

MUNI
SCI



HR EXCELLENCE IN RESEARCH

CoKriging

27.3.2024

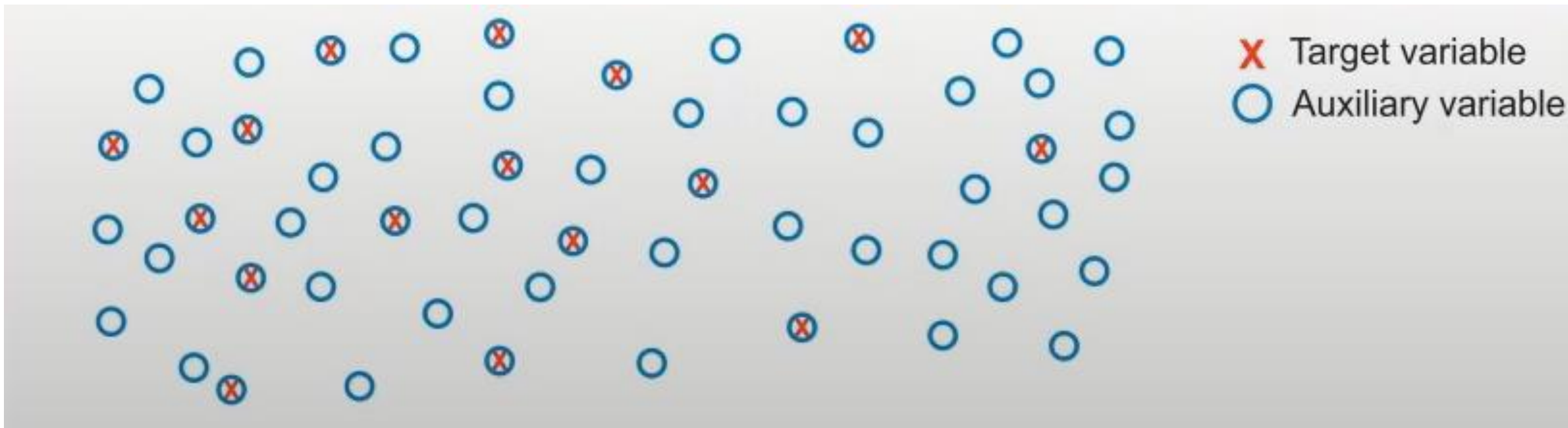
Komentář k dotazům 1. protokolu

- Odlehlé hodnoty \neq extrémně odlehlá hodnota
- V oborech jako je meteorologie, mohou být mírné odchylky od normálu v ohonech zcela normální a očekávané kvůli povaze extrémů počasí
- Není nutnost ji odstranit (nicméně např. u RBF by to mohl být problém, ve výpočtu je RBF citlivá na odlehlé hodnoty)
- Nesplňuje to ale podmínku normálního rozdělení
- Je dobré také takovou hodnotu vyhledat, a ujistit se jestli je to opravdu chyba nebo ne (zde to byla plechová střecha, takže ani to nebyla chyba v měření, ale fyzikální jev v DPZ)

CoKriging

- Rozšíření Krigingu
- Používá se když, máme dvě (nebo více datových sad) ve stejném území, o kterých předpokládáme, že se sebou nějakým způsobem souvisí - jsou korelované
- Pro data, která chceme interpolovat však máme pouze omezené množství výskytů (třeba z důvodu náročné akvizice dat - ať již finanční nebo například neexistujících měření v rámci celého území).
- Naproti tomu druhá datová sada pokrývá určitým způsobem celé území. Pokud jsou tyto dvě datové sady závislé (tím, že se mění jedna, očekáváme, že se podobně bude měnit i druhá), můžeme tato data využít pro interpolaci požadované veličiny.

- Možné použití
 - Teplota – nadmořská výška
 - Nadmořská výška podzemní vody - výška terénu
- Pomocná proměnná musí být korelovaná s modelovanou proměnnou geograficky i vlastnostma (negativní nebo pozitivní korelace)
- Některé vzorky by měly být měřeny společně



Cvičení 1 - Kriging

- Ve cvičení použijeme datovou sadu, se kterou jsem již pracovali
- Oregon.gdb
 - Data teplot jsou ve stupních Fahrenheita. Je zde i DEM

Postup při zpracování

- 1) - Zjistěte, zda vstupní data mají normální rozdělení => histogram, QQplot
 - zda obsahují trend a je vhodné je podrobit transformaci,
 - zda obsahují outliery,
 - jestli pole bodů vykazuje izotropii.
 - Výsledků explorační analýzy využijte při volbě vhodného teoretického modelu při strukturní analýze.

- 2) Pomocí nástroje Geostatistical Wizard zvolte metodu interpolace krigování (Kriging/CoKriging) a v rámci ní metodou tzv. základního krigování (Ordinary Kriging) s pomocí modelování semivariogramu vytvořte mapu interpolovaných hodnot.

– **Transformation type:** (Kriging předpokládá, že máme data podle normálního rozdělení, bez nějakých trendů v datech, nastavením typu transformace pomáháme řešit tyto problémy transformací dat do formy, která lépe vyhovuje předpokladům modelu kriging)

– **1. Žádná (výchozí)**

- Vhodné, pokud máme data s normálním rozdělením nebo jsou dostatečně blízko, tak, že přínos transformace je zanedbatelný

– **2. Log (Logaritmická)**

- Aplikuje logaritmickou transformaci na data. To je zvláště užitečné, když jsou data pozitivně šikmá (dlouhý ocas vpravo). Může stabilizovat varianci a udělat rozdělení symetričtější, což je blíže normálnímu rozdělení, které kriging předpokládá.

– **3. Normal Score**

- Transformuje data tak, aby měla standardní normální rozdělení. Tato transformace je univerzálnější a může být použita, když data významně odchyľují od normality jakýmkoli způsobem. Data se seřadí a poté se aplikuje transformace, aby rozdělení těchto řad následovalo normální rozdělení.

– Arcsin

- Může být užitečná pro data, která jsou mírně šikmá doprava. Podobně jako logaritmická transformace, má za cíl snížit šikmost a stabilizovat varianci.

– Box-Cox

- flexibilnější možnost, která zahrnuje logaritmické a odmocninné transformace jako speciální případy. Snaží se identifikovat nejlepší lambda (λ) parametr, který přivádí data k normalitě. Tato možnost vyžaduje, aby data byla striktně kladná.

– Výběr vhodné transformace závisí na charakteristikách dat a konkrétních požadavcích na prostorovou analýzu.

– Nejlepšího přístupu je ovšem někdy dosaženo experimentováním s různými transformacemi

– Decluster before transformation

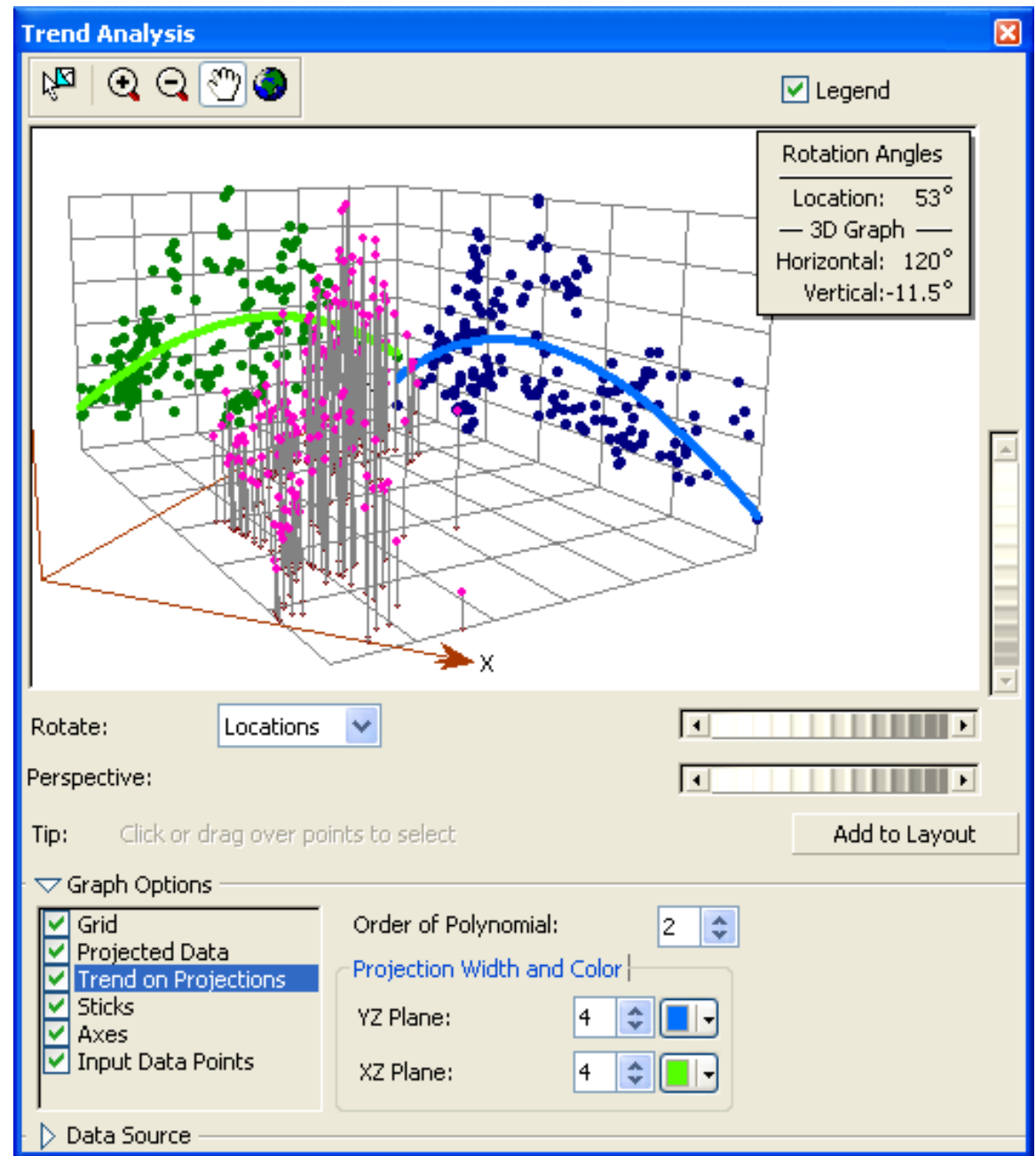
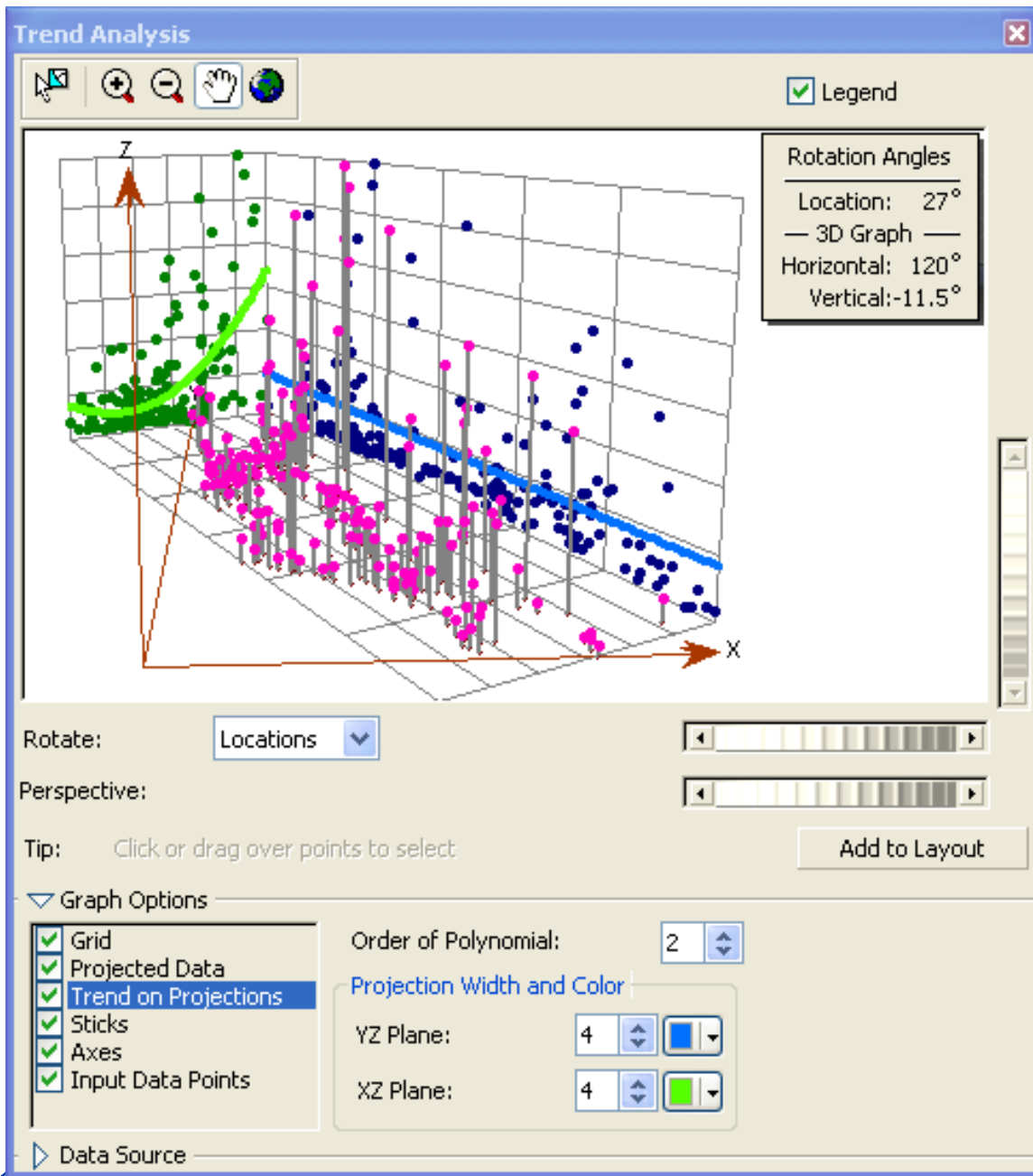
- „spravedlivější a reprezentativnější“ zohlednění různých částí studované oblasti (typicky nižší váhy v oblastech s vysokou hustotou vzorků a vyšší s nižší hustotou vzorků)

Proč použít declustering?

- Vyrovnání rozdílů v prostorovém rozložení dat, což je důležité pro kvalitní prostorovou analýzu a modelování.

– Order of trend removal

- k zohlednění a eliminaci velkých, systematických trendů v datech před samotnou interpolací
- stupeň polynomu, který se použije k modelování a odstranění trendu z dat před interpolací.
- **0. řád:** Neodstraňuje se žádný trend; předpokládá se, že data jsou stacionární bez potřeby úpravy.
- **1. řád:** Lineární trend. Používá se lineární model pro odstranění trendů, které se mění lineárně s lokací.
- **2. řád:** Kvadratický trend. Odstraní složitější trendy, které se mohou měnit rychleji než lineárně s lokací.
- příliš nízký řád nemusí adekvátně odstranit existující trendy v datech, zatímco příliš vysoký řád může vést k nadměrnému vyhlazení dat a ztrátě důležitých informací.



Normal score transformations

- Cílem je nastavit spolehlivou metodologii pro prostorovou interpolaci, i když výchozí rozdělení dat není ideální
- **Approximation method** - výběr specifického přístupu k aproximaci (odhadu) hodnot v nevzorkovaných místech na základě dostupných datových bodů.
 - Volba metody aproximace závisí na velikosti a povaze datové sady
 - Pro malé až středně velké datové sady může být vhodné použít přesné metody, pro velké datové sady jsou preferovány aproximační metody
 - Kompromis mezi potřebou přesnosti a dostupnými výpočetními zdroji
- **Number of modifiers** - Větší počet modifikátorů může umožnit detailnější modelování trendů a variabilit v datech, ale také zvyšuje riziko přetrénování modelu, kdy model příliš přesně odpovídá vzorkovacím datům a ztrácí schopnost generalizace

- 3) V následné strukturní analýze provedte několik nastavení parametrů vhodného teoretického modelu semivariogramu. Prozkoumejte případnou anizotropii vašich vstupních dat. Vaším cílem je nalézt vhodný teoretický model semivariogramu, jehož parametry budou vstupovat do vlastní interpolace metodou krigování jako váhy. Volte různé modely, hodnoty dosahu (range), prahu (sill) a zbytkového rozptylu (nugget). Volit můžete též hodnoty vzdálenosti (lag) - tuto hodnotu mohu zjistit pomocí nástroje **Average Nearest Neighbor**, na kterou se spojují obdobně vzdálené body při výpočtu empirických hodnot semivariancí a také počet hodnot lag.
- **Tip - anizotropie**
 - Anizotropii povolíte změněním parametru Anisotropy na "true". Vliv anizotropie pak můžete zkoumat při změně "show search direction" opět na "true" a podobně jako tomu bylo u průzkumové analýzy, můžete zakliknutím a potažením myši měnit zkoumaný směr.

- 4) Vhodnost nastavení parametrů semivariogramu kontrolujte hodnotami průměrné chyby predikce (Mean Prediction Error) – ideálně nula a průměrné čtvercové chyby (Root Mean Square Prediction Error) – čím menší, tím lepší odhad.
- 5) Vytvořte mapu směrodatné chyby predikce (pravým tlačítkem na vytvořený korigovaný povrch Change Output to Prediction Standard Error). Z ní určete, která část zpracovávaného území vykazuje největší chyby predikce. Analogicky se vrátíte k původní mapě.
- 6) Obdobně vyzkoušejte Simple a Universal kriging. U metody Simple kriging použijte Normal Score Transformation (defaultní) a u Universal kriging odstranění trendu (trend se v prvním kroce odstraní, poté se modeluje semivariogram na reziduích a po provedení interpolace se odstraněný trend opět přičte).
- **Tip**
 - Všechny tři uvedené druhy krigingu umožňují kromě samotných interpolací počítat i další druhy povrchů - kvantilové a pravděpodobostní mapy.

- 7) Vyzkoušejte EBK
- 8) Použijte nástroj **Extract Multi Values To Points** (do bodové vrstvy teplot se propíše hodnota nadmořské výšky z rastru) k zjištění korelace teploty a nadmořské výšky.

Vyzkoušejte EBK Regression Prediction.

- Source Dataset: vstupní data
 - Dependent Variable: parametr, který interpolujeme (Mean Aug Temp)
 - Raster 1: proměnná, která pomůže zpřesnit interpolaci teplot (Elevation)
- 9) Porovnejte mezi sebou jak EBK a EBKRP, tak předchozí 3 metody krigingu.

Cvičení 2 - CoKriging

- Budeme interpolovat teplotu v okolí Vsetína pomocí jiných proměnných, které máme k dispozici
- Data (Vsetin_CoKriging):
 - Data v polygonové pravidelné čtvercové síti, nicméně postup bude úplně stejný jako u vrstev bodových
 - Jedná se o okolí Vsetína, v atributové tabulce jsou různé atributy, jenž budeme využívat pro CoKriging:
 - **T_avg** - teplota z letního mobilního měření
 - **T_Avg_WIN** - teplota ze zimního mobilního měření
 - **ELEV** - průměrná, minimální a maximální nadmořská výška kvadrátu
 - **NDVI** - průměrná, minimální a maximální hodnota NDVI v kvadrátu
 - procentuální zastoupení propustných povrchů (BSF – budovy, ISF – nepropustné povrchy, PSF – vegetace)

Postup vypracování

- Úprava dat:
 - 1. převést hodnoty „0“ u teplot na „No-Data“ (SHP -> geodatabáze)
 - 2. Field calculator - None
- 1. Průzkumová analýza
- 2. Pomocí nástroje Geostatistical Wizard zvolte metodu interpolace krigování (Kriging/CoKriging) a vyberte aspoň dva nebo více atributů. Prvním bude jedna z měřených teplot, které jsou změřené pouze na malé části území, další atributy zvolte svého uvážení. Použijte metody Ordinary CoKriging, Simple CoKriging a Universal Cokriging (principy metod jsou stejné).

- 3) Vyzkoušejte které atributy dávají lepší a logičtější výsledky. Výsledky vizuálně porovnejte s dostupnými mapovými podklady (Ortofoto, Základní mapy), zdali jsou logicky přesné.
- 4) Na závěr se u každého povrchu nezapoměňte podívat na mapu Prediction Standart Error, stejně jako v minulém cvičení. Podívejte se, kde metody vykazují vyšších hodnot chyb - asi nás příliš nepřekvapí, že nejnižší chyby budou blízko bodů z nichž máme přímá měření.

Popisná statistika bodových objektů

- Otázky, na které se zaměřuji nástroje, které budeme dnes používat:
 - Kde je střed?
 - Jaký je tvar a orientace dat?
- **Central Feature:** Nejvíce středová poloha prvků (body, linie, polygon), vypočtena pomocí např. Euklidovské vzdálenosti.
- **Mean Center:** Identifikuje geografický střed (ze souřadnic X,Y a Z pokud je k dispozici) pro sadu prvků.
- **Median Center:** Identifikuje umístění, které minimalizuje celkovou euklidovskou vzdálenost k prvkům v souboru dat. Odlehlým hodnotám dává menší váhu.
- **Directional Distributions:** Vytváří standardní odchylkové elipsy nebo elipsoidy pro shrnutí prostorových charakteristik geografických prvků: centrální tendence, rozptyl a směrové trendy.

Zadání

- Ve studovaném území byly metodami pozemního spektrometrického měření (Metoda 1) na vybraných odběrných bodech a metodami geofyzikálního leteckého průzkumu (Metoda 2) zjištěny zvýšené koncentrace radioizotopu ^{238}U . Lokality se zvýšenou koncentrací tvoří ve zpracovávaném území dvě dobře odlišitelné oblasti (severní a jižní).
 - 1) Zjistěte, jak se liší hodnoty průměrné polohy pro obě oblasti a to pro data naměřená oběma metodami.
 - 2) Je předpoklad, že lokality jsou vázány na hlavní **geologické** zlomy v oblasti. S využitím směrodatné elipsy odchylek charakterizujte rozdíly ve směrovosti a proměnlivosti obou lokalit.

Postup vypracování

- 1) K vypracování využijte program ArcMap/ArcPRO, sadu nástrojů ArcToolbox – Spatial Statistics Tools – Measuring Geographic Distribution
- 2) Pomocí souboru polygonů U_high.shp nejprve vyberte ze všech měření (U238jtsk.shp) pouze ta se zvýšenou koncentrací.
- 3) Výpočet pro jednotlivé lokality (severní a jižní) provedete tak, že nejprve danou podmnožinu bodů vyberete. Vzniknou tedy dva výběry dat – sever a jih.
- 4) Do výsledné mapy vykreslete polohu průměrného středu pro obě lokality a obě použité metody. Zjistěte praktický rozdíl mezi metodami Central Feature, Median Center a Mean Center. Dále vykreslete hlavní a vedlejší poloosy pro elipsy odchylek (Directional Distribution) obou lokalit. Stručně interpretujte hodnoty vypočtených charakteristik.

5) Postupně vypočtete průměrné středy a elipsy směrodatných odchylek. Poznamenejte si a zjistěte, kde lze vypočíst prostorovou statistiku vypočtených charakteristik a jaké rozšířené funkce jsou k dispozici (zejména váhy u metody Mean Center).

— Cíle

- Budete mít tedy celkově čtyři vsrvtvy na kterých budete provádět zmíněné operace. Vždy jižní a severní lokalitu pro obě dvě metody. Na těchto vrstvách si vyzkoušíte uvedené nástroje a porovnáte (alespoň vizuálně) rozdíly mezi metodami měření.

Seminární práce 2

- Vstupní data:
 - stejná jako při první sem. práci („**Brno_LST_trenovaci.shp**“)
- Zadání:
 - 1) Proveďte podrobnou průzkumovou analýzu datového souboru.
 - 2) Proveďte interpolaci povrchového teplotního ostrova města Brna čtyřmi typy krigování: Ordinary, Simple, Universal a Empirical Bayesian Kriging. Zvolené parametry jednotlivých interpolací, včetně komentáře, uveďte do textu a nastavení podpořte argumenty. Součástí textu budou výsledky všech interpolací a jejich komentář.
 - 3) Stručně popište rozdíly mezi jednotlivými typy krigingu. Přiložte i screenshoty nastavení vašich semivariogramů a uveďte jak jste při jeho tvorbě postupovali (opět stačí stručně, nemusíte popisovat každý parametr).

- 4) Proveďte validaci výsledků pomocí nezávislého datasetu. Tyto výsledky okomentujte a vytvořte k nim vizualizaci.
- 5) Výsledky srovnejte s výstupy první seminární práce, buď graficky anebo tabelárně. Diskutujte vhodnost použitých metod v obou seminárních pracích a doporučte nejvhodnější metodu pro interpolaci teplotního ostrova města.
- 6) Při interpretaci výsledků interpolací a validací vycházejte z dostupných znalostí které máte o území, nebojte se využívat i dalších zdrojů, například podkladové mapy.

- **Struktura textu:**
 - Práce bude jasně strukturovaná (úvod, zpracování, závěr; formální a typografická úprava, číslování příloh, atd.) a je vhodné, abyste popsali každý krok, který jste provedli.
 - Když použijete nějaké nastavení, tak napište proč. Stejně tak u komentářů – když vyslovíte závěr, podpořte jej fakty.
 - Do závěru shrňte poznatky z celé práce.
 - Není nutné dělat pro každý grafický výstup mapu, stačí výsledek interpolace/validace a legenda.