

Manuál

**VYHLÁDÁVANIE KOLOKAČNÝCH
VZOROV**
pre priestorové dáta

Simona BUČKOVÁ
Olomouc 2022



KATEDRA GEOINFORMATIKY
Univerzita Palackého v Olomouci

OBSAH

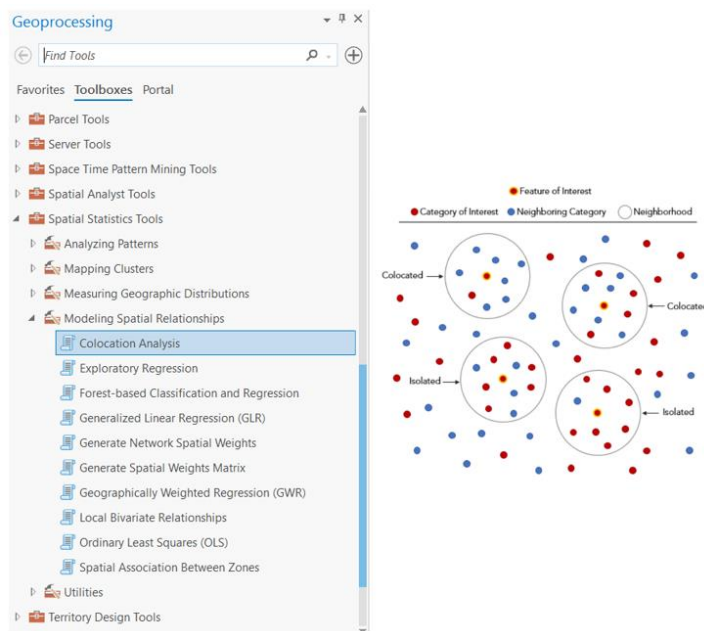
1 Úvod.....	3
1.1 Kolokačná analýza (Colocation Analysis)	3
2 Predpríprava dát.....	6
2.1 Nákladná loď a obchodné plavidlo	6
2.2 Pirátske násilie a podozrivé priblíženie	8
3 Tvorba .swm súboru	9
4 Vyhľadávanie kolokačných vzorov.....	11
4.1 Priestorové váhy zo súboru (<i>Get spatial weights from file</i>)	11
4.2 K najbližších susedov (<i>K Nearest Neighbors</i>)	12
4.3 Vzdialenostné pásmo (<i>Distance Band</i>) a časový interval.....	13
5 Vizualizácia kolokačných vzorov.....	15
6 Interpretácia kolokačných vzorov.....	18
6.1 Nákladná loď a obchodné plavidlo	18
6.2 Pirátske násilie a podozrivé priblíženie	22

Tento návod je prílohou diplomovej práce s názvom
Aplikácia vyhľadávania kolokačných vzorov v priestorových dátach.

Autorom všetkých obrázkov v návode je S. Bučková.

1 ÚVOD

Tento manuál ukazuje na konkrétnych dátach praktické použitie nástroja Colocation Analysis zo sady nástrojov Spatial Statistics (viď Obr. 1), je určený pre akéhokoľvek užívateľa programu ArcGIS Pro. Okrem základného nastavenia ukazuje aj varianty voľby parametrov, demonštruje výsledky analýzy a interpretáciu výsledkov. V návode sa bude pracovať s dátami s názvom ASAM, celým názvom Anti-shipping Activity Messages, čiže správy o aktivitách smerované proti námorníctvu. Dáta sú dostupné na stránke Maritime Safety Information¹, avšak je nutné podotknúť, že sa aktualizujú týždenne.



Obr. 1 Umiestnenie nástroja Colocation Analysis a nápoveda (zdroj: ArcGIS Pro)²

1.1 Kolokačná analýza (Colocation Analysis)

Nástroj priestorovej štatistiky meria vzorce priestorových kolokácií a počíta lokálny kolokačný kvocient. Určuje pre všetky prvky záujmu či sú prvky susednej kategórie viac, alebo menej prítomné pri porovnaní s celkovou priestorovou distribúciou. Akceptovaná je len bodová reprezentácia, buď v jednom datase, alebo v dvoch separátnych.

Vzťah medzi skúmanými kategóriami nie je symetrický, v prípade obrátenia poradia vstupných bodov, bude výsledok rozdielny. Veľký dôraz je potrebné klásť pri výbere kategórií, ktoré môžu byť v potenciálnom priestorovom vzťahu. Pri nesprávnom výbere vhodných kategórií sa môžu stratiť informácie, ktoré sú relevantné pre výskum a taktiež sa stráca informácia o vzťahu k iným okolitým kategóriám prvkov. Priestorový vzťah je možné definovať tromi spôsobmi:

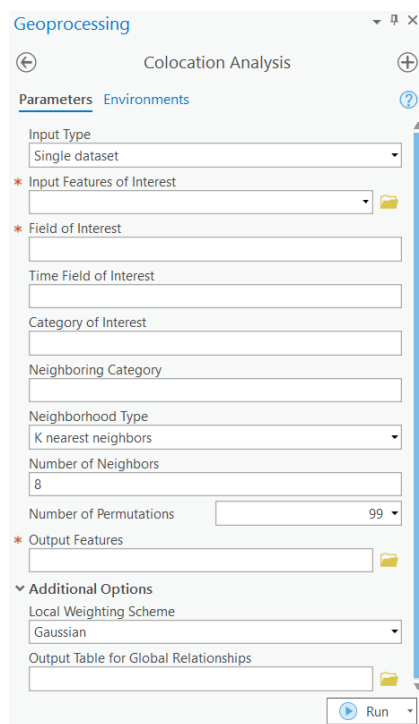
- K najbližších susedov
- Vzdialenostné pásmo
- Matica váh

Jednotlivé metódy pre definovanie priestorového vzťahu sú vysvetlené v ďalších kapitolách. Základné nastavenie nástroja je možné vidieť na Obr. 2. V prvom poli typ vstupu (*Input Type*) sa volí či sa jedná o jeden dataset s oboma skúmanými kategóriami, alebo je každá kategória uložená samostatne. Následne sa nalistuje

¹ <https://msi.nga.mil/Piracy>

² <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/colocationanalysis.htm>

vrstva s prvkami záujmu (*Input Features of Interest*) a z atribútovej tabuľky sa zvolí stĺpec, ktorý kategóriu reprezentuje (*Field of Interest*). Taktiež je možné zvoliť aj časové pole (*Time Field of Interest*), ak sa v dátach nachádza. Potom je potrebné vybrať hodnotu zo stĺpca, ktorý určuje skúmanú kategóriu (*Category of Interest*) a zvoliť, ktoré prvky budú susedné (*Neighboring Category*). Ďalším krokom je voľba metódy spomínanej vyššie a jej hodnôt. V prípade matice váh sa nalistuje externý súbor, ktorý je možné zostrojiť pomocou nástroja Generovanie priestorovej matice váh (*Generate Spatial Weights Matrix*). Predposledným krokom je vybratie zložky, alebo geodatabázy, kde bude výsledný vzor uložený (*Output Features*). Posledným krokom je možnosť voľby lokálnej váhovej schémy (*Local Weighting Scheme*). Dostupné sú tri možnosti, a to Bikvadratická (*Bisquare*), Gaussovská (*Gaussian*) a Žiadna (*None*). Voliteľným poľom je tabuľka globálnych vzťahov (*Output Table for Global Relationships*), ktorú je potrebné pomenovať a vybrať jej úložisko.



Obr. 2 Základné nastavenie kolokačnej analýzy pre Single dataset

Čo je potrebné:

- ArcGIS Pro, min. verzia 2.8
- Vstupné dáta

Vstupné vrstvy:

- *Primárna vrstva upravená* – vrstva vstupuje do kolokačnej analýzy
- *Primárna celá vrstva neupravená* – vrstva stiahnutá z MSI
 - ASAM.shp
- *Primárna celá vrstva upravená* – v správnom súradnicovom systéme
 - ASAM_UTM.shp
- *Primárna vrstva upravená* – len skúmané kategórie v oblasti Arabského polostrova
 - CargoShip_MerchantVessel_UTM.shp
 - PirateAssault_SuspiciousApproach_UTM.shp

DÔLEŽITÉ!

Pre rovnaké výsledky je potrebné, aby vstupné dáta boli v rovnakom súradnicovom systéme

V rámci tohoto návodu boli všetky vrstvy uložené do súradnicového systému Yemen NGN 1996 UTM Zone 39N (SRS: 2090).

V zložke sa okrem manuálu nachádza projekt a ďalšie súbory:

V zložke Colocation_Analysis sa nachádza projekt programu ArcGIS Pro s názvom „Projekt“. Okrem toho sú dostupné aj ďalšie zložky, a to „Shp“, „Grafy“, „Obrazky“ a geodatabáza „data“.

- *Neupravené vstupné vrstvy* – vrstvy v pôvodnom súradnicovom systéme v zložke „Shp“
 - *CargoShip_MerchantVessel.shp*
 - *PirateAssault_SuspiciousApproach.shp*
- *Vstupné súbory* – súbory .swm získané z nástroja Generate spatial Weights Matrix
 - *CM_FixedDistance_5oS_10km.swm*
 - *CM_SpaceTimeWindow_5R_50km.swm*
 - *CM_KNearestNeighbors_1oS.swm*
 - *PS_Inverse_Distance_1oS_100km.swm*
 - *PS_SpaceTimeWindow_5R_50km.swm*
 - *PS_KNearestNeighbors_1oS.swm*
- *Výstupné vrstvy* – triedy prvkov v geodatabáze „data“ získané z nástroja Colocation Analysis
 - *CM_FixedDistance_5oS_10km_CA*
 - *CM_SpaceTimeWindow_5R_50km_CA*
 - *CM_KNearestNeighbors_1oS_CA*
 - *CM_DistanceBand_5R_50km_CA*
 - *PS_Inverse_Distance_1oS_100km_CA*
 - *PS_SpaceTimeWindow_5R_50km_CA*
 - *PS_KNearestNeighbors_1oS_CA*
 - *PS_KNearestNeighbors_CA_1oS_CA*
 - *PS_Distance_band_5R_100km_CA*
- *Grafy* – obrázkové (.png) súbory exportované z výstupov analýzy v zložke „Grafy“
 - *CM_FixedDistance_5oS_10km_CA.png*
 - *CM_SpaceTimeWindow_5R_50km_CA.png*
 - *CM_KNearestNeighbors_1oS_CA.png*
 - *CM_DistanceBand_5R_50km_CA.png*
 - *PS_Inverse_Distance_1oS_100km_CA.png*
 - *PS_SpaceTimeWindow_5R_50km_CA.png*
 - *PS_KNearestNeighbors_1oS_CA.png*
 - *PS_KNearestNeighbors_CA_1oS_CA.png*
 - *PS_Distance_band_5R_100km_CA.png*

Vysvetlivky:

- CM – Cargo Ship a Merchant Vessel
- PS – Pirate Assault a Suspicious Approach
- 10S – desať susedných prvkov
- 5R – päť rokov

2 PREDPRÍPRAVA DÁT

Pred samotným hľadaním kolokačných vzorov je nutné dátové sady upraviť. V zložke „Shp“ je dostupná aj vrstva po úprave, avšak jej príprava je taktiež podstatná.

2.1 Nákladná loď a obchodné plavidlo

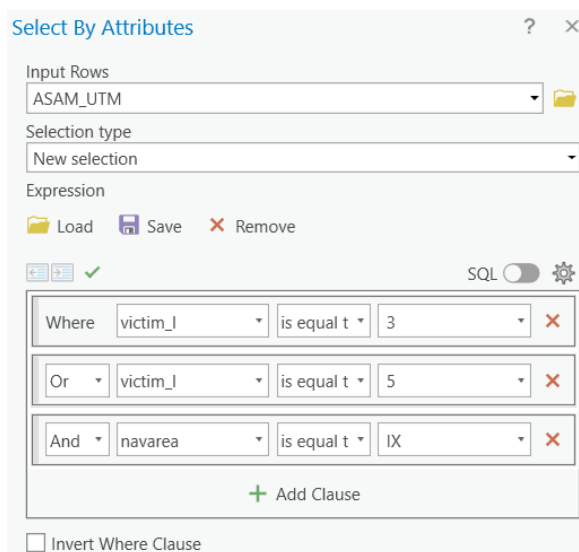
Ako už bolo spomenuté v prvej kapitole, dáta je potrebné nastaviť na rovnaký súradnicový systém. Dáta sa rozprestierajú v okolí viacerých štátov, no podstatná časť dát je v okolí Jemenu, a preto bol zvolený pás UTM práve pre tento štát. Prvým krokom je pridanie nových stĺpcov **X** a **Y** do atribútovej tabuľky. Po kliknutí pravým tlačidlom myši na vrstvu sa otvorí panel, kde je jednou z možností atribútová tabuľka (*Attribute Table*). Po jej otvorení sa v hornej lište možností nachádza tlačidlo pridať (*Add*). Po pomenovaní nového stĺpca, je potrebné nastaviť dátový typ (*Data Type*) na „**Double**“ a číselný formát (*Numeric Format*) na „**Numeric**“. Otvorí sa nové okno, kde je možné zmeniť zarovnanie (*Alignment*) vpravo s hodnotou „**o**“. V hornej lište nástroja je nutné zmeny uložiť kliknutím na tlačítko Uložiť (*Save*).

Po kliknutí pravým tlačidlom myši na novo vytvorený stĺpec súradnice X sa zobrazí panel s možnosťou vypočítať geometriu (*Calculate Geometry*). V poli vlastnosť (*Property*) je voľba X súradnice (*Point x-coordinate*) a ako súradnicový systém (*Coordinate System*) stačí ponechať voľbu súčasnej mapy (*Current Map*). Rovnakým spôsobom sa vypočíta aj druhá súradnica Y.

Ďalším krokom je použitie nástroja definovanie projekcie (*Define Projection*), ktorý je možné nájsť v záložke „**Geoprocessing**“. Vstup tvorí zatiaľ jediná vrstva, a to „**ASAM**“. Ako súradnicový systém bude zvolený „**Yemen_NGN_1996_UTM_Zone_39N**“ s kódom „**2090**“.

Pre dokončenie nastavenie správneho súradnicového systému je nutné použiť nástroj previesť súradnicový zápis (*Convert Coordinate Notation*). Vstupnou tabuľkou (*Input Table*) je vrstva „**ASAM**“, následne novo vypočítané súradnice **X** a **Y** a posledným krokom je nový formát súradníc (*Output Coordinate Format*) s voľbou UTM zóny (*UTM zones*). Po kliknutí na tlačidlo spustiť (*Run*) sa do atribútovej tabuľky pridá nové pole s názvom „**UTM_ZONES**“. Novo vzniknutú vrstvu uložíme s názvom „**ASAM_UTM**“. Teraz už len stačí zmeniť aj súradnicový systém projektu, a to tak, že po kliknutí v pravej časti programu na obsah (*Contents*) a možnosti (*Properties*), kde je možnosť súradnicového systému s kódom 2090. (*Coordinate Systems*).

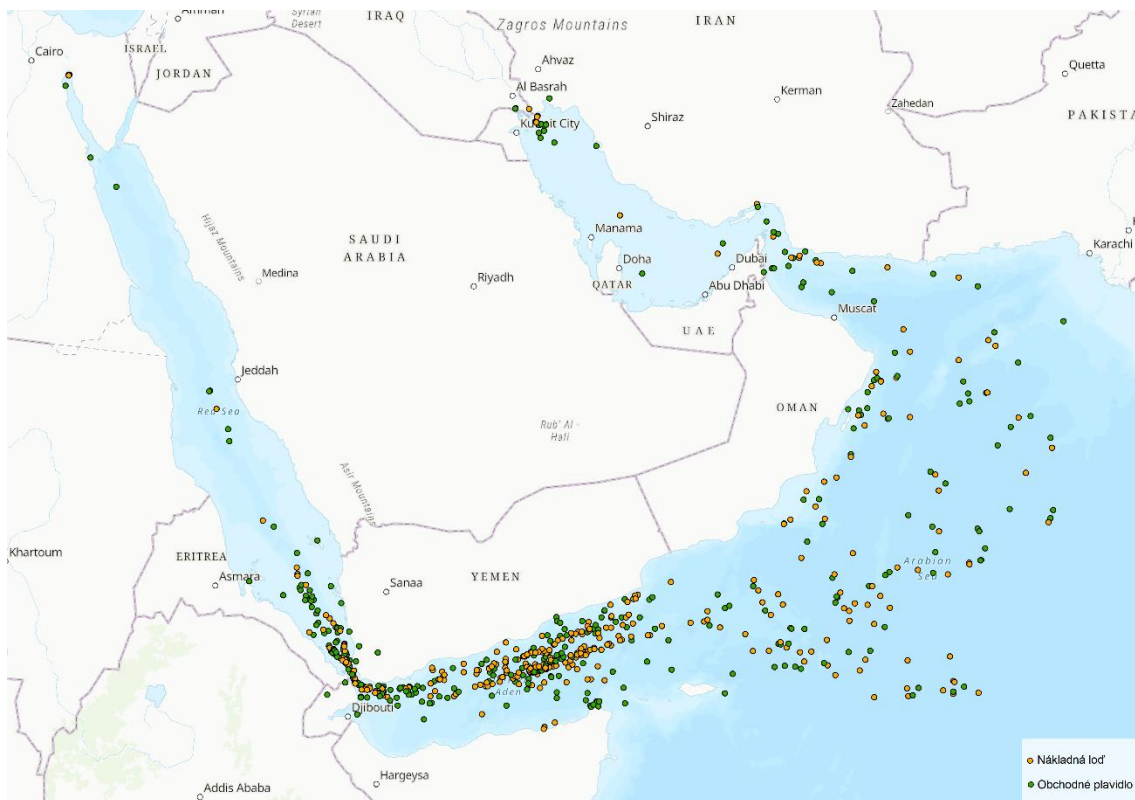
Teraz už len stačí vybrať konkrétne kategórie, ktoré sa ukrývajú v stĺpci „**victim_I**“ (typ obeť). Stačí použiť nástroj výber podľa atribútov (*Select By Attributes*) a vybrať ako vstup vrstvu „**ASAM_UTM**“ a po kliknutí na tlačidlo nový výraz (*New expression*) zadať následné vlastnosti (viď Obr. 3).



Obr. 3 Výber podmnožiny dát podľa atribútov

Po úprave zostalo len 697 záznamov s vybranými typmi obetí – Cargo Ship a Merchant Vessel.

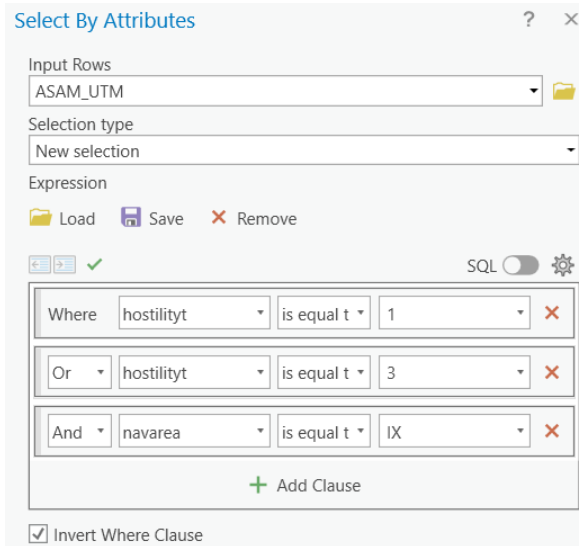
Študovaná oblasť spadá do navigačnej oblasti s hodnotou „IX“, a preto bola do výberu pridaná táto hodnota. Teraz, keď sú vybrané len konkrétne záznamy, stačí len kliknúť pravou myšou na pôvodnú vrstvu „ASAM_UTM“ a zvoliť možnosť výber (*Selection*) a zvoliť vytvoriť vrstvu z vybraných prvkov (*Make Layer From Selected Features*). Novú vrstvu stačí premenovať po dvojkliku na názov „CargoShip_MerchantVessel_UTM“ (viď Obr. 4).



Obr. 4 Oblasť skúmaných dát – nákladná loď a obchodné plavidlo

2.2 Pirátske násilie a podozrivé priblíženie

Rovnaký postup je opakovateľný aj pre tvorbu druhej vstupnej vrstvy s názvom „**PirateAssault_SuspiciousApproach_UTM**“. Jediná zmena spočíva pri použití nástroja výber podľa atribútov, kde namiesto „Victim List“ bude použitá kategória „**hostilityt**“ (typ nepriateľov) s hodnotami „1“ (Pirate Assault) a „3“ (Suspicious Approach) (viď Obr. 5).

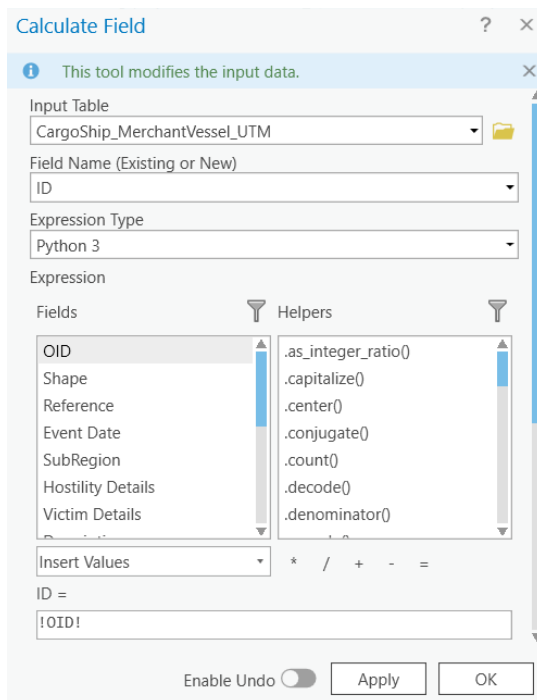


Obr. 5 Výber podmnožiny dát podľa atribútov


Po úprave zostalo len 1 226 záznamov s vybranými typmi nepriateľov – Pirate Assault a Suspicious Approach.


3 TVORBA .SWM SÚBORU

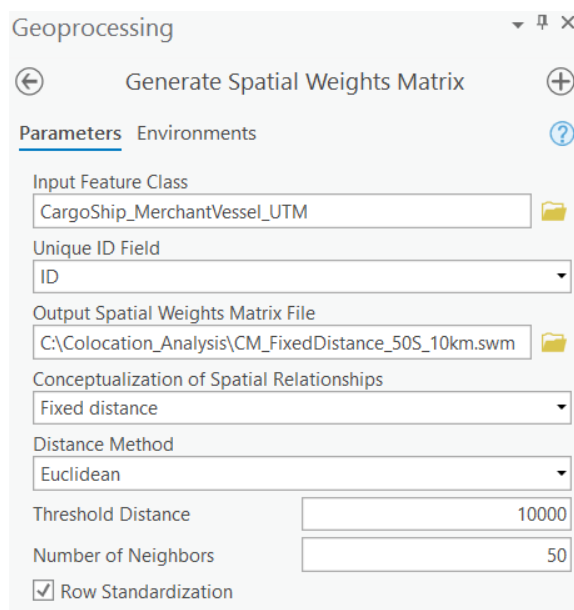
Jednou z možností definovania priestorového vzťahu je použitie externého súboru. Pre tvorbu tejto priestorovej matice váh slúži nástroj generovanie priestorovej matice váh (*Generate Spatial Weights Matrix*). Skúmané prvky neboli rozdelené do dvoch samostatných vrstiev, práve kvôli tomuto nástroju, pretože vstupom musia byť ako prvky záujmu, tak aj susedné prvky. Čo je však ešte potrebné vyriešiť je identifikátor záznamov. Dátová vrstva ho síce má, ale nástroj ho neakceptuje. Preto stačí vytvoriť nový stĺpec pomocou pridať (*Add*) a vybrať dátový typ „short“. Následne stačí pravým tlačidlom kliknúť na novo vytvorený stĺpec s názvom „ID“ a zvoliť počítať pole (*Calculate Field*), kde stačí zadať nasledovné parametre (viď Obr. 6):



Obr. 6 Tvorba stĺpca s ID identifikátorom

Potom už je možné pristúpiť k samotnému generovaniu súboru, vstupnou triedou (*Input Feature Class*) je vrstva „**CargoShip_MerchantVessel_UTM**“ so záznamami o nákladnej lodi (3 - Cargo Ship) a obchodnom plavidle (5 - Merchant Vessel). Je však nutné nalistovať vrstvu pomocou ikony , inak nástroj ohlási chybu. Ďalším poľom je novo vzniknuté ID a následne miesto, kde bude súbor uložený. Parametre pre konceptualizáciu priestorového vzťahu sa nastavujú ako je uvedené na Obr. 7. Názov výstupného súboru je zložený z metódy a hodnôt parametrov. Napríklad „**CM_FixedDistance_50S_10km.swm**“ znamená použitie metódy *Fixed Distance* s 50 susednými prvkami a medznou vzdialenosťou desiatich kilometrov.

V prípade, že nástroj prebehne, no zobrazí sa varovanie v podobe , to znamená, že niektoré prvky nemajú v definovanom okolí prvky susednej kategórie.



Obr. 7 Generovanie súboru matice priestorových váh

Takto bude zostrojených aj ďalších päť súborov definujúcich priestorové vzťahy medzi skúmanými kategóriami a parametrami (viď Tab. 1).

Tab. 1 Parametrizácia súborov .swm

názov súboru	metóda	počet susedov	vzdialenosť	časový interval
<i>CM_FixedDistance_50S_10km.swm</i>	Fixed Distance	50	10 km	-
<i>CM_SpaceTimeWindow_5R_50km.swm</i>	Space Time Window	-	50 km	5 rokov
<i>CM_KNearestNeighbors_10S.swm</i>	K Nearest Neighbors	10	-	-
<i>PS_Inverse_Distance_10S_100km.swm</i>	Inverse Distance	10	100 km	-
<i>PS_SpaceTimeWindow_5R_50km.swm</i>	Space Time Window	-	50 km	5 rokov
<i>PS_KNearestNeighbors_10S.swm</i>	K Nearest Neighbors	10	-	-

4 VYHLÁDÁVANIE KOLOKAČNÝCH VZOROV

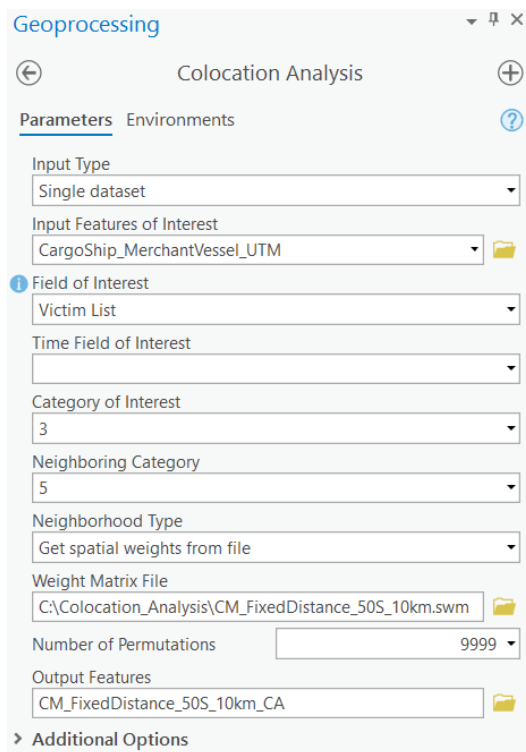
Teraz už možné pristúpiť k samotnému nástroju kolokačnej analýzy, (*Colocation Analysis*). Nástroj je možné nájsť pomocou hľadacieho okna, alebo po kliknutí na toolbox priestorovej štatistiky (*Spatial Statistics*), kde sa následne otvorí okno s týmto nástrojom. V nasledujúcich kapitolách sú popísané analýzy s rôznym nastavením parametrov.

4.1 Priestorové váhy zo súboru (*Get spatial weights from file*)

Prvým poľom je typ vstupu (*Input Type*), čiže či sa jedná o jeden dataset, alebo dva. Dostupné sú možnosti:

- Single dataset – jedna vrstva obsahuje obidve skúmané kategórie
- Datasets without categories – dve samostatné vrstvy
- Two datasets – dve vrstvy, ktoré v atribútovej tabuľke obsahujú skúmané kategórie

Dáta plavidiel a útokov sú v jednej dátovej vrstve, ktorá reprezentuje obidve kategórie, preto je nutné zvoliť prvú možnosť „**Single dataset**“. Ďalším krokom je voľba prvkov záujmu (*Input Features of Interest*), čo je vrstva „**CargoShip_MerchantVessel_UTM**“. Následne je potrebné zvoliť skúmanú kategóriu (*Category of Interest*), čo je voľba „**3**“ (Cargo Ship) a susednú kategóriu (*Neighboring Category*) s hodnotou „**5**“ (Merchant Vessel). Teraz je nutné vybrať metódu „**Get spatial weights from file**“ a nalistiť konkrétny .swm súbor. Medzi posledné kroky patrí nastavenie presnosti čiže počtu permutácií (*Number of Permutations*) s hodnotou „**9999**“ a uloženie výstupnej vrstvy ako „**FixedDistance_50S_10km_CA**“ vid' Obr. 8.



Obr. 8 Kolokačná analýza s metódou priestorových váh

Rovnakým spôsobom je možné zostrojiť ďalšie kolokačné vzory, ich parametre sú zhrnuté v Tab. 2:

Tab. 2 Parametrizácia nástroja Colocation Analysis

názov súboru	metóda	počet susedov	vzdialenosť	časový interval	počet permutácií
CM_FixedDistance_50S_10km_CA	Fixed Distance	50	10 km	-	9999
CM_SpaceTimeWindow_5R_50km_CA	Space Time Window	-	50 km	5 rokov	9999
CM_KNearestNeighbors_10S_CA	K Nearest Neighbors	10	-	-	9999
CM_DistanceBand_5R_50km_CA	Distance Band	-	50 km	5 rokov	9999
PS_Inverse_Distance_10S_100km_CA	Inverse Distance	10	100 km	-	9999
PS_SpaceTimeWindow_5R_50km_CA	Space Time Window	-	50 km	5 rokov	9999
PS_KNearestNeighbors_10S_CA	K Nearest Neighbors	10	-	-	9999
PS_KNearestNeighbors_CA_10S_CA*	K Nearest Neighbors	10	-	-	9999
PS_Distance_band_5R_100km_CA	Distance Band	-	100 km	5 rokov	9999

* Kolokačný vzor len z nástroja Colocation Analysis, bez externého súboru.

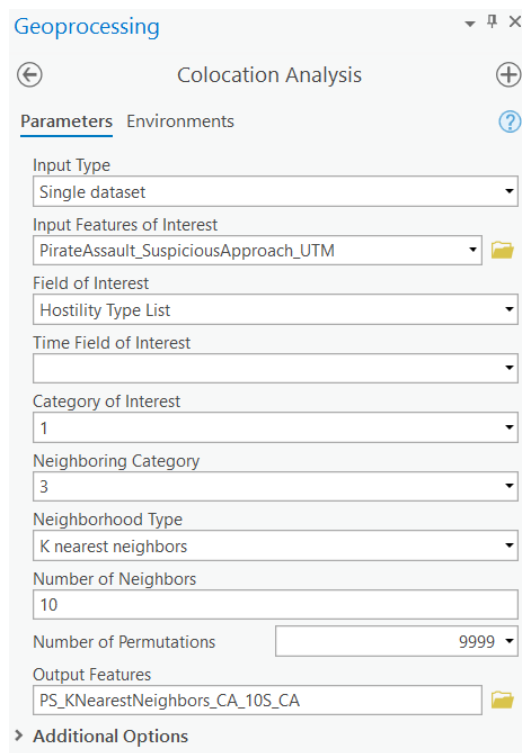
Doteraz však bola využitá len možnosť „**Get spatial weights from file**“, no primárne kolokačná analýza disponuje ďalšími dvoma metódami, a to:

- K najbližších susedov
- Vzďialenosťné pásmo

4.2 K najbližších susedov (*K Nearest Neighbors*)

Susedných prvkov bude zahrnutých presne toľko, koľko užívateľ zadá (za predpokladu, že taký počet prvkov je v dátovej vrstve dostupných). Počet susedov je špecifikovaných parametrom počtu susedných prvkov (*Number of Neighbors*). Vplyv suseda v analýze je vážený na základe vzdialenosti od najvzdialenejšieho suseda. V prípade, že užívateľ nezadá žiadnu hodnotu, bude použitých východných osem prvkov v okolí. V tomto návode bola však využitá táto metóda ako v rámci nástroja pre generovanie externého .swm súboru, tak aj v samotnom nástroji. Je ale dôležité si uvedomiť, že výsledky aj pri totožnom nastavení K najbližších susedov v primárnom nástroji (*Colocation Analysis*) a v nástroji pre generovanie matice s metódou K najbližších susedov nie sú rovnaké, avšak sú si podobné.

Práve preto boli pre typy útokov zostrojené obe možnosti. Stačí zvoliť skúmanú vrstvu „**PirateAssault_SuspiciousApproach_UTM**“ ako typ vstupu (*Input Type*), v poli záujmu (*Field of Interest*) kliknúť na „**Hostility Type List**“ a zvoliť kategórie. Posledným krokom je voľba správnej metódy a počet susedných prvkov (*Number of Neighbors*). Vzhľadom na to, že pri generovaní .swm súboru s rovnakou metódou bolo použitých desať prvkov, taktiež nastavíme hodnotu „**10**“ (viď Obr. 9). Potom už len stačí spustiť analýzu pomocou tlačidla spustiť (*Run*).



Obr. 9 Kolokačná analýza s metódou K najbližších susedov

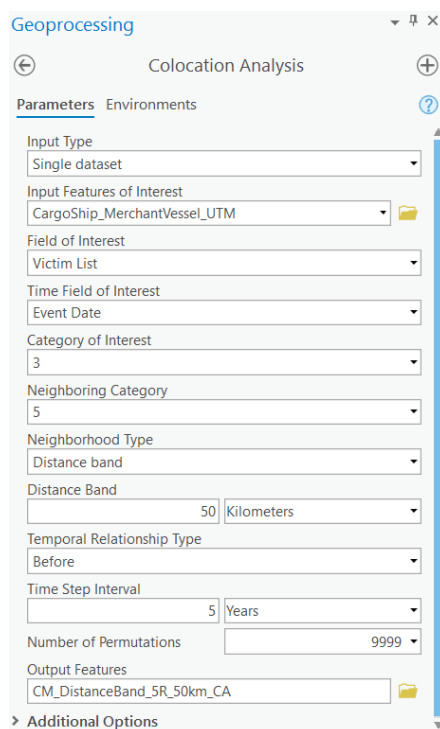
4.3 Vzďalienostné pásmo (*Distance Band*) a časový interval

Každý prvok bude analyzovaný v kontexte susedných prvkov. Parameter vzďalienostného pásma (*Distance Band*) určí susedné prvky v intervale kritickej vzďalienosti. Prvky v susednej kategórii, ktoré sa v ňom nachádzajú, dostanú váhu jeden a majú vplyv na výpočet kolokačného vzoru. Ostatné prvky nemajú vplyv a získavajú váhu nula.

Pre obidve študované vstupné vrstvy „**CargoShip_MerchantVessel_UTM**“, „**PirateAssault_SuspiciousApproach_UTM**“ bola použitá aj táto metóda definovania priestorového vzťahu. Navyše bol využitý aj stĺpec „**Event Date**“, takže vzťah bol skúmaný aj z časového hľadiska. Je však logické, že výsledný kolokačný vzor nebude až taký pestrý, čo sa týka významne kolokovaných bodov. Dôvodom je pomerne prísna podmienka, ako vzďalienosti, tak aj časového intervalu.

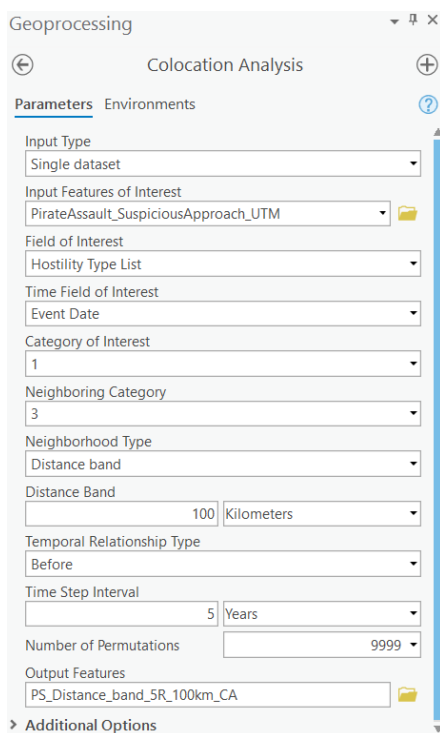
Postup je nasledovný, v prvom rade vybrať druh vstupu (*Input Type*) ako „**Single**“ a vybrať prvky záujmu (*Input Feature of Interest*), a to „**CargoShip_MerchantVessel_UTM**“. Potom je potrebné zvoliť, ktorá z dostupných atribútových vlastností bude analyzovaná, čiže pole záujmu (*Field of Interest*) ako „**Victim List**“. Nižšie je nutné zvoliť kategóriu záujmu (*Category of Interest*) s hodnotou „**3**“ (Cargo Ship) a susednú kategóriu (*Neighboring Category*), kde je hodnota „**5**“ (Merchant Vessel).

Ďalej je potrebné zvoliť správnu metódu, zadať vzďalienosť „**50**“ a vo vedľajšom poli vybrať jednotku vzďalienosti „**Kilometers**“. Potom v poli časové pole záujmu (*Time Field of Interest*) kliknúť na stĺpec s časom „**Even Date**“ a nižšie nastaviť interval časového kroku (*Time Step Interval*) na hodnotu „**5**“ a jednotku „**Years**“. Čo sa týka časového okna, bude skúmané retrospektívne, čiže aké udalosti sa udiali päť rokov dozadu, preto v poli typ časového vzťahu (*Temporal Relationship Time*) bude zvolená možnosť „**Before**“. Počet permutácií nie je potreba od predošlých nastavení meniť, stačí ponechať najvyššiu hodnotu. Posledný krokom je uloženie výslednej vrstvy ako „**CM_DistanceBand_5R_50km_CA**“. Parametrizácia je taktiež zhrnutá na Obr. 10.



Obr. 10 Kolokačná analýza s metódou vzdialenostného pásma a časovým intervalom

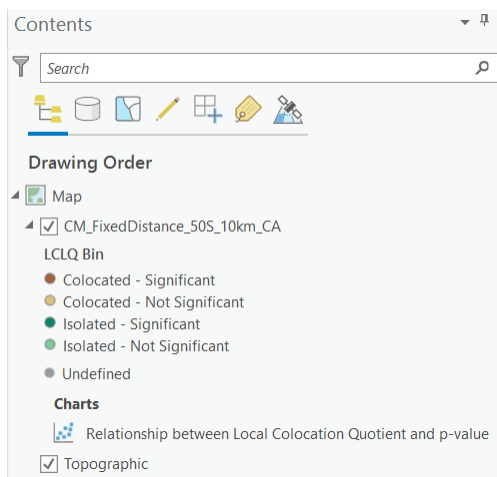
Veľmi podobným spôsobom je možné skonštruovať aj kolokačný vzor pre typy útokov „*PS_Distance_band_5R_100km_CA*“ (viď Obr. 11).



Obr. 11 Kolokačná analýza s metódou vzdialenostného pásma

5 VIZUALIZÁCIA KOLOKAČNÝCH VZOROV

Výsledok nástroja Colocation Analysis sa po vygenerovaní objaví v ľavej časti programu ako nová vrstva. Tú je možné nájsť v oblasti obsahu (Contents), ako je možné pozorovať na Obr. 12. Pomocou krížika vedľa názvu vzoru je možné vrstvu vypnúť a zapnúť.



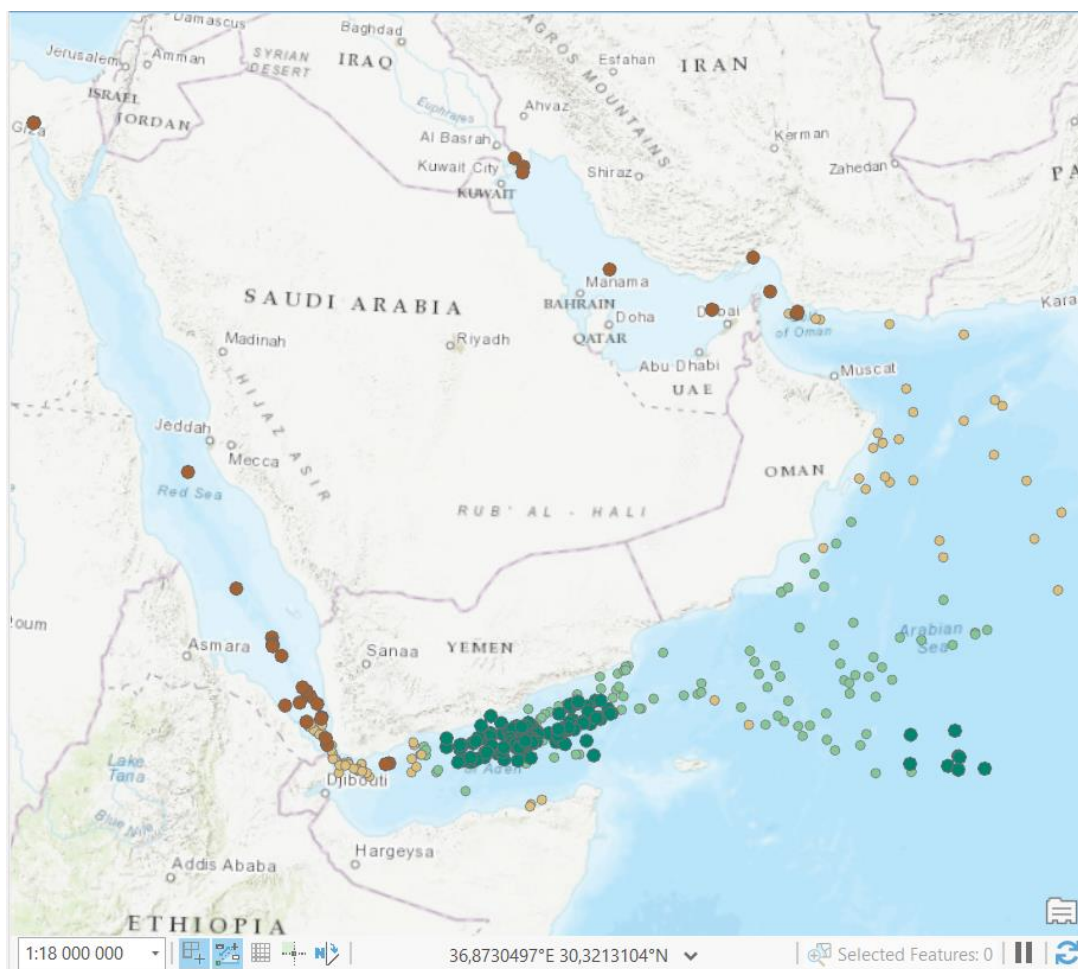
Obr. 12 Oblasť obsahu projektu v ArcGIS Pro

Analýza rozdelí skúmané body vo vstupnej vrstve do piatich kategórií na základe spočítanej hodnoty lokálneho kolokačného kvocientu (Local Colocation Quocient, LCLQ) na:

- Kolokované – Významne
- Kolokované – Nevýznamne
- Izolované – Významne
- Izolované – Nevýznamne
- Nedefinované

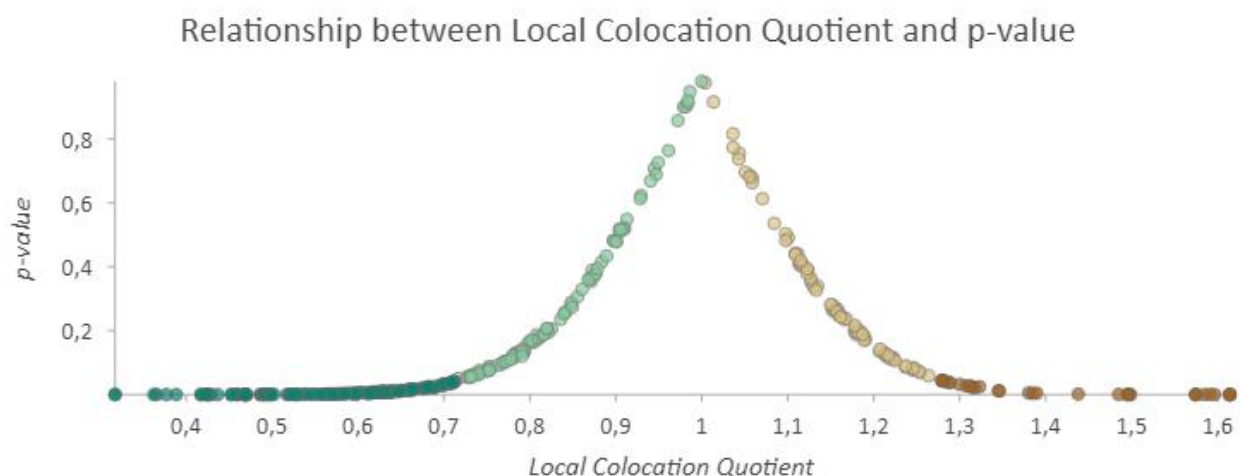
Nemusia sa však všetky kategórie vo výslednej vrstve objaviť, čo ale nie je chyba. Každopádne najlepšie výsledky sú pri čo najväčšom počte významne kolokovaných bodov, no všetko závisí od vstupných dát a konkrétnej parametrizácie. Netreba však zabúdať na fakt, že sa môžu objaviť falošné vzory, ktoré sú síce významne kolokované, no nedávajú zmysel. Identifikácia takéhoto falošného výsledku nie je vždy jednoduchá, avšak keď sa užívateľ rozumie do študovanej problematiky, je to o niečo jednoduchšie. Každopádne momentálne neexistuje odporúčenie, ktoré by pomohlo tieto falošné vzory nájsť.

Výsledné priestorové rozdelenie bodov môže vyzerat' napríklad ako na Obr. 13. Ďalšie výsledky sú dostupné v kapitole päť.



Obr.13 Kolokačný vzor – metóda fixnej vzdialenosti s desiatimi susednými prvkami a medznou vzdialenosťou desať kilometrov
CM_FixedDistance_5oS_10km_CA

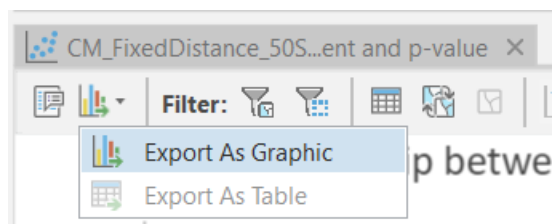
Ako je možné vidieť v oblasti obsahu (*Contents*), okrem samotnej vrstvy kolokačného vzoru nástroj vytvoril aj graf s názvom „Relationship between Local Colocation Quocient and p-value“ (viď Obr. 14). Je ho možné otvoriť dvojklikom, alebo po kliknutí pravým tlačidlom myši s možnosťou otvoriť (*Open*). Podstata grafu spočíva v tom, že zobrazuje vzťah lokálnych kolokačných kvocientov na ose X a ich príslušnú p-hodnotu (*p-value*) – *hladinu významnosti* na ose Y, ktorá sa odvíja od hodnoty počtu permutácií. Vo všetkých analýzach bola použitá najvyššia hodnota 9999, kedy najmenšia možná p-hodnota je rovná 0,0002.



Obr. 14 Graf vzťahu medzi LCLQ a p-hodnotou

Graf ukazuje hodnoty kolokačných kvocientov všetkých bodov vstupnej kategórie a ich hodnotu p-value. Hodnota LCLQ menšia než 1 znamená izolovaný bod, hodnota vyššia než 1 znamená bod kolokovaný. Nižšia hodnota p-value (menšia než 0,05) znamená významnú kolokáciu alebo izoláciu. Vyššia hodnota p-value (nad 0,05) znamená nevýznamnú kolokáciu alebo izoláciu.

Graf je možné aj exportovať a uložiť do ľubovoľného zložky. Stačí len kliknúť na ikonu grafu (viď Obr. 15) a zvoliť možnosť exportovať ako grafiku (*Export As Graphic*). Dostupné sú tri možnosti, a to ako .svg, .jpg a .png. V zložke je priečinok Grafy, kde sú všetky grafy z výstupných vrstiev kolokačnej analýzy.



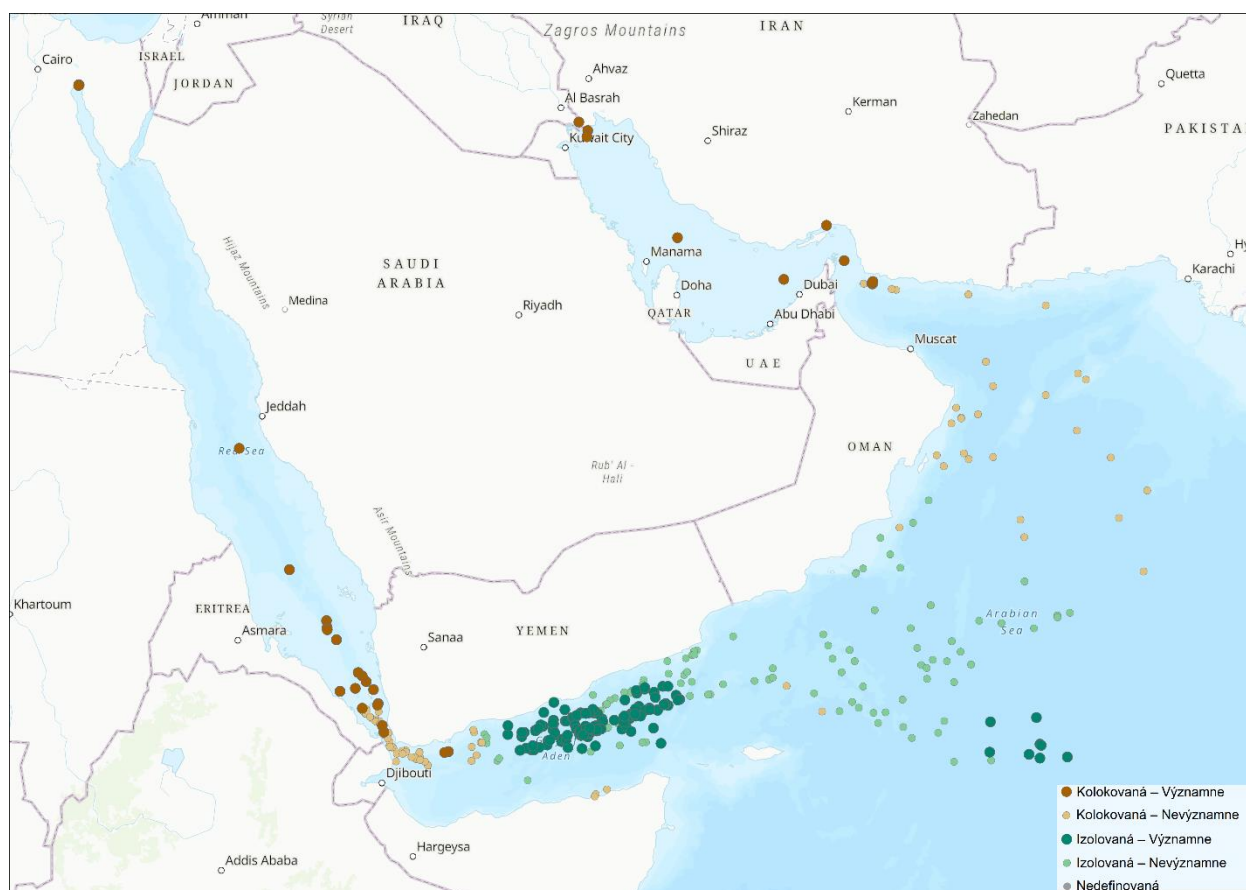
Obr. 15 Export grafu vrstvy kolokačnej analýzy

6 INTERPRETÁCIA KOLOKAČNÝCH VZOROV

6.1 Nákladná loď a obchodné plavidlo

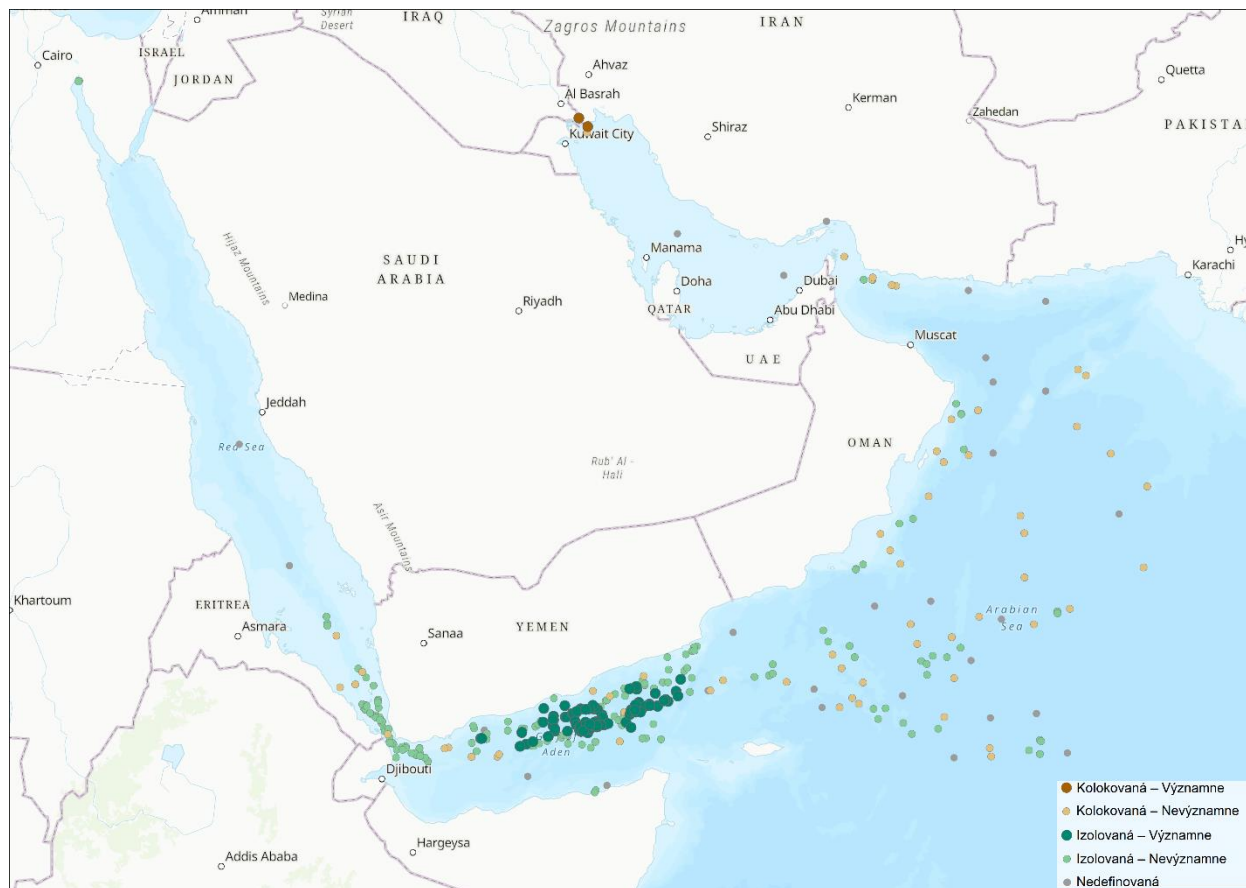
Ako už bolo zmienené, výsledný kolokačný vzor je veľmi závislý na parametrizácii nástroja. Tá je totiž na užívateľovi, ktorý by mal mať znalosti v skúmanej problematike. Samozrejme netreba zabúdať na konkrétne dáta, od ktorých sa výsledok odvíja taktiež. V tomto prípade bola vyšetrovaná úvaha či v okolí napadnutej nákladnej lode sa vyskytuje záznam s obchodným plavidlom.

Medzi výsledkami metód fixnej a inverznej vzdialenosti nie je žiaden rozdiel, samozrejme za predpokladu, že sú pri parametrizácii použité rovnaké hodnoty. Výsledky vyzerajú v tomto nastavení zaujímavo, v miestach zálivov, kde dochádza k zúženiu otvoreného oceánu a objavuje sa pevnina. Naopak v oblasti Arabského mora je väčšina lodí vyhodnotených za izolovaných (viď Obr. 16). Druhou rozsiahlou oblasťou významne izolovaných bodov je Adenský záliv, ktorého lode boli častokrát označované za izolované. Každopádne z celkového hľadiska je kolokovaných len niečo málo než 30 % všetkých nákladných lodí, ale je možné konštatovať, že sú isté oblasti, kde existuje významný vzťah medzi týmito plavidlami – obeťami útokov.



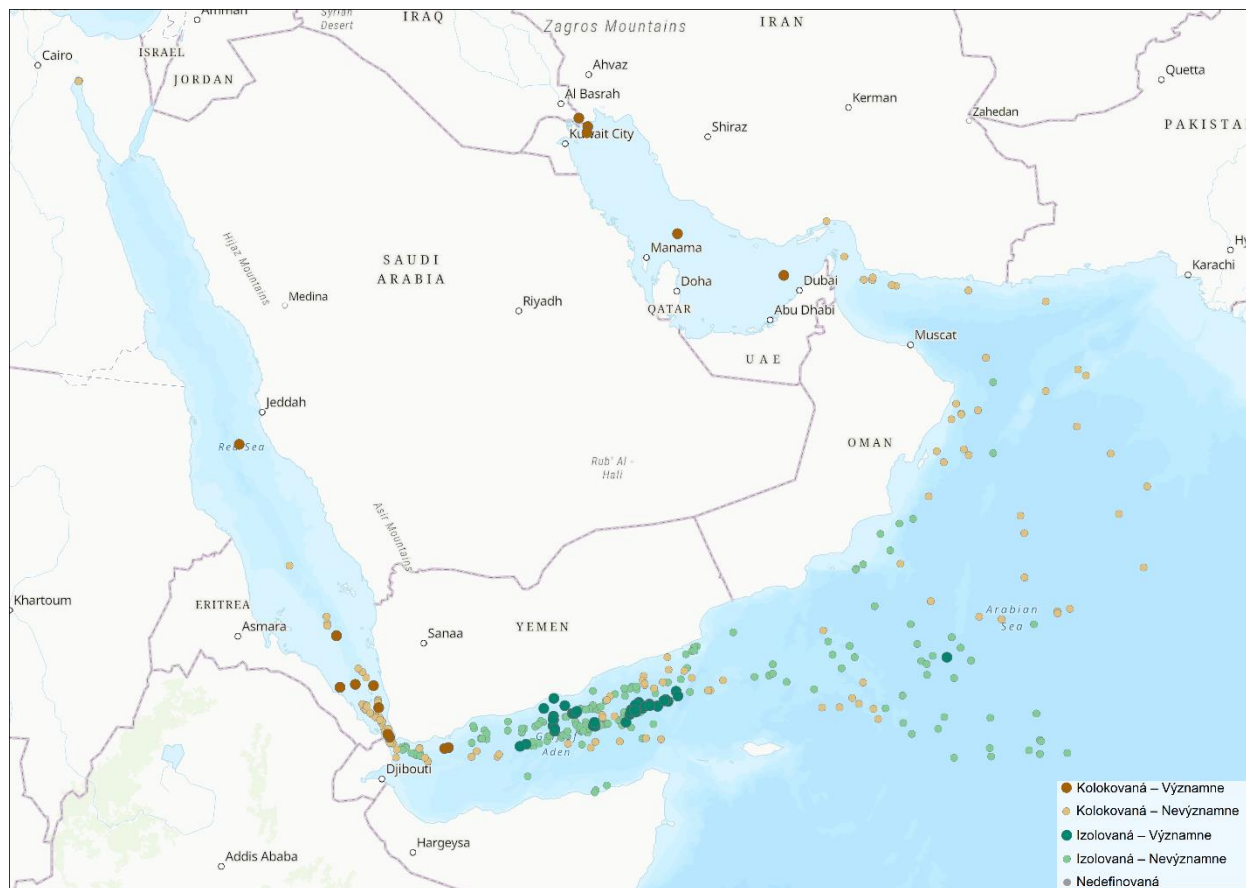
Obr. 16 Kolokačný vzor – metóda fixnej vzdialenosti s desiatimi susednými prvkami a medznou vzdialenosťou desať kilometrov
CM_FixedDistance_50S_10km_CA

Pri použití metódy časopriestorového okna do analýzy okrem priestoru vstupuje aj čas, to znamená, že výsledok je výrazne prísnejší. Aj pri pomerne veľkom časovom intervale piatich rokov je veľká časť záznamov izolovaných (viď Obr. 17). Štandardne významne izolované lode sú v oblasti Adenského zálivu, každopádne vzhľadom na čas sa v Arabskom mori objavujú kolokované body. Tie sú vo väčšine výsledkov izolované, no taktiež sú vo väčšine nastavení hodnotené len z hľadiska priestoru. Z celkového hľadiska je kolokovaných len necelých 17 % lodí a významne kolokovaných je menej než 1 %. To znamená, že v časovom intervale piatich rokov medzi skúmanými plavidlami nie je až taký výrazný vzťah.



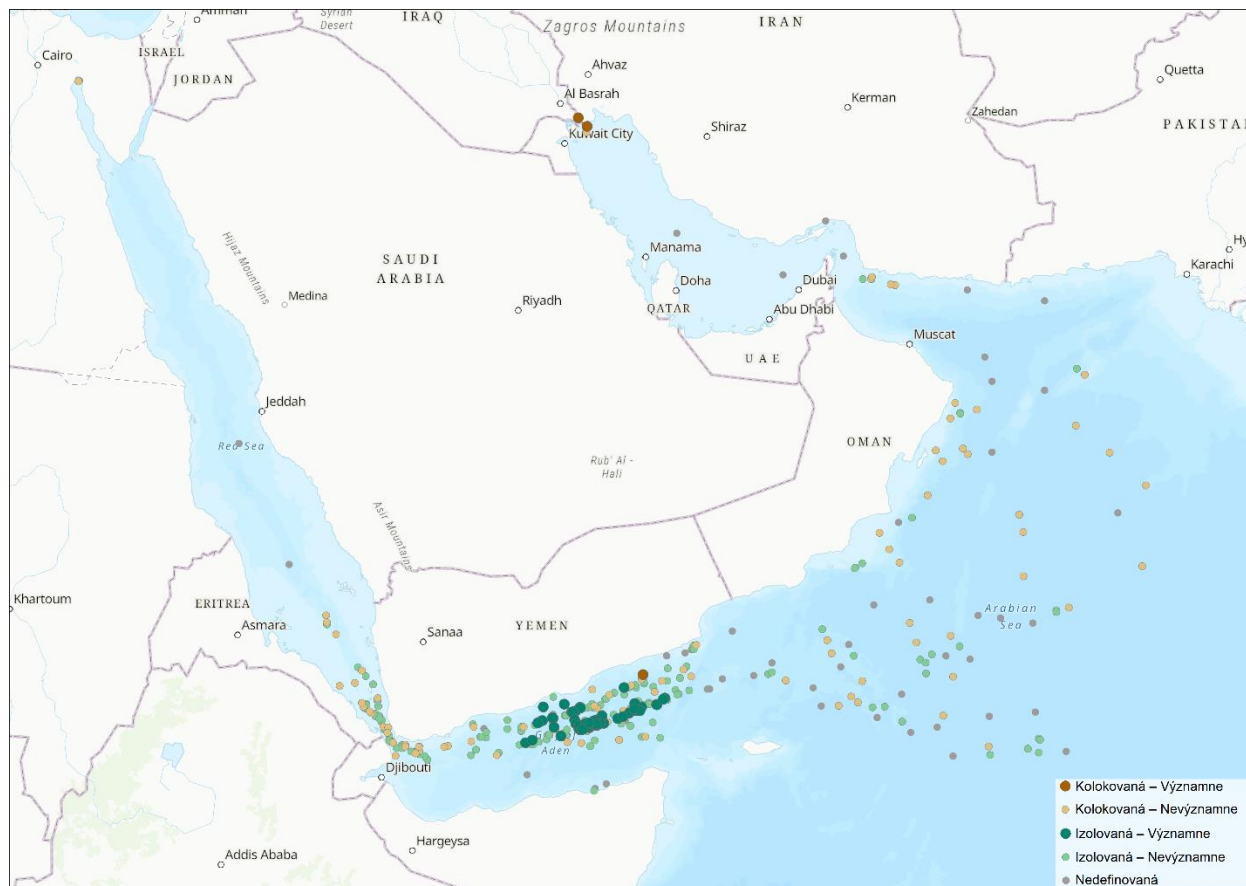
Obr. 17 Kolokačný vzor – metóda časopriestorového okna s medznou vzdialenosťou 50 kilometrov a časovým intervalom päť rokov
CM_SpaceTimeWindow_5R_50km_CA

Poslednou skúmanou metódou pomocou externého súboru je K najbližších susedov, pri parametrizácii nie je možné pracovať s časovým údajom. Každopádne 10 najbližších prvkov je pomerne nízka hodnota, vzhľadom na veľkosť vstupných dát. Všetko ale záleží na užívateľovi a na veľkosti študovaného okolia. Pri porovnaní s prechádzajúcimi metódami sa stále množstvo kolokovaných bodov pohybuje okolo tretiny, konkrétne 33 % záznamov je nevýznamne kolokovaných a 4,5 % je kolokovaných významne. Tie sa tradične nachádzajú v okolí Perzského zálivu, prielivu Báb el-Mandeb, ale aj v južnej oblasti Arabského mora, kde predtým boli skúmané nákladné lode najmä izolované (viď Obr. 18).



Obr. 18 Kolokačný vzor – metóda K najbližších susedov s desiatimi susednými prvkami
CM_KNearestNeighbors_10S_CA

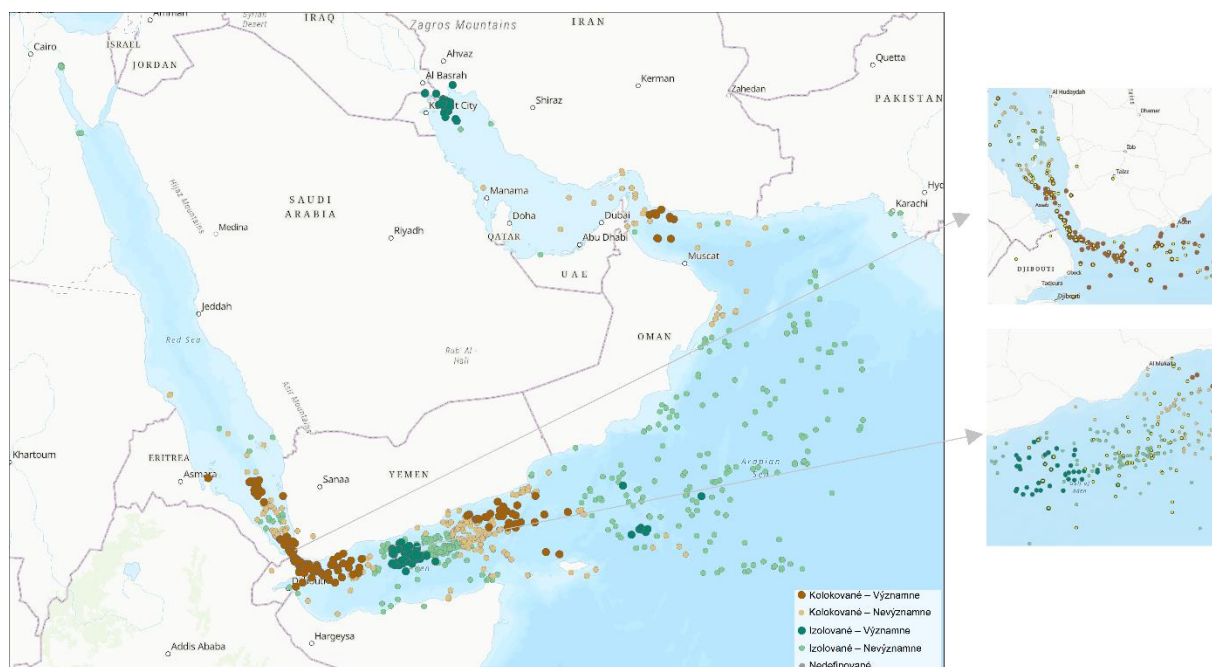
Vzdialenostné pásmo je už metóda konceptualizácie priestorových vzťahov v rámci nástroja kolokačnej analýzy. Okrem medznej vzdialenosti, je však možné nastaviť aj časový interval. Výsledný kolokačný vzor sa nápadne podobá časopriestorovému oknu. Nie je to prekvapujúce, keďže rovnako používa pri výpočte priestorovú aj časovú zložku. Významne kolokovaných bodov je pomerne málo, kolokovaných je necelých 26 %. Objavujú sa aj nedefinované záznamy, typické pre metódy obsahujúce časový interval (viď Obr. 19). Taký záznam v podstate nesplnil definovanú podmienku a jeho lokálny kolokačný kvocient je rovný nule.



Obr. 19 Kolokačný vzor – metóda vzdialenostného pásma so vzdialenosťou 50 kilometrov a časovým intervalom päť rokov
CM_DistanceBand_5R_50km_CA

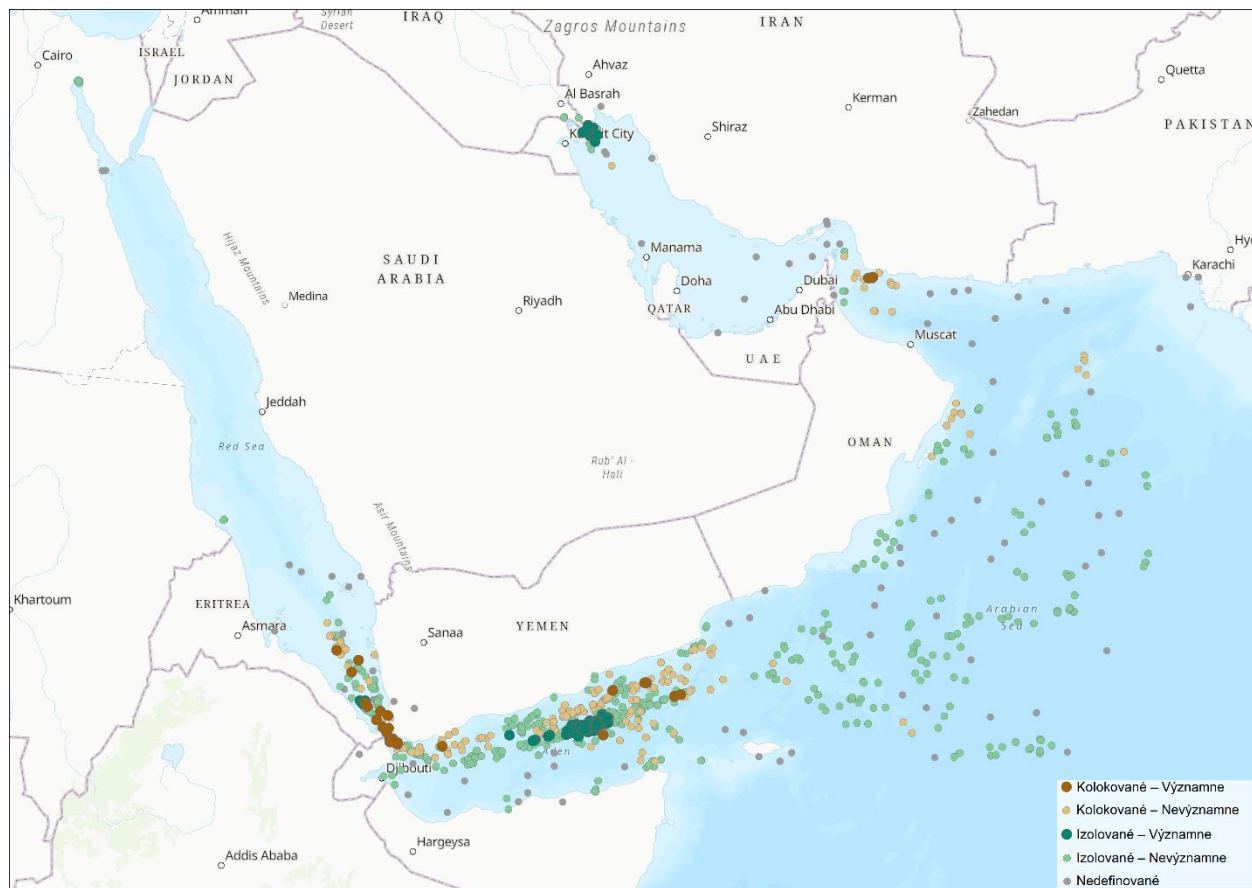
6.2 Pirátske násilie a podozrivé priblíženie

Cieľom týchto skúmaných kategórií bolo zistiť či sú záznamy s podozrivým priblížením kolokované s pirátskym násilím a či je medzi nimi priestorový či časový vzťah. Výsledný vzor je veľmi zaujímavý, očividne existujú lokality, kde sú nepriateľské aktivity v priestorovom vzťahu. Takými miestami je južná časť Červeného mora, samotný prieliv Báb el-Mandeb, Ománsky záliv a východná časť Adenského zálivu (viď Obr. 20). Celkovo je kolokovaných necelých 47 %, čo je takmer polovica. Takmer všetky pirátske útoky v Arabskom mori sú izolované, zdá sa, že v týchto oblastiach nie je medzi sledovanými kategóriami kladný vzťah. Faktom však ostáva, že pri analýze bola použitá pomerne veľká medzná vzdialenosť, ktorá mohla byť dôvodom takýchto pozitívnych výsledkov. V ostatných nastaveniach sú kolokované záznamy, no ich podiel sa pohybuje okolo 35 %.



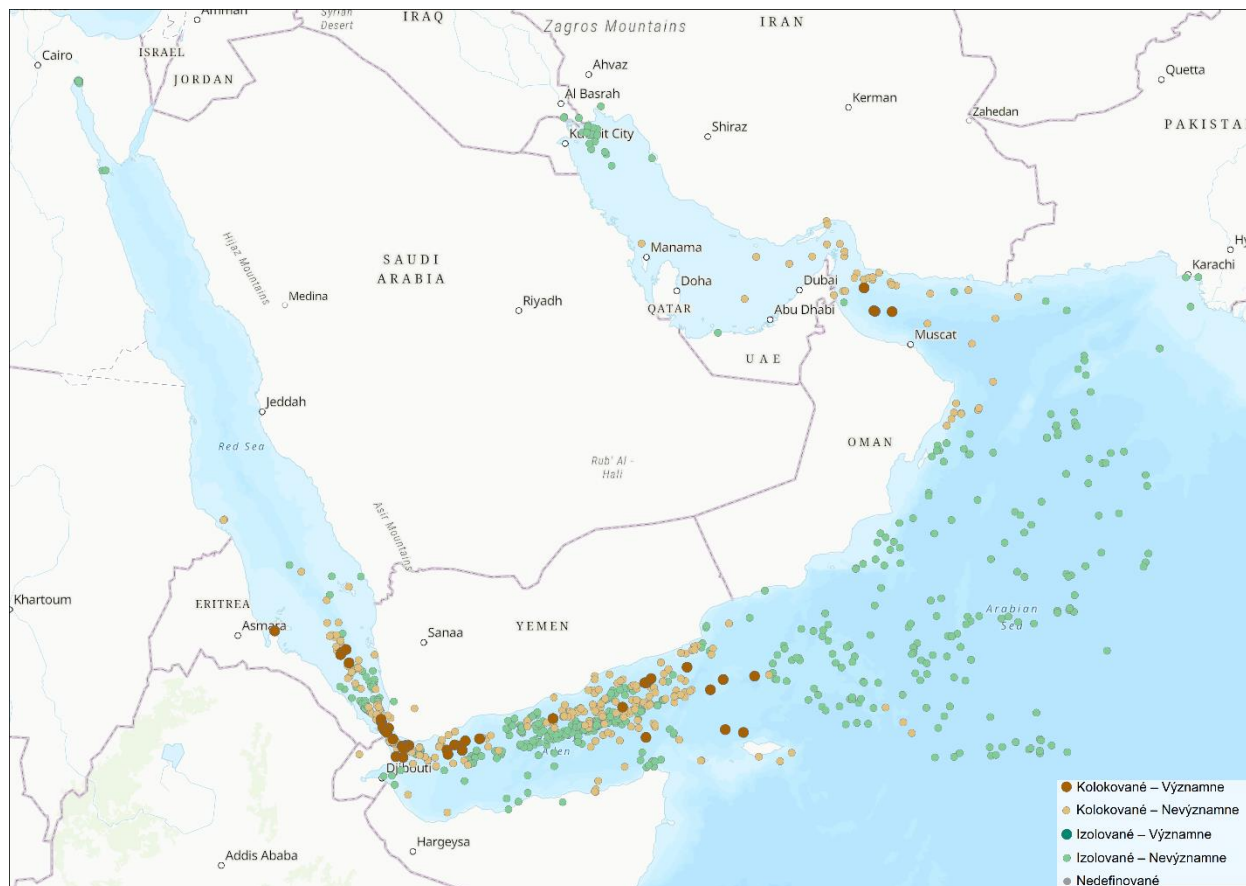
Obr. 20 Kolokačný vzor – metóda inverznej vzdialenosti s desiatimi susednými prvkami a medznou vzdialenosťou 100 kilometrov
PS_Inverse_Distance_10S_100km_CA

Pri použití časového intervalu spolu s priestorovou podmienkou dochádza zvyčajne ku zvýšenému výskytu nedefinovaných záznamov. Každopádne pri nastavení piatich rokov a medznej vzdialenosti 50 kilometrov je len 11,2 % pirátskych útokov označených za nedefinovaných, no na druhej strane je 55,2 % útokov nevýznamne izolovaných. Zdá sa, že v oblasti Arabského mora nemajú útoky ani typy plavidiel časopriestorový vzťah. Na druhej strane opäť sa objavujú rovnaké lokality kolokovaných záznamov, typicky prieliv Báb el-Mandeb a Ománsky záliv (viď Obr. 21). Novo kolokovanou lokalitou je Adenský záliv, avšak aj tu sa objavujú izolované body či už významne, alebo nevýznamne. Z celkového pohľadu je len 23 % pirátskych útokov označených za kolokovaných, no veľkú rolu hrá časový aspekt.



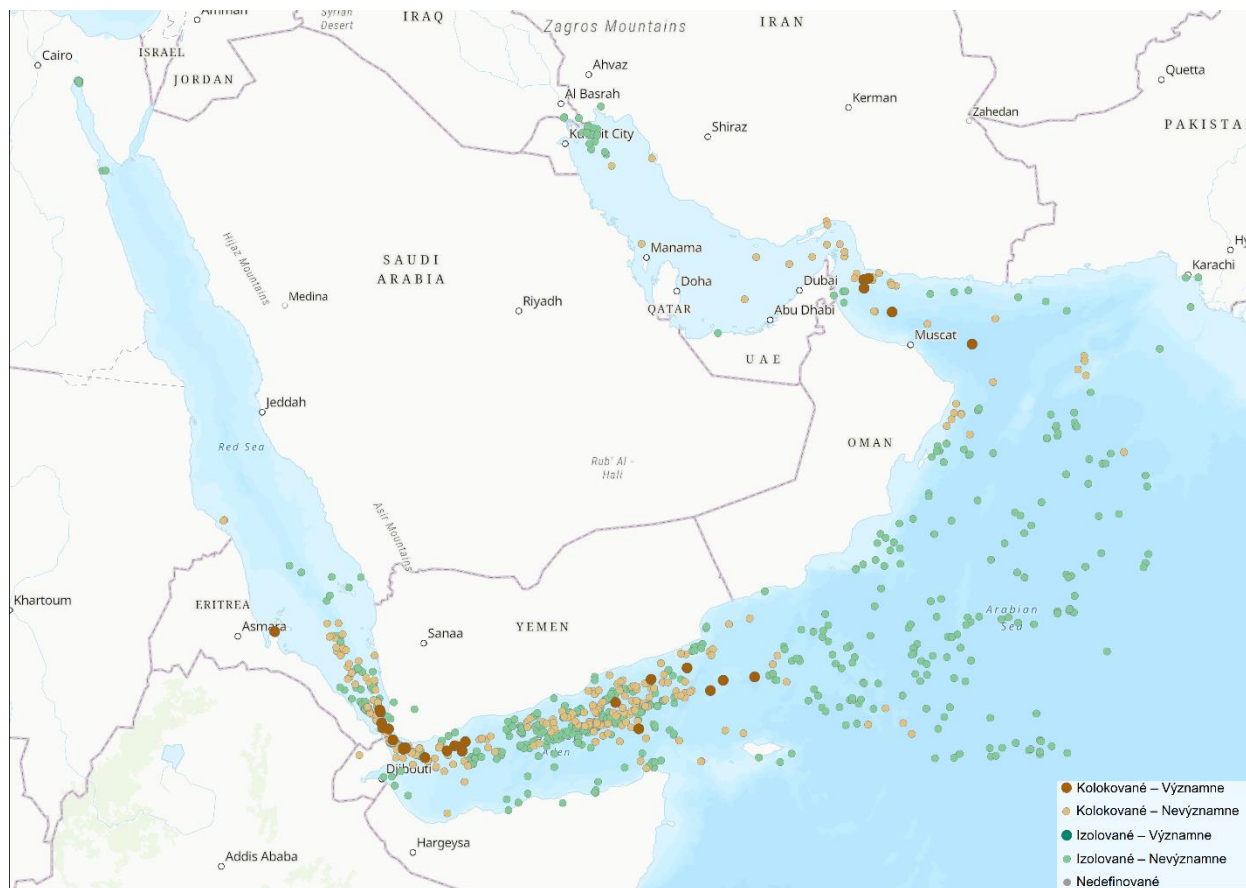
Obr. 21 Kolokačný vzor – metóda časopriestorového okna s medznou vzdialenosťou 50 kilometrov a časovým intervalom päť rokov
PS_SpaceTimeWindow_5R_50km_CA

Pri použití metódy K Najbližších susedov je výsledný kolokačný vzor podobný tomu predchádzajúcemu, kolokované útoky sú najviditeľnejšie v prielive a Ománskom zálive, celkovo tvoria 37,4 %. Taktiež sa objavuje pomerne veľká časť kolokovaných bodov v Adenskom zálive (viď Obr. 22). Očividne sú to miesta, kde majú skúmané aktivity priestorový vzťah. Je viac než pravdepodobné, že záznamy s podozrivým priblížením mohli byť prvotne iniciované ako pirátsky útok, no z istých dôvodov sa útočníci stiahli, alebo boli chytení. Oblasť Arabského mora takmer neobsahuje žiadne iné útoky, než tie izolované. Je to oblasť, ktorá je už vzdialenejšia od pevniny, v porovnaní so zálivmi, ktoré sú medzi pevninami. Práve to môže byť dôvodom častejšej izolovanosti, útoky môžu pochádzať z druhej strany pevniny, častokrát z Afriky.



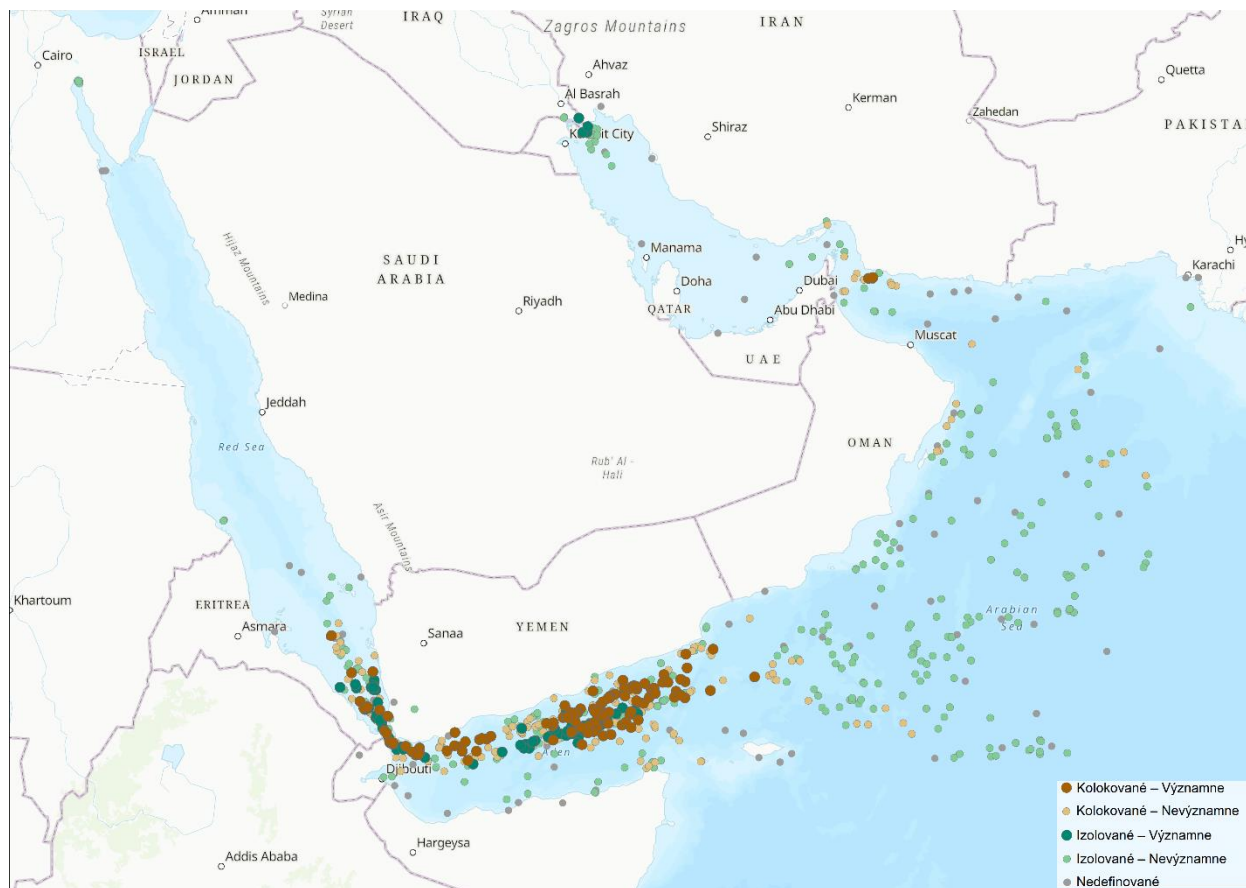
Obr. 22 Kolokačný vzor – metóda K najbližších susedov s desiatimi susednými prvками PS_KNearestNeighbors_10S_CA

Opäť použitie metódy K najbližších susedov, avšak teraz nie s pomocou externého súboru. V kapitole o metóde K najbližších susedov je zmienka o fakte, že výsledky nie sú totožné pri použití externého súboru s touto metódou a pri použití klasického nástroja kolokačnej analýzy. Už pri pohľade na priestorovú distribúciu je možné pozorovať rozdiely. V oblasti prieplyvu pár významne kolokovaných útokov zmizlo, rovnako aj v Adenskom zálive. Na druhú stranu mierny nárast týchto útokov sa objavil v Ománskom zálive (viď Obr. 23). Oblasť Arabského mora ostala takmer nezmenená, a taktiež sa vo vzore neobjavujú žiadne nedefinované záznamy.



Obr. 23 Kolokačný vzor – metóda K najbližších susedov s desiatimi susednými prvkami
PS_KNearestNeighbors_CA_10S_CA

Poslednou metódou v rámci tohto manuálu je vzdialenostné pásmo, kde je možné uplatniť aj atribútovú zložku s dátumom útoku. Z celkového počtu pirátskych útokov je kolokovaných 27,4 %, čo nie je viac než vo väčšine predchádzajúcich kolokačných vzorov. Rovnako ostávajú oblasti kolokovaných záznamov, no istá zmena tu nastala. Pomerne veľká časť významne kolokovaných záznamov je v Adenskom zálive, kde sa doteraz objavovali len v malom počte. Arabské more ostáva stále v majorite útokov nevýznamne kolokované (viď Obr. 24). Vzhľadom na časový aspekt v analýze je možné opäť pozorovať nedefinované záznamy, no je ich menej než 10 %.



Obr. 24 Kolokačný vzor – metóda vzdialenostného pásma so vzdialenosťou 100 kilometrov a časovým intervalom päť rokov
PS_Distance_band_5R_100km_CA