

Statistický popis prostorového uspořádání bodů

10.4.2024

Google Earth Engine

The background of the image is a server room. Rows of server racks are visible, each filled with numerous server units. The room is dimly lit, with a strong blue glow emanating from the server units, creating a futuristic and high-tech atmosphere. The perspective is from a low angle, looking down a central aisle between the racks.

Google Earth Engine combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities and makes it available for scientists, researchers, and developers to detect changes, map trends, and quantify differences on the Earth's surface

Desktopové vs. cloudové platformy

Desktopové



Cloudové platformy



and

Google Earth Engine

- Revoluční technologie pro zpracování satelitních dat
- Otevřená vědecká platforma s možností snadného přístupu k datům
- Tvorba a sdílení metod/algoritmů
- K dispozici obrovský výpočetní výkon
- Široká a živá komunita výzkumníků
- Přibližuje lidem družicové snímky a metody na jejich zpracování
- Velký výzkumný a aplikační potenciál

Datové typy

- **ee.Image** – raster, 1 nebo více pásem; každé pásmo má vlastní jméno, měřítko (scale), projekci a masku + metadata jako množinu atributů (properties).
- **ee.ImageCollection** – kolekce souvisejících snímků (ee.Image) se stejnou charakteristikou a vlastnostmi. Vlastní ID.
- **ee.Feature** – vektor, jako objekt GeoJSON. Dvě vlastnosti: 1) Geometry - geometrie (případně NULL) a 2) Property - metadata, které obsahuje jednotlivé atributy.
- **ee.Geometry** - Point, LineString (a list of points), LinearRing (a closed LineString), and Polygon (a list of LinearRings where the first is a shell and subsequent rings are holes). MultiPoint, MultiLineString, and MultiPolygon. MultiGeometry (GeoJSON GeometryCollection).
- **ee.FeatureCollection** – kolekce souvisejících objektů (ee.Feature).
- **ee.Array** - pole s 1D vektory, 2D matice, 3D cubes nebo vícedimenzionální prostory
- **ee.Number, ee.String, ee.Date, ee.List, ee.Dictionary,...**

Geoprostorové funkce

- **Filter** - by bounds, within distance, date, day-of-year, metadata ...
- **Reducer** - způsob agregace dat v čase, prostoru, pásmech, polích a dalších datových strukturách, např. mean, max, min, lineární regrese, histogram, zonal statistics
- **Join** - kombinování prvků z různých kolekcí (např. ImageCollection nebo FeatureCollection) na základě podmínky určené ee.Filter, např. simple, inner, outer, inverted, spatial join ...
- **Základní GIS funkce** - clip, buffer, intersect, union, dissolve, ...
- **Chart** - integrace Google Charts. Column, pie chart, scatter plot, histogram, time-series, ...
- **Export** - export obrázků, mapových dlaždic, tabulek a videí
- **Machine Learning** - Řízené a neřízené per-pixel klasifikace, OBIA. Např. CART, random forests, bayes, SVM, k-means, cobweb
- **Pokročilé funkce** – Fmask, Change Detection algoritmy, ...

více než 800 předdefinovaných funkcí... a stále se rozšiřující možnosti

Webové rozhraní GEE

The image shows a screenshot of the Google Earth Engine web interface. The interface is divided into several sections, each labeled with a Czech text and an arrow pointing to the corresponding part of the UI:

- Správce skriptů** (Scripts manager): Points to the 'Scripts' tab in the top navigation bar.
- Dokumentace funkcí** (Function documentation): Points to the search bar at the top.
- Vyhledávač míst a datasetů** (Places and datasets search): Points to the search bar at the top.
- Inspektor pixelu** (Pixel inspector): Points to the 'Inspector' tab in the right sidebar.
- Běžící úkoly** (Running tasks): Points to the 'Tasks' tab in the right sidebar.
- Správce assetů** (Assets manager): Points to the 'Assets' tab in the top navigation bar.
- Editor JavaScript kódu** (JavaScript code editor): Points to the central area where code is written.
- Konzole** (Console): Points to the 'Console' tab in the right sidebar.
- Mapové okno** (Map window): Points to the map area at the bottom of the interface.

The map at the bottom shows the United States with various cities and state names labeled. The interface also includes a top navigation bar with 'Scripts', 'Docs', and 'Assets' tabs, and a right sidebar with 'Inspector', 'Console', and 'Tasks' tabs.

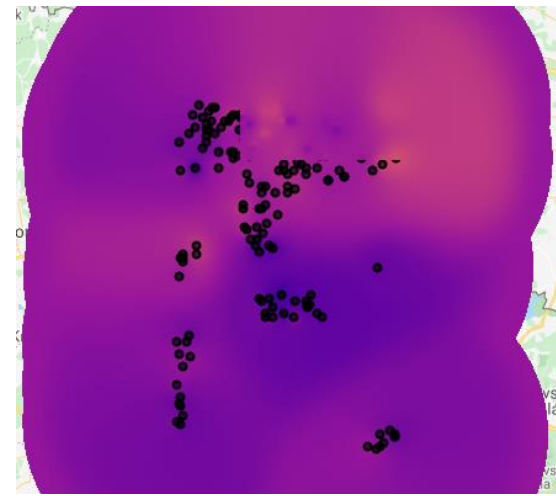
<https://code.earthengine.google.com/>

Interpolace v GEE - IDW

```
var interpolated = samples.inverseDistance({
  range: 2e4,
  propertyName: 'ch4',
  mean: stats.get('mean'),
  stdDev: stats.get('stdDev'),
  gamma: 0.3});

var band_viz = {
  min: 1800,
  max: 1900,
  palette: ['0D0887', '5B02A3', '9A179B', 'CB4678', 'EB7852', 'FBB32F', 'F0F921']};

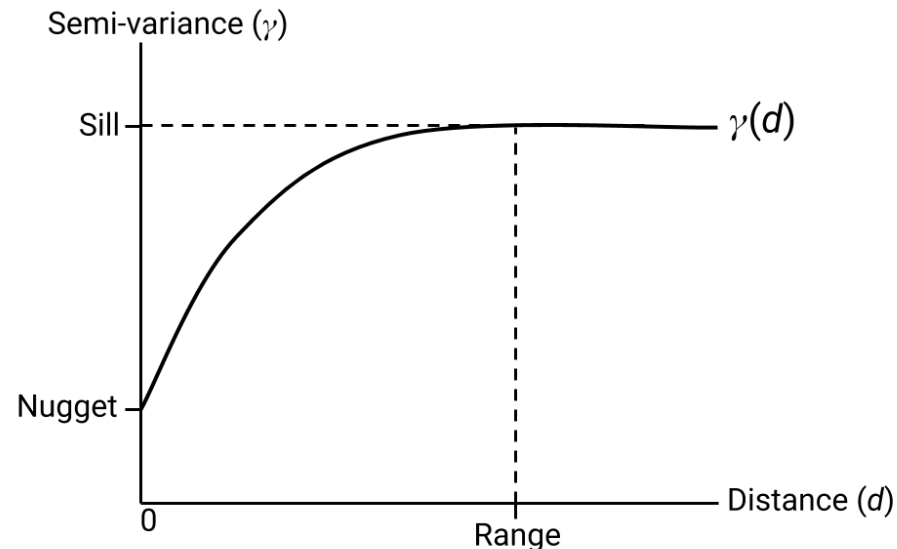
Map.centerObject(aoi, 7);
Map.addLayer(ch4.clip(aoi), band_viz, 'CH4');
Map.addLayer(interpolated, band_viz, 'CH4 interpolovaný');
Map.addLayer(samples, {}, 'Vzorky')
```



Interpolace v GEE - Kriging

```
var samples = sst.addBands(ee.Image.pixelLonLat())
  .sample({region: geometry, numPixels: 1000})
  .map(function(sample) {
    var lat = sample.get('latitude');
    var lon = sample.get('longitude');
    var sst = sample.get('sst');
    return ee.Feature(ee.Geometry.Point([lon, lat]), {sst: sst});
  });
```

```
var interpolated = samples.kriging({
  propertyName: 'sst',
  shape: 'exponential',
  range: 100 * 1000,
  sill: 1.0,
  nugget: 0.1,
  maxDistance: 100 * 1000,
  reducer: 'mean',
});
```



Analýza vzorů

- Proč analyzovat vzory?
- Identifikace geografických vzorů nám umožní lépe pochopit, jak se geografické jevy chovají
- Jak je analyzovat?
- Mapováním získáme určitou představu o celkovém vzoru prvků a jejich přidružených hodnotách
- Nicméně, celkovou kvantifikaci vzoru získáme pomocí statistického výpočtu, což nám umožní porovnat vzory pro různá rozdělení či časová období
- Často jsou také tyto analýzy výchozím bodem pro hlubší analýzy

Spatial statistics toolbox/Analyzing patterns

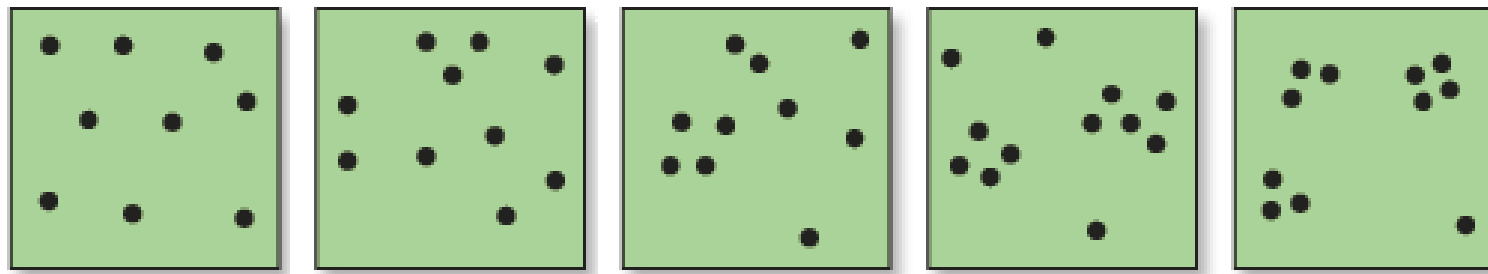
- Nástroje v této sadě nástrojů využívají tzv. **Interferenční statistiky**, což znamená vytváření generalizovaných závěrů o celkové populaci na základě omezeného množství data té populace
 - Její metody nám umožňují odhadovat parametry populace
 - Testovat hypotézy a posuzovat nejistoty těchto odhadů a testů pomocí p-hodnot, což nám dává pravděpodobnost, že nulová hypotéza je správná (že pozorovaný vzor je jednoduše jednou z mnoha možných verzí úplné prostorové náhodnosti)
- Využívá se toho převážně pokud potřebujeme mít velkou důvěru v určitém rozhodnutí (např. když se na základě toho rozhodnutí dělají právníkové případy, veřejná bezpečnost atd., zkrátka když je potřeba to rozhodnutí mít statisticky potvrzeno)
- Typicky se snažíme odpovědět na otázky jako:
 - Jsou prvky v datové sadě nebo hodnoty spojené s prvky v datové sadě prostorově shlukované?
 - Stává se shlukování postupem času více či méně intenzivní?

Nástroje dostupné v Arcgis Pro

- Average Nearest Neighbor
- High/Low Clustering
- Incremental Spatial Autocorrelation
- Multi-Distance Spatial Cluster Analysis (Ripley's k-function)
- Spatial Autocorrelation

Average Nearest Neighbor

- Vypočítá stupeň prostorového shlukování bodových objektů
- Hodnotí se vzdálenost mezi každým bodem a jeho nejbližším sousem a tyto vzdálenosti se porovnávají s očekávanými vzdálenostmi v případě náhodného rozložení.



Dispersed



Clustered

Nearest Neighbor Index <1

Rozptýlené rozložení

1

Náhodné rozložení

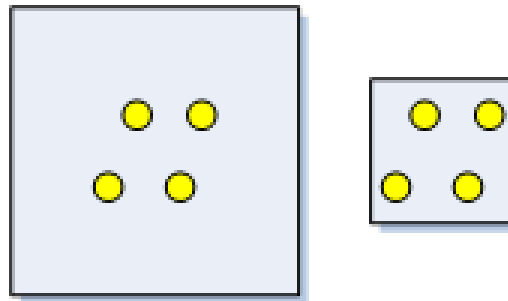
>1

Shlukové rozložení

- Nástroj Average Nearest Neighbor vrací pět hodnot:
 - Pozorovanou střední vzdálenost
 - očekávanou střední vzdálenost
 - index nejbližšího souseda
 - z-skóre
 - p-hodnotu
- Hodnoty jsou zapsány jako zprávy ve spodní části podokna Geoprocessing během provádění nástroje a předány jako odvozené výstupní hodnoty pro potenciální použití v modelech nebo skriptech.
- Lze vytvořit také HTML report s grafickým shrnutím výsledků

– Z-score a p-hodnoty:

- Míry statistické významnosti, podle nich se můžeme rozhodnout zda zamítnout či potvrdit nulovou hypotézu
- Nulová hypotéza zde tvrdí, že jsou prvky rozmístěny náhodně
- Je třeba ale zmínit, že statistická významnost této metody je silně ovlivněna velikostí studijní oblasti (malé změny v hodnotě parametru Area mohou vést ke značným změnám ve výsledcích z-skóre a p-hodnoty)
- Proto je tento nástroj nejúčinnější pro porovnávání různých prvků v dané studijní oblasti. (Níže uvedený obrázek je klasickým příkladem toho, jak mohou být identické distribuce prvků rozptýleny nebo seskupeny v závislosti na specifikované studijní oblasti)



— Nearest Neighbor Index:

- poměr pozorované střední vzdálenosti k očekávané střední vzdálenosti
- očekávaná vzdálenost je průměrná vzdálenost mezi sousedy v hypotetickém náhodném rozdělení
- pokud je index menší než 1, vzorek vykazuje shlukování; pokud je index větší než 1, trend směřuje k rozptylu

— Area:

- Pokud není hodnota parametru Area zadána, použije se oblast minimálního ohraničujícího obdélníku kolem vstupních prvků

— Vstupní data by měla být v projekčním souřadnicovém systému

— Nástroj je nejvhodnější použít na bodové prvky

Cvičení

- Budeme zkoumat teoretická rozložení na příkladu drogové kriminality v londýnských čtvrtích
- Data:
 - Stáhnout z IO
 - Použít SS ETRS 1989 UTM Zone 30N
 - London_LAU2.shp – administrativní dělení Velké Británie v úrovni LAU2
 - London_Crimi_2016-02.shp – bodová hlášení přestupků a trestných činů v únoru 2016 ("Crime_type" = 'Drugs')

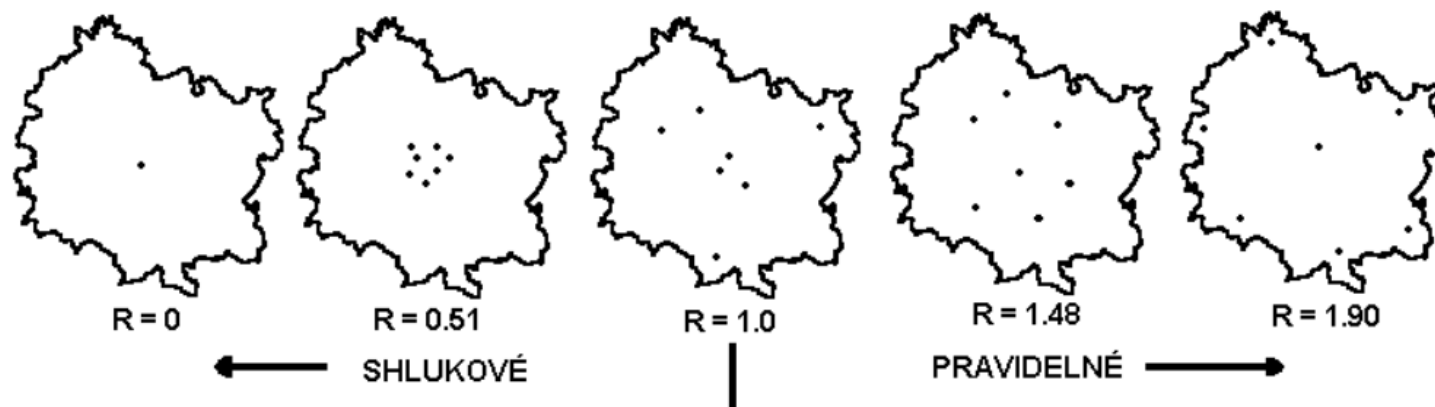
- Zadání:
 - Zjistěte, jestli je drogová kriminalita v londýnských čtvrtích Camden Town (with Primrose Hill), Shepherd's Bush Green a Harlesden prostorově náhodně rozložená nebo se váže na určitou oblast.
 - Pomocí vhodné charakteristiky popište, k jakému z teoretických rozložení (shlukové či pravidelné) se vámi zjištěné uspořádání ve čtvrtích blíží (udejte statistickou významnost).
 - Stručně interpretujte hodnoty vypočtených charakteristik.
 - K hodnocení prostorového uspořádání sídel použijte metodu nejbližšího souseda a kvadrátovou analýzu.

– Postup vypracování:

– **Metoda nejbližšího souseda** - Používaná statistika je poměrem vzdáleností

($R = r\text{-obs} / r\text{-exp}$)

- Čím je hodnota $R < 1$, tím více se prostorové rozložení bodů blíží rozložení shlukovému ($r\text{-obs} < r\text{-exp}$)
- Čím je hodnota $R > 1$, tím více se prostorové rozložení bodů blíží rozložení pravidelnému ($r\text{-obs} > r\text{-exp}$)



$R = 0$ zcela shlukové uspořádání
 $R = 1$ náhodné uspořádání
 $R = 2,149$ zcela pravidelné uspořádání

- Je-li z-score $< -1,96$ či z-score $> 1,96$, potom vypočtený rozdíl mezi pozorovaným a náhodným uspořádáním je statisticky významný – tedy není náhodný a naopak.

- 1) Zadaný úkol vyřešte pro všechny 3 čtvrtě (zmíněné v zadání) metodou nejbližšího souseda v prostředí ArcMap. Metodu lze spustit pomocí ArcToolbox – Spatial Statistics Tools – Analyzing Patterns – Average Nearest Neighbour.
- Tip:
 - Nezapomeňte, že chceme pracovat pouze s drogovou kriminalitou, tudíž prvním vaším krokem by mělo být nějakým způsobem vyfiltrovat pouze tyto záznamy z atributové tabulky. Způsobů, jak tohle udělat, je mnoho, takže zvolte jaký uznáte za vhodné.
 - V dalším kroku najděte pouze ty městské části, které nás zajímají a pro každou si vytvořte vlastní shapefile.
 - Takto vzniklými shapefily pak ořežte bodovou vrstvu kriminality.
 - Na konci toho zpracování byste tedy měli mít 3 plošné vrstvy (musíte si je přeuložit, jinak nebude fungovat tvorba sítě) pro požadované londýnské čtvrti a 3 bodové vrstvy s drogovou kriminalitou vztahující se k těmto čtvrtím.
 - Bodové vrstvy budou poté vstupovat do nástroje v Arcgis.

- 2) V nástroji Average Nearest Neighbour pak vkládejte bodové vrstvy, do políčka Area byste měli vložit dopočítanou hodnotu plochy dané čtvrti (toho lze snadno dosáhnout přes možnost Calculate Geometry při kliknutí na sloupeček v atributové tabulce, musíte si jen vytvořit nový sloupeček, kam se area bude ukládat). Zaškrtněte také políčko Generate Report - Arcgis vám potom vytvoří přehledný html soubor s výsledky (cestu k němu se dozvíte v oknu Results).

— Kvadrátová analýza

- 1) Nejprve je nadefinována síť kvadrátů (čtverců). Kvadráty generujte pomocí nástroje Create Fishnet přibližně v rozsahu čtvrtě. Tato síť se přeloží přes studovanou oblast. Pro nadefinování sítě musíte určit počet buněk v síti.
 - první varianta: vyzkoušejte výpočet, kdy počet buněk je roven přibližně polovině počtu bodů (například mám 48 bodů, polovina z nich je 24, tudíž území rozdělím buňky na 4×6, 6×4, 3×8 nebo 8×3 - to záleží na tvaru území)
 - druhá varianta: vyzkoušejte postup, kdy velikost jedné buňky a počet buněk jsou odvozeny z následujících vztahů: velikost kvadrátu = $(2 \cdot A)/n$, kde A je plocha studované oblasti a n počet analyzovaných bodů
 - velikost strany jedné buňky (to co se zadává do nástroje Create Fishnet) je $\sqrt{2A/n}$
- Tip:
 - V nástroji Create Fishnet vyplňujete buď pole Cell size nebo Number of Rows/Columns - v závislosti na požadované variantě. Také si dole zatrhněte Geometry Type - Polygon (chceme plošnou vrstvou) a odznačte Create Label Point - tuto vrstvu nepotřebujeme.

2) Zjistěte, kolik sídel je v každém kvadrátu (nástroj Spatial Join)

— Tip:

- Target Feature - Fishnet
- Join Feature - drugs

3) Pomocí souboru z Excelu vypočtete testovací kritéria a kritickou hodnotu

— Cíl - Excel

- Sloupeček Reálný počet nám říká, kolik je kvadrátů s daným počtem přestupků.
- V modelovém případě je tedy 8 kvadrátů které mají 0 přestupků, dále 4 kvadráty které mají 1 přestupek atd.
- Vy musíte správně vyplnit sloupečky Reálný počet (dle toho co vám vyšlo pomocí nástroje Spatial Join).
- Shlukové rozdělení tak, že v prvním řádku pro 0 přestupků je 24 kvadrátů a v posledním řádku je 1 kvadrát (hodnoty platí pro modelový případ).
- Ostatní buňky jsou v souboru řešené pomocí excelových funkcí, takže do nich nemusíte zasahovat, pokud to nebude vyloženě nutné kvůli nějaké úpravě.
- Ještě také budete muset upravit sloupeček Počet bodů v kvadrátu tak, aby odpovídal daným fishnetům pro dané čtvrtě.

— Hladina významnosti a kritická hodnota:

$$0,01 = 1,63\sqrt{n}$$

$$0,05 = 1,36\sqrt{n}$$