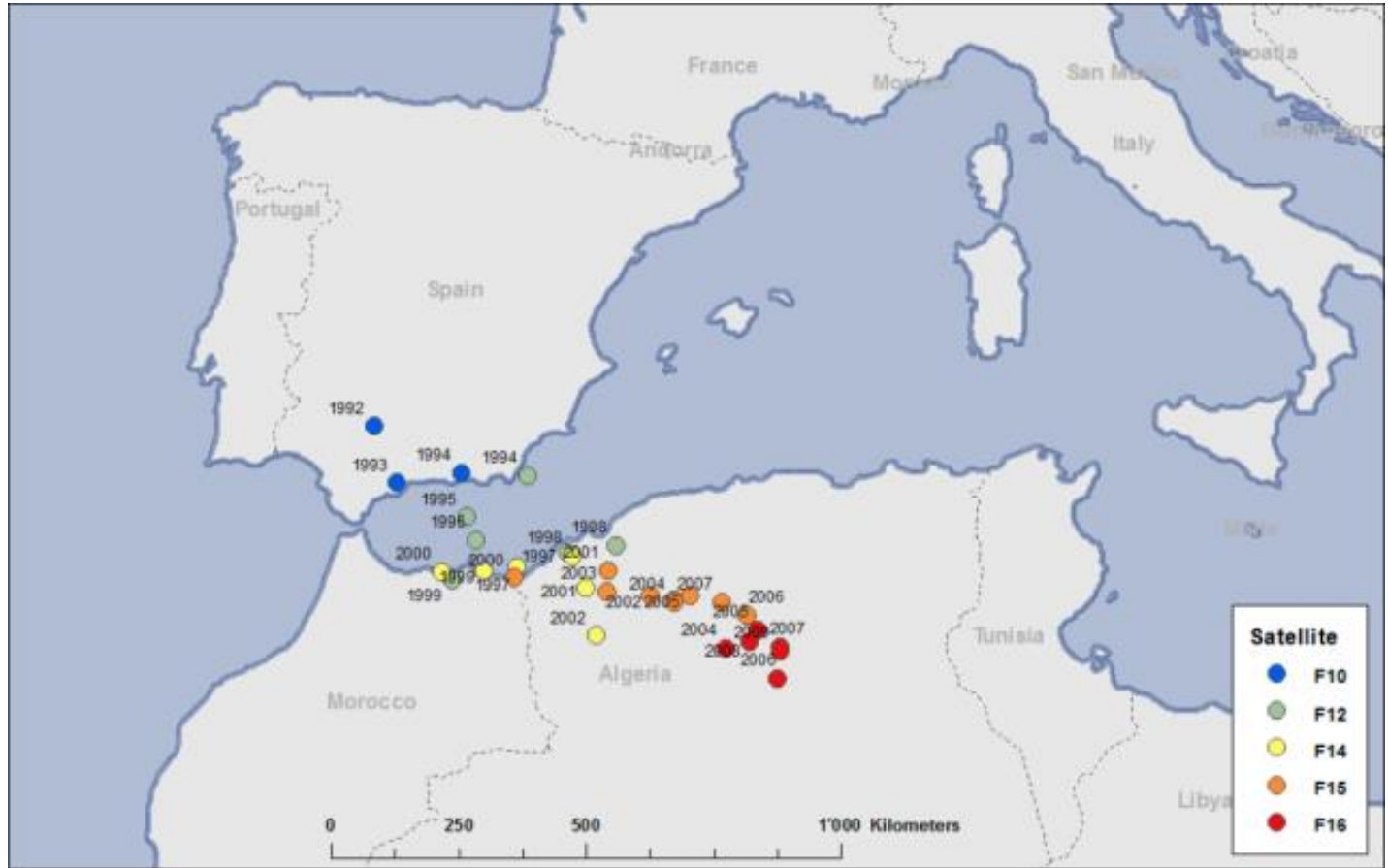
$$(\bar{x}_{wmc}, \bar{y}_{wmc}) = \left(\frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} \right)$$

Souřadnice bodů X a Y jsou váženy **hodnotou sledované proměnné W**.

Např. počet obyvatel ve městech, hodnota znečištění v místě měření apod.

CHARAKTERISTIKY POLOHY

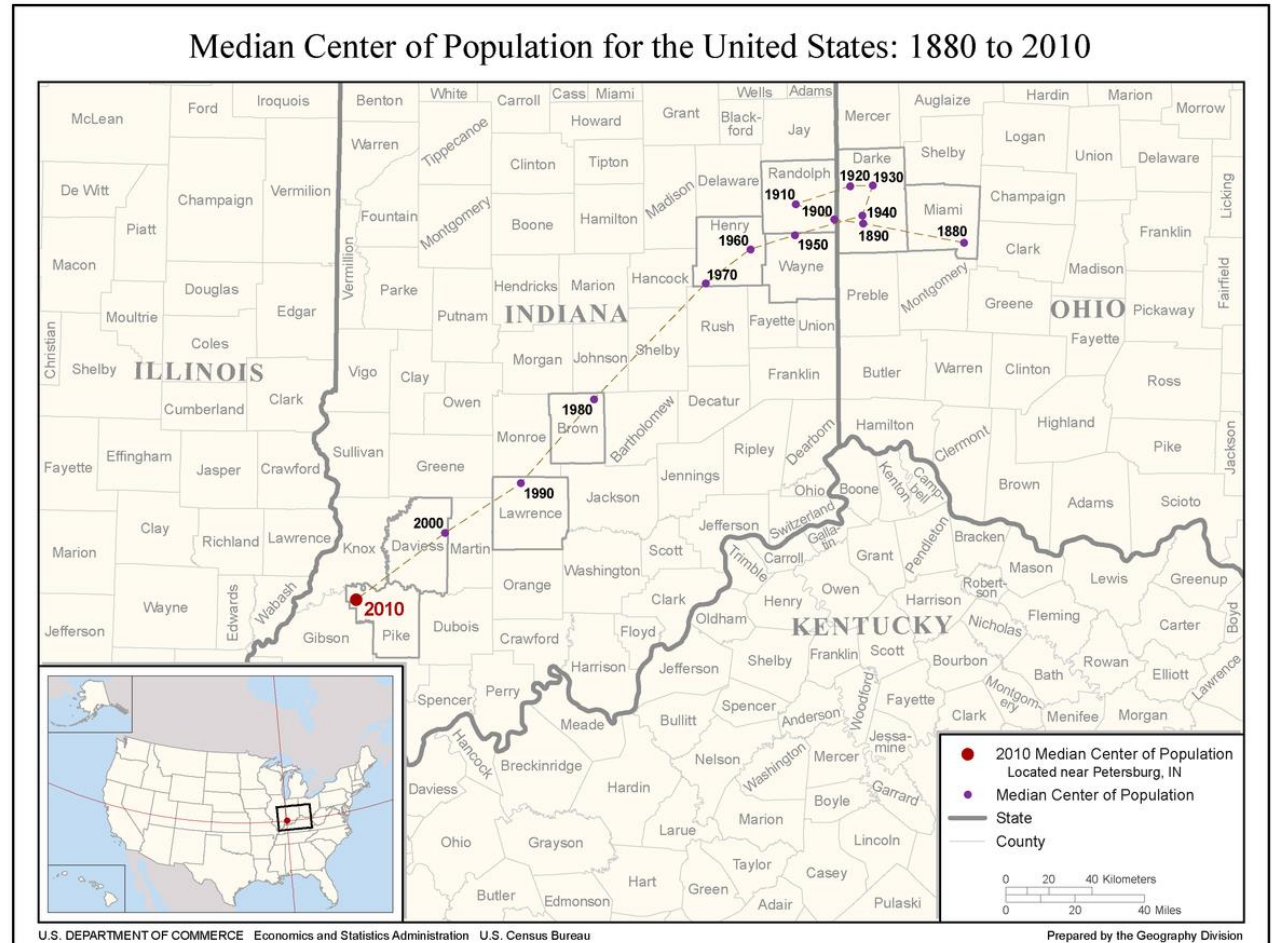
**Vážený
průměrný
střed**



Vážený průměrný střed světelných zdrojů na Zemi od roku 1992 do roku 2009

CHARAKTERISTIKY POLOHY

**Vážený
mediánový
střed**



Vážený mediánový střed populace USA od roku 1880 do roku 2010

U.S. Census Bureau. Median centers of population for the United States, 1880-2010. USA, 2011.

CHARAKTERISTIKY POLOHY

Geograficky vážený průměr (*geographically weighted mean*)

Doposud uvedené charakteristiky lze považovat za **globální** (počítají se ze všech bodů).

Lokální charakteristiky se počítají pro konkrétní místo (jejich hodnoty lze zakreslit do mapy).

Nepočítají se ze všech bodů, ale jen z vybrané množiny bodů (bodů z blízkého okolí prostřednictvím geografického vážení).

Základní princip geografického vážení:

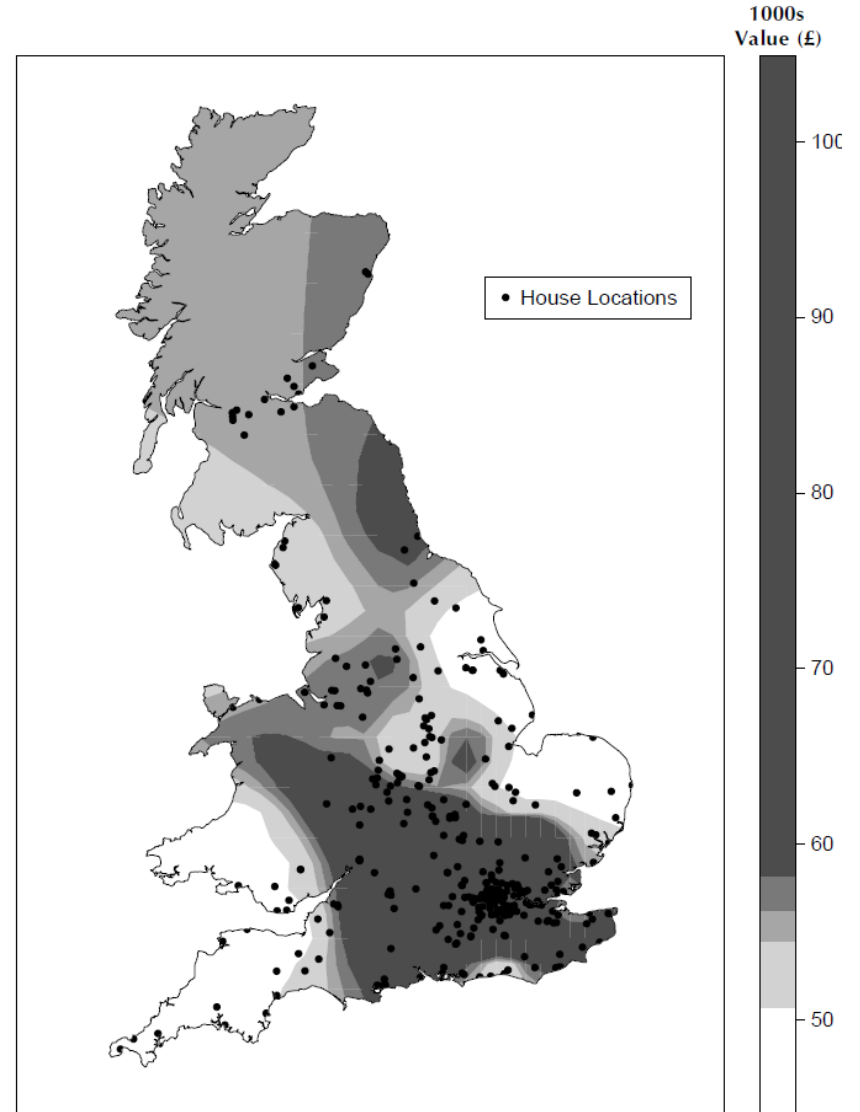
čím blíže k místu výpočtu, tím vyšší váha.

CHARAKTERISTIKY POLOHY

Geograficky vážený průměr

Geograficky vážený průměr cen nemovitostí ve Velké Británii

FOTHERINGHAM, A. S., BRUNSDON, C., and M. CHARLTON. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Chichester: John Wiley & Sons, 2002.



CHARAKTERISTIKY VARIABILITY

Geostatistické charakteristiky variability

Popisují rozmístění (distribuci) bodů kolem středu (míry polohy).

- směrodatná vzdálenost (vážená směrodatná vzdálenost),
- směrodatná elipsa odchylek,
- směrodatná odchylka.

CHARAKTERISTIKY VARIABILITY

Směrodatná vzdálenost (*standard distance*)

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

x_i a y_i jsou souřadnice i -tého bodu, x_{mc} a y_{mc} jsou souřadnice průměrného středu, n je počet bodů.

Zobrazuje se ve formě kružnice kolem průměrného středu s poloměrem o hodnotě směrodatné vzdálenosti.

CHARAKTERISTIKY VARIABILITY

Vážená směrodatná vzdálenost (*weighted standard distance*)

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

f_i jsou váhy i -tého bodu.

I pro sm. vzdálenost lze počítat váženou variantu, kdy vahami jsou hodnoty zkoumané proměnné. Středem kružnice je vážený průměrný střed.

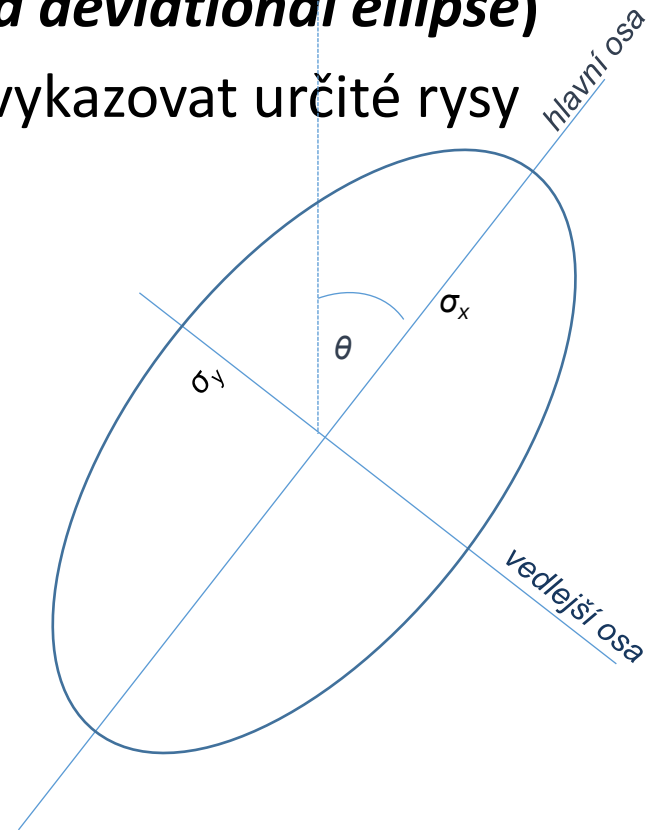
CHARAKTERISTIKY VARIABILITY

Směrodatná elipsa odchylek (*standard deviational ellipse*)

Prostorové rozdělení jevů může často vykazovat určité rysy směrovosti (directional bias).

Místo kruhu je vhodnější použití elipsy se třemi parametry:

- úhel otočení elipsy θ ,
- směrodatná odchylka podél hlavní osy elipsy σ_x ,
- směrodatná odchylka podél vedlejší osy elipsy σ_y .

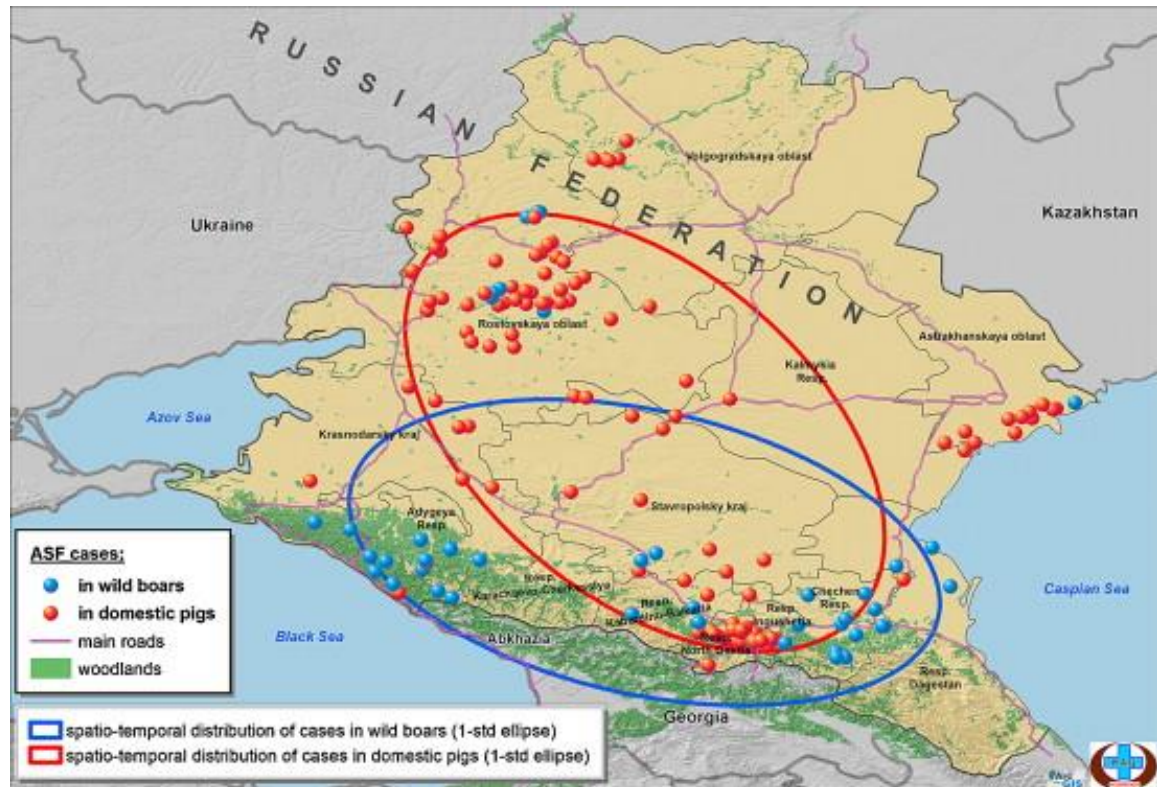


CHARAKTERISTIKY VARIABILITY

Pomocí geostatistických charakteristik polohy a variability lze porovnávat prostorovou variabilitu více souborů mezi sebou.

Porovnání výskytu horečky u domácích a divokých prasat v Severní Osetii – Rusko

V.M. Gulenkin, F.I. Korennoy, A.K. Karaulov, S.A. Dudnikov, Cartographical analysis of African swine fever outbreaks in the territory of the Russian Federation and computer modeling of the basic reproduction ratio, *Preventive Veterinary Medicine*, Volume 102, Issue 3, 1 December 2011, Pages 167-174, ISSN 0167-5877

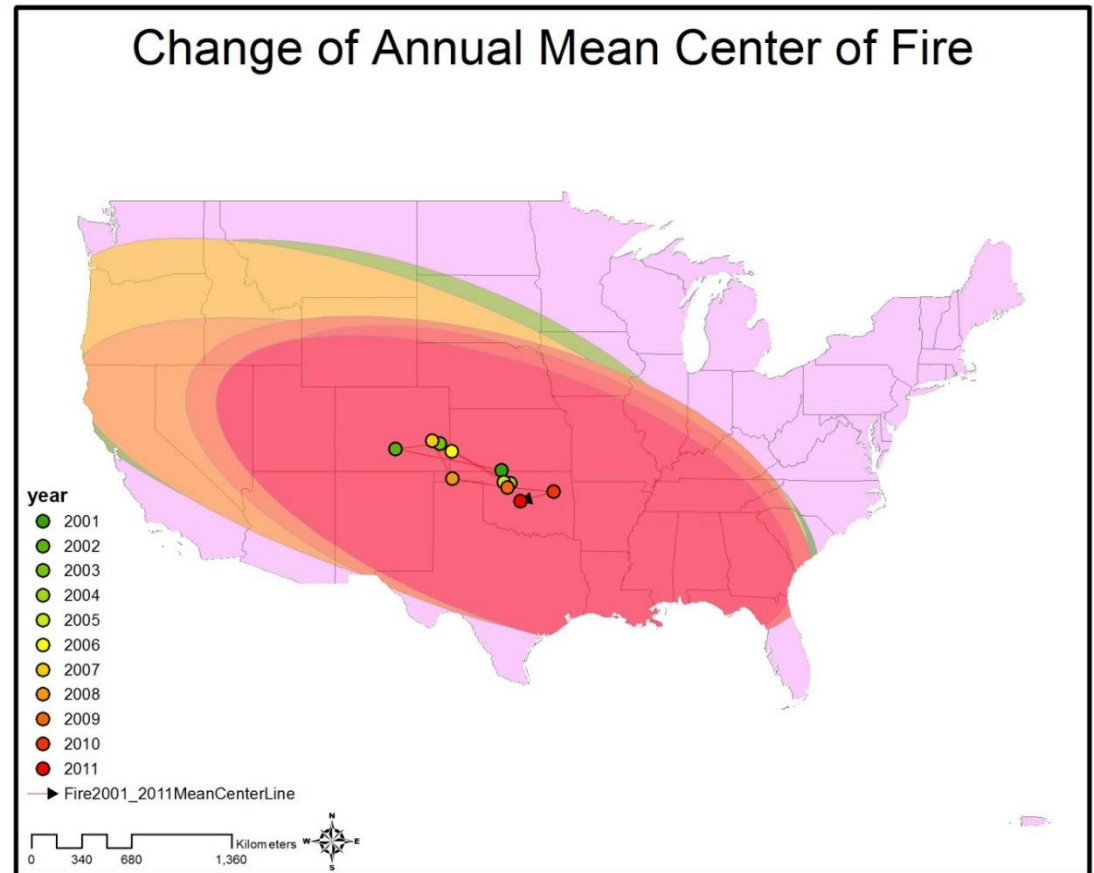


CHARAKTERISTIKY VARIABILITY

Pomocí geostatistických charakteristik polohy a variability lze porovnávat prostorovou variabilitu více souborů mezi sebou.

*Porovnání vývoje požárů
vegetace v USA mezi roky
2001 až 2011.*

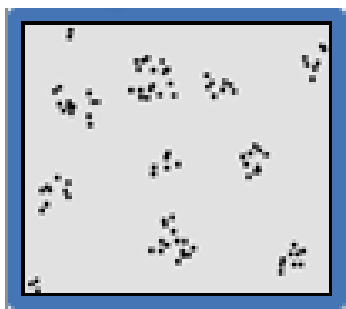
Satoshi Miyazawa. [Note] Hotspot Analysis of Fires
in U.S. during 2001 to 2011 based on Vegetation
Type [online]. Dostupné z:
<http://www.koitaroh.com/search/label/gis>



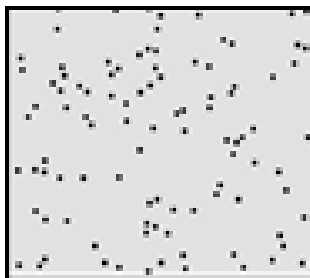
PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Rozlišujeme tři základní typy prostorového uspořádání bodů:

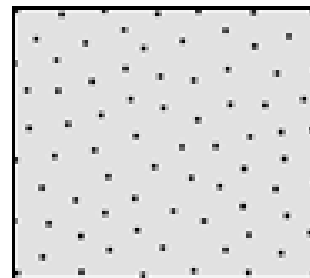
- Shlukové (Clustered)
- Rozptýlené (Dispersed)
- Náhodné (Random)



Shlukové



Náhodné



Rozptýlené

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Základní metody statistického popisu prostorového uspořádání bodů:

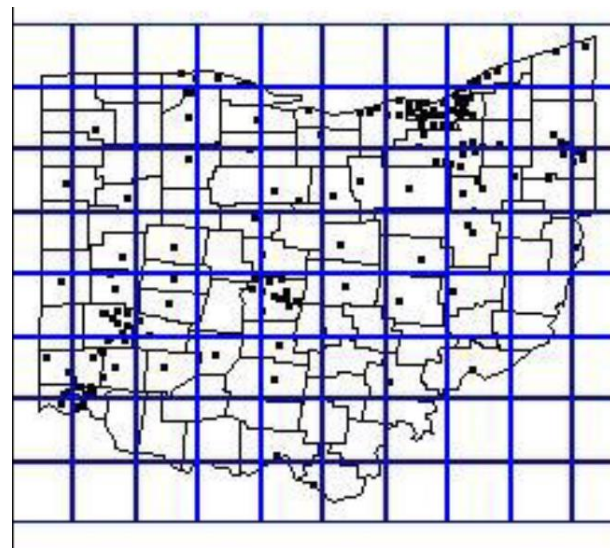
- Analýza kvadrátů (podobnost polohy bodů).
- Průměrný nejbližší soused (podobnost poloha bodů) - *ANN*.
- Metody prostorové autokorelace (podobnost hodnot proměnné s ohledem na polohu bodů) – Moranův index I , Gearyho poměr C , Getis-Ordovo obecné G .

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Analýza kvadrátů (*Quadrat Analysis*)

Hodnocení změn hustoty bodů v prostoru.

- Studované území se rozdělí pravidelnou sítí na buňky a spočte se počet bodů v každé buňce.
- Rozdělení četností buněk s určitým počtem bodů se porovná s náhodným rozdělením četností (nejčastěji s Poissonovým rozdělením pomocí Kolmogorov-Smirnova testu).
- Buňky mohou mít tvar čtverce, kruhu, šestiúhelníku. Tvar a velikost musí být v území konstantní.



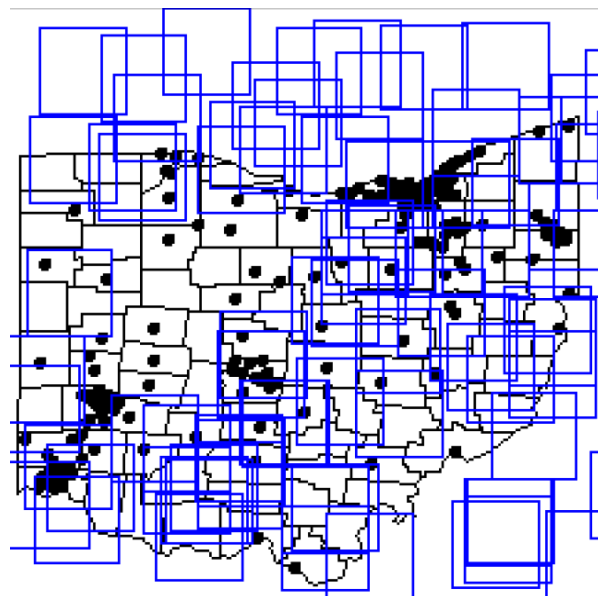
PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Analýza kvadrátů (*Quadrat Analysis*)

Extrémní možnosti:

- Extrémně shlukové uspořádání – většina bodů v jedné či několika málo buňkách.
- Extrémně pravidelné – ve všech buňkách přibližně stejně

Analýzu lze modifikovat tak, že se buňky stejné velikosti náhodně rozmístí po studované ploše.



PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Průměrný nejbližší soused (*Average Nearest Neighbor – ANN*)

Založena na průměrné minimální vzdálenosti mezi body.

$$ANN = D_O / D_E$$

D_O je průměrná minimální vzdálenost $D_O = \sum d_i / n$,

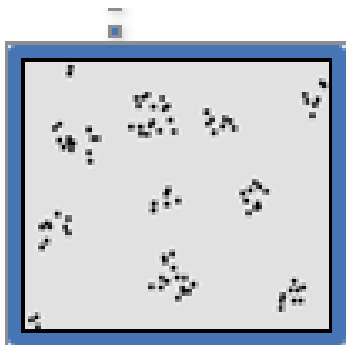
D_E je očekávaná průměrná vzdálenost = $0,5 / \text{odmocnina}(n/A)$,

A je velikost území (plocha).

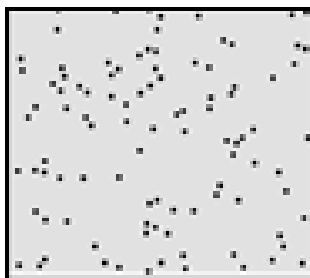
PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Průměrný nejbližší soused (*Average Nearest Neighbor – ANN*)

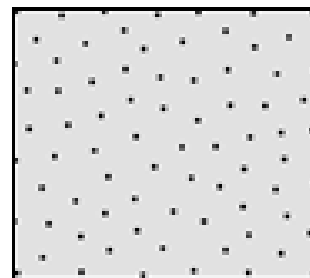
- $ANN > 1$ Rozptýlené uspořádání.
- $ANN = 1$ Náhodné uspořádání.
- $ANN < 1$ Shlukové uspořádání.



Shlukové



Náhodné



Rozptýlené

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Metody prostorové autokorelace

Většina jevů se v prostoru mění spojitě. Blízké body budou mít podobné hodnoty studovaného jevu a naopak (*First law of geography* - Tobler, 1970).

Moranův index I (Moran's I) je koeficientem prostorové autokorelace. Kombinuje míru podobnosti proměnné a míru podobnosti polohy bodů.

Míra podobnosti mezi dvěma body se počítá jako:

$$c_{ij} = (x_i - \bar{x}) \cdot (x_j - \bar{x})$$

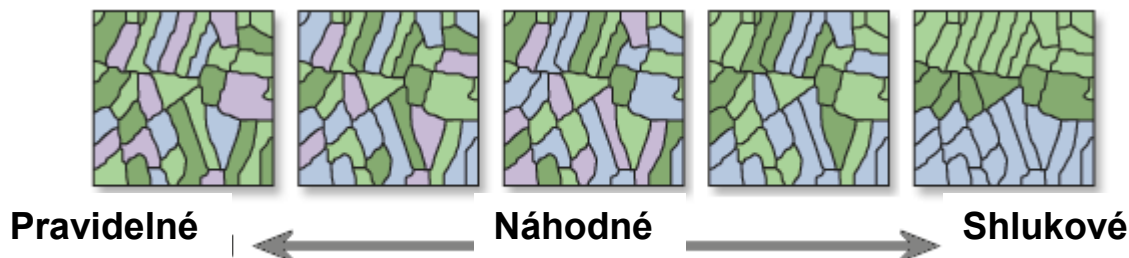
kde x_i resp. x_j jsou hodnoty proměnné v bodě i resp. j ,
 $\bar{x}$ je průměrná hodnota proměnné.

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Moranův index I (*Moran's I*)

- $I > 0$ přesněji $I > E(I)$ *Shlukové uspořádání*, sousední body vykazují podobné hodnoty jako zkoumaný bod (autokorelace).
- $I = 0$ přesněji $I \approx E(I)$ *Náhodné uspořádání*, body nevykazují znaky podobnosti (bez autokorelace).
- $I < 0$ přesněji $I < E(I)$ *Rozptýlené uspořádání*, sousední body vykazují rozdílné charakteristiky (autokorelace).

$$E(I) = -1/(n-1).$$



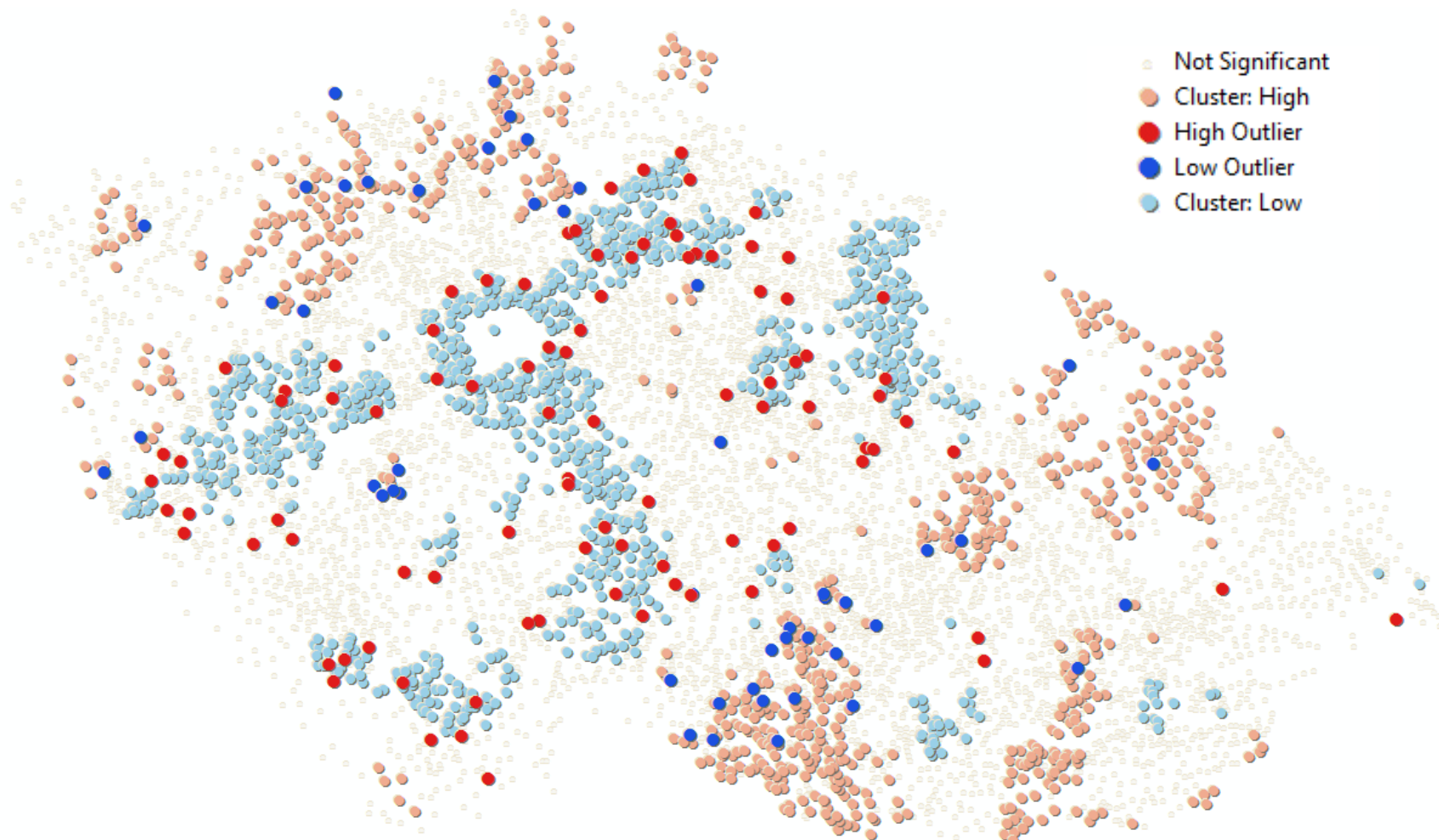
PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Lokální Moranův index I

Lokální varianta indexu se počítá pro každý bod zvlášť! Pomocí jeho standardizované hodnoty (Z-skóre), lze do mapy zakreslit statisticky významné:

- Oblasti bodů bez prostorové autokorelace.
- Shluky bodů s vysokými hodnotami proměnné
- Shluky bodů s nízkými hodnotami proměnné.
- Lokálně odlehlý bod s vysokou hodnotou proměnné.
- Lokálně odlehlý bod s nízkou hodnotou proměnné.

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ



Míra nezaměstnanosti v obcích ČR
interpretace lokálního Moranova indexu ($I=0,28$; $Z=73$) ²⁹

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Gearyho poměr C (*Geary's Ratio*)

Další z koeficientů prostorové autokorelace.

Jeho lokální varianta dokáže identifikovat body, které se nepodobají svému okolí (odlehle hodnoty).

Míra podobnosti mezi dvěma body se počítá jako:

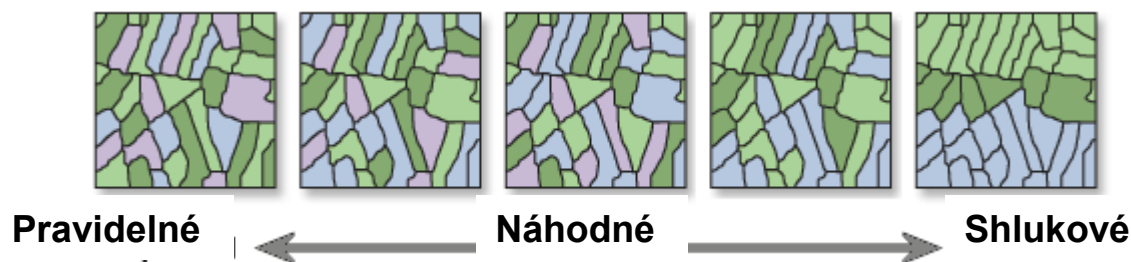
$$c_{ij} = (x_i - x_j)^2$$

kde x_i resp. x_j jsou hodnoty proměnné v bodě i resp. j .

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Gearyho poměr C (*Geary's Ratio*)

- $0 < C < 1$ *Shlukové uspořádání*, sousední body vykazují podobné hodnoty jako zkoumaný bod (autokorelace).
- $C \approx 1$ *Náhodné uspořádání*, body nevykazují znaky podobnosti (bez autokorelace).
- $1 < C < 2$ *Rozptýlené uspořádání*, sousední body vykazují rozdílné charakteristiky (autokorelace).



PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Getis-Ordovo obecné G (*Getis-Ord General G*)

Používá se pro tzv. analýzu *Hot-Spot*, tedy k identifikaci shluků s vysokými (*hot spot*) a nízkými (*cold spot*) hodnotami.



Shluk nízkých
hodnot (cold spot)



Náhodné

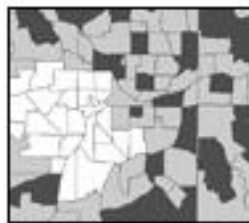


Shluk vysokých hodnot
(hot spot)

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

Getis-Ordovo obecné G (*Getis-Ord General G*)

- $C < 1$ *Shluk nízkých hodnot (Cold-spot)*, sousední body vykazují podobné nízké hodnoty jako zkoumaný bod (autokorelace).
- $C \approx 1$ *Náhodné uspořádání*, body nevykazují znaky podobnosti (bez autokorelace).
- $C > 1$ *Shluk vysokých hodnot (Hot-spot)*, sousední body vykazují podobné vysoké hodnoty jako zkoumaný bod (autokorelace).



Shluk nízkých hodnot (cold spot)



Náhodné



Shluk vysokých hodnot (hot spot)

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ

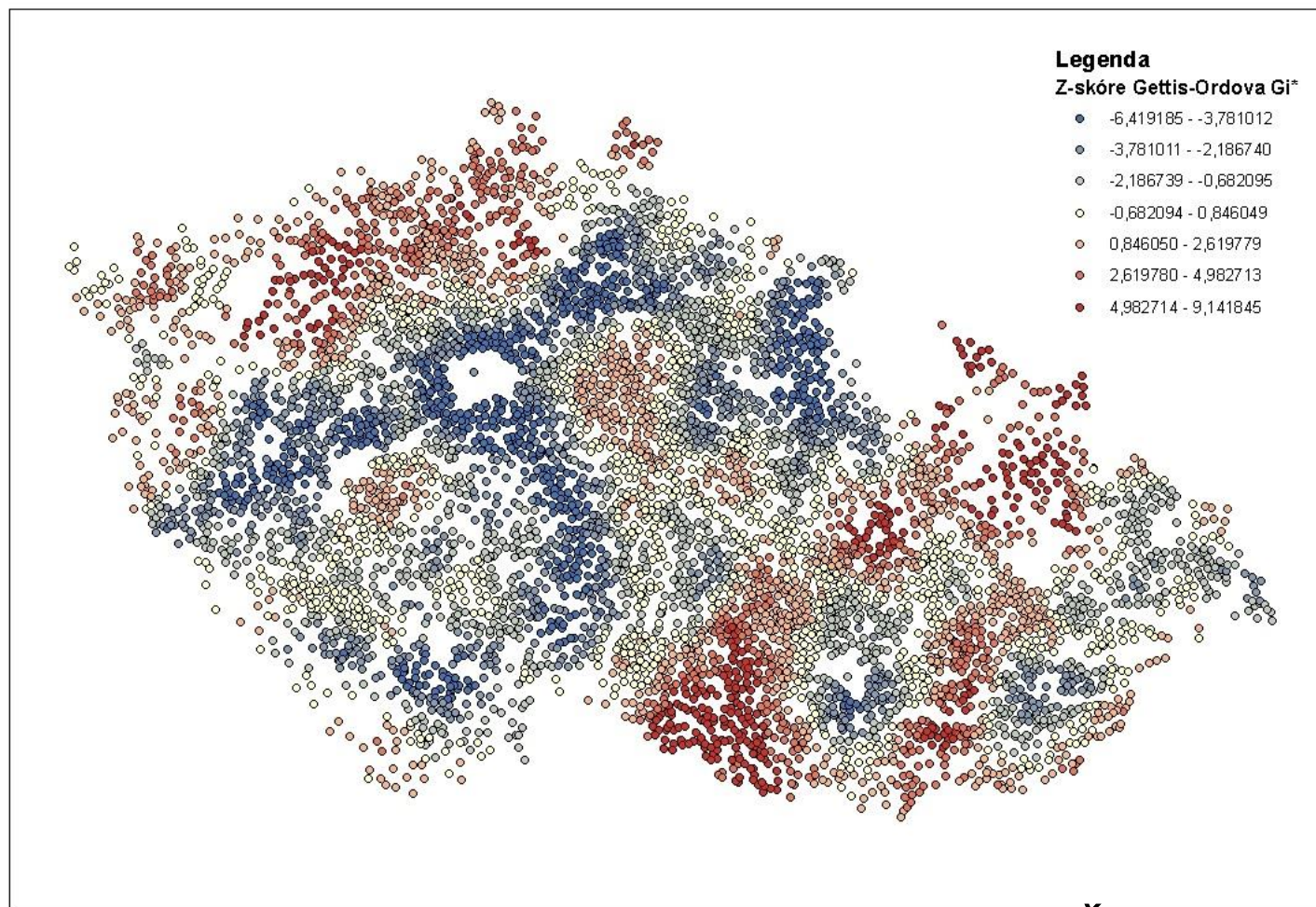
Getis-Ordovo lokální G_i^* (*Getis-Ord G_i^)**

Lokální varianta indexu se počítá pro každý bod zvlášť! Vhodný nástroj pro interpretaci prostorově nespojitých jevů (jevů, které není vhodné interpolovat).

Pomocí jeho standardizované hodnoty (Z-skóre), lze do mapy zakreslit statisticky významné:

- Oblasti bodů bez prostorové autokorelace.
- Shluky bodů s vysokými hodnotami proměnné (Hot-spot).
- Shluky bodů s nízkými hodnotami proměnné (Cold-spot).

PROSTOROVÉ USPOŘÁDÁNÍ BODŮ



Míra nezaměstnanosti v obcích ČR
z-skóre Getis-Ordova lokálního G_i^ ($G=0,28$; $Z=73$)*

ZDROJE (*pro celý původní materiál*)

Použité zdroje:

CHARLTON, Martin and A Stewart FOTHERINGHAM. *Geographically Weighted Regression. A Tutorial on using GWR in ArcGIS 9.3*. National University of Ireland Maynooth, 2009.

ESRI. *ArcGIS 10.1 Help*. Esri. Copyright © 1995-2012.

FOTHERINGHAM, A. S., BRUNSDON, C., and M. CHARLTON. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Chichester: John Wiley & Sons, 2002.

LLOYD, Christopher D. *Local models for spatial analysis. 2nd ed.* CRC Press Taylor & Francis Group, 2011. ISBN 978-1-4398-2919-6.

SCHEJBAL, Ctirad. *Úvod do geostatistiky*. Ostrava: VŠB, 1996. ISBN 80-7078-325-7.