

Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

**AGREGAČNÍ A DISAGREGAČNÍ METODY
V ARCGIS PRO**

Diplomová práce

Bc. Jan ZAPLETAL

Vedoucí práce doc. RNDr. Jaroslav BURIAN, Ph.D.

Olomouc 2021
Geoinformatika

ANOTACE

Práce se zaměřuje na umožnění automatického zpracování agregačních a disagregačních metod v prostředí ArcGIS Pro. Hlavním cílem práce bylo implementovat obě analytické metody do vytvářeného nástroje. Dalším cílem práce bylo vytvořený nástroj otestovat na několika datových sadách různého prostorového rozsahu a vybrané přepočty vizualizovat a interpretovat. Vytvořený nástroj byl sestaven tak, aby umožňoval práci s územně správnými jednotkami i gridovými sítěmi. Nedílnou součástí práce bylo k sestavenému nástroji vytvořit manuál pro jeho použití.

Na základě studia problému agregace a disagregace byl navržen vhodný postup přepočtu tak, aby jej bylo možné aplikovat na přepočet dat do libovolných plošných územních jednotek. Tento postup byl implementován do vytvořených nástrojů.

Výsledkem práce je toolbox do ArcGIS Pro obsahující šest skriptů v jazyce Python. Vytvořené skripty byly zpracovány tak, aby byly schopny provádět přepočty pro extenzivní i intenzivní vstupní proměnnou. Součástí výsledků jsou také vizualizace přepočtů testovaných ukazatelů. Doplňujícím výstupem práce je validace areálových interpolací počtu obyvatel do administrativních jednotek a gridových sítí. Vybrané výsledky přepočtů byly vizualizovány podle kartografických pravidel a přiloženy v tištěné podobě. Vytvořený toolbox včetně manuálu je k dispozici na webových stránkách práce. Výsledky jsou prezentovány na webových stránkách diplomové práce a na informačním posteru, veškeré výstupy jsou dostupné na přiloženém DVD.

KLÍČOVÁ SLOVA

agregace; disagregace; skript; pomocná data; vizualizace

Počet stran práce: 59

Počet příloh: 10 (z toho 6 vázaných a 4 volné)

ANOTATION

This thesis is focused on enabling automatic processing of aggregation and disaggregation methods in ArcGIS Pro environment. Main aim of the thesis was to implement both analytical methods into a created processing tool. Next aim was to test the created tool on several datasets of various spatial resolutions and to visualize and interpret some results of aggregation and disaggregation calculations. The tool was designed to be able to work with territorial administrative units and grid cells. Last but not least, a handbook for the tool was created.

A suitable method of data recalculation was proposed based on the research of aggregation and disaggregation methods. This method is applicable to the conversion of indicators into any territorial units and it was implemented into the developed set of tools.

Main result of the thesis is a toolbox for ArcGIS Pro containing six Python scripts. The scripts were designed to allow calculations for both extensive and intensive input variable. The results also include visualisations of some of the indicators tested in calculations. As an additional output of the thesis, several areal interpolation techniques for converting population to administrative units or grid cells were validated on reference data. Some of the calculated indicators were visualised according to the cartographic rules and their hard copies were appended to this thesis. The toolbox is available on the thesis website, handbook inclusive. The results are presented on the thesis website and the poster, the outputs are available on the attached DVD.

KEYWORDS

aggregation; disaggregation; ancillary data; script; visualisation

Number of pages: 59

Number of appendixes: 10

Prohlašuji, že

- diplomovou práci včetně příloh, jsem vypracoval(a) samostatně a uvedl(a) jsem všechny použité podklady a literaturu,

- jsem si vědom(a), že na moji diplomovou práci se plně vztahuje zákon č.121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo,

- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užívat (§ 35 odst. 3),

- souhlasím, aby jeden výtisk diplomové práce byl uložen v Knihovně UP k prezenčnímu nahlédnutí,

- souhlasím, že údaje o mé diplomové práci budou zveřejněny ve Studijním informačním systému UP,

- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé diplomové práce v rozsahu § 12 odst. 4 autorského zákona,

- použít výsledky a výstupy mé diplomové práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne

Bc. Jan Zapletal

podpis autora

Děkuji vedoucímu práce doc. RNDr. Jaroslavu BURIANOVÍ, Ph.D. za cenné podněty, rady a připomínky během vypracovávání práce. Dále děkuji mé rodině a přátelům za podporu po celou dobu mého studia.

Za poskytnutá data děkuji Katedře geoinformatiky.

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2019/2020

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Bc. Jan ZAPLETAL**
Osobní číslo: **R190018**
Studijní program: **N1301 Geografie**
Studijní obor: **Geoinformatika**
Téma práce: **Agregační a disagregační metody v ArcGIS Pro**
Zadávající katedra: **Katedra geoinformatiky**

Zásady pro vypracování

Cílem diplomové práce je vytvořit v prostředí ArcGIS Pro nástroj (model, skript) pro automatizaci vybraných agregačních a disagregačních metod. Dalším cílem je tento nástroj prakticky otestovat na několika datových sadách různého rozsahu (stát-kraj-ORP-město). V rešeršní části práce se student zaměří na přehled existujících metod a SW řešení (extenze, model, skript). Dále vybere datové sady vhodné pro agregaci a disagregaci, navrhne vhodné metody zpracování a ty prakticky implementuje a otestuje. Sestavený nástroj bude pracovat jak s územně správními jednotkami, tak také s gridovými sítěmi. Výstupem práce bude nástroj pro agregaci a disagregaci určený pro prostředí ArcGIS Pro, manuál pro jeho použití, agregovaná a disagregovaná data a vhodná mapová vizualizace doplněná interpretací výsledků.

Celou práci, tj. text včetně všech příloh, posteru, výstupů, zdrojových i vytvořených dat, map, programových kódů a databází, student odevzdá v digitální podobě na paměťovém nosiči připevněném k deskám práce s popisem (jméno, název práce, Katedra geoinformatiky UP, rok). Text práce s přílohami odevzdá ve dvou svázaných výtiscích na sekretariát katedry ve stanoveném termínu. O práci student vytvoří webovou stránku v souladu s pravidly dostupnými na stránkách katedry. Práce bude zpracována podle obecných zásad (Voženílek, 2002) a závazné šablony pro kvalifikační práce na KGI. Povinnou přílohou práce je poster formátu A2.

Rozsah pracovní zprávy: **max. 50 stran**
Rozsah grafických prací: **dle potřeby**
Forma zpracování diplomové práce: **tištěná**

Seznam doporučené literatury:

Šablona KGI pro psaní bakalářských a magisterských prací

VOŽENÍLEK, V.: Diplomové práce z geoinformatiky, Vydavatelství UP, Olomouc, 2002, 60 s.

BATISTA E SILVA, Filipe, Javier GALLEGO a Carlo LAVALLE, 2013. A high-resolution population grid map for Europe. *Journal of Maps*. 9(1), 16-28. ISSN 17445647.

BIELECKA, Elzbieta, 2005. A dasymetric population density map of poland. *Proceedings of the 22nd International Cartographic Conference*. (48 22), 9-15.

EICHER, Cory L. a Cynthia A. BREWER, 2001. Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science*. 28(2), 125-138. ISSN 1523-0406.

EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS, 2011. GEOSTAT 1A – Representing Census data in a European population grid.

FLOWERDEW, Robin, Mick GREEN a Evangelos KEHRIS, 1991. Using areal interpolation methods in geographic information systems. *Papers in Regional Science*. 70(3), 303-315. ISSN 10568190.

GALLEGO, Francisco Javier, 2010. A population density grid of the European Union. *Population and Environment*. 31(6), 460-473. ISSN 01990039.

- GOODCHILD, Michael a Nina Siu-ngan LAM, 1980. Areal interpolation: A variant of the traditional spatial problem. *Geo-Processing*. 1, 297-312.
- HOLT, James B. a Hua LU, 2011. Dasymetric Mapping for Population and Sociodemographic Data Redistribution. In: *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*. s. 195-210. ISBN 9780470749586.
- HOLT, James B., C.P. LO a Thomas W. HODLER, 2004. Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science*. 31(2), 103-121. ISSN 1523-0406.
- HUANG, Zhengdong, Henk OTTENS a Ian MASSER, 2007. A doubly weighted approach to urban data disaggregation in GIS: A case study of Wuhan, China. *Transactions in GIS*. 11(2), 197-211. ISSN 13611682.
- KRIVORUCHKO, Konstantin, Alexander GRIBOV a Eric KRAUSE, 2011. Multivariate Areal Interpolation for Continuous and Count Data. *Procedia Environmental Sciences*. 3, 14-19. ISSN 18780296.
- LANGFORD, Mitchel, 2006. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. *Computers, Environment and Urban Systems*. 30(2), 161-180. ISSN 01989715.
- LI, Tiebei a Jonathan CORCORAN, 2011. Testing dasymetric techniques to spatially disaggregate the regional population forecasts for South East Queensland. *Journal of Spatial Science*. 56(2), 203-221. ISSN 14498596.
- LI, Tiebei, David PULLAR, Jonathan CORCORAN a Robert STIMSON, 2007. A comparison of spatial disaggregation techniques as applied to population estimation for South East Queensland (SEQ), Australia. *Applied GIS*. 3(9). ISSN 18325505.
- LI, Xiaoma a Weiqi ZHOU, 2018. Dasymetric mapping of urban population in China based on radiance corrected DMSP-OLS nighttime light and land cover data. *Science of The Total Environment*. 643, 1248-1256. ISSN 00489697.
- MENNIS, Jeremy, 2009. Dasymetric Mapping for Estimating Population in Small Areas. *Geography Compass*. 3(2), 727-745. ISSN 17498198.
- NAGLE, Nicholas N., Barbara P. BUTTENFIELD, Stefan LEYK a Seth SPIELMAN, 2014. Dasymetric Modeling and Uncertainty. *Annals of the Association of American Geographers*. 104(1), 80-95. ISSN 00045608.
- ŠIMBERA, J., 2019. Neighborhood features in geospatial machine learning: the case of population disaggregation. *Cartography and Geographic Information Science*. 1-16. ISSN 1523-0406.
- ZANDBERGEN, Paul A., 2011. Dasymetric Mapping Using High Resolution Address Point Datasets. *Transactions in GIS*. 15(SUPPL. 1), 5-27. ISSN 13611682.

Vedoucí diplomové práce: **doc. RNDr. Jaroslav Burian, Ph.D.**
Katedra geoinformatiky

Datum zadání diplomové práce: **7. října 2019**
Termín odevzdání diplomové práce: **6. května 2021**

L.S.



doc. RNDr. Martin Kubala, Ph.D.
děkan

prof. RNDr. Vít Voženílek, CSc.
vedoucí katedry

OBSAH

SEZNAM POUŽITÝCH ZKRATEK	9
ÚVOD	10
1 CÍLE PRÁCE.....	11
2 METODY A POSTUP ZPRACOVÁNÍ.....	12
2.1 Použité metody	12
2.2 Použitý software a data	14
2.3 Postup zpracování.....	14
3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY	16
3.1 Metoda agregace	16
3.2 Metoda disagregace.....	20
4 TVORBA NÁSTROJŮ.....	27
4.1 Aggregate Features To Polygon	27
4.2 Aggregate Polygon To Polygon.....	28
4.3 Create Ancillary Layer	30
4.4 Disaggregate Polygon To Polygon	31
4.4.1 Vektorová verze	32
4.4.2 Rastrová verze.....	34
5 TESTOVÁNÍ NÁSTROJŮ	37
5.1 Aggregate Features To Polygon	38
5.2 Aggregate Polygon To Polygon.....	40
5.3 Create Ancillary Layer	42
5.4 Disaggregate Polygon To Polygon	43
6 SROVNÁNÍ METOD.....	46
6.1 Volba metod a vstupních dat	46
6.1.1 Zdrojová a referenční data	46
6.1.2 Pomocná data	47
6.2 Přepoččet dat	48
6.3 Validace	50
7 VÝSLEDKY	53
7.1 Nástroje pro agregaci a disagregaci dat	53
7.2 Agregovaná a disagregovaná data	54
7.3 Vizualizace	55
8 DISKUZE	56
9 ZÁVĚR	59
POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE	
PŘÍLOHY	

SEZNAM POUŽITÝCH ZKRATEK

Zkratka	Význam
CLC	CORINE Land Cover
ČR	Česká republika
ČSÚ	Český statistický úřad
GIS	Geografický informační systém
MAUP	Modifiable Areal Unit Problem
OSM	OpenStreetMap
PDF	Portable Document Format
PSČ	poštovní směrovací číslo
RTAE	Relative Total Absolute Error
RÚIAN	Registr územní identifikace, adres a nemovitostí
SHP	ESRI Shapefile
SLDB	Sčítání lidu, domů a bytů
SO ORP	správní obvod obce s rozšířenou působností
TIFF	Tagged Image File Format
ÚAP	územní analytické podklady

ÚVOD

Metoda agregace je analytická metoda seskupující hodnoty jevu z nižších územních jednotek do vyšších. Na vstupu do agregace se nachází detailní bodová nebo polygonová data, při agregaci pak dochází k výpočtu statistického souhrnu ze všech dílčích částí územního celku. Velké množství statistických dat je v dnešní době prezentováno v agregované podobě, nejčastěji do administrativních nebo statistických jednotek. Hlavním důvodem agregace publikovaných dat je zejména jejich anonymizace a ochrana osobních údajů, dalším důvodem může být snížená náročnost na úložiště vytvořené vrstvy nebo jednoduchost další práce s územními jednotkami. Je ale nutné si uvědomit, že agregování dat je rovněž formou generalizace, extrémy v datech se v závislosti na velikosti cílových územních jednotek stírají. Dalším nedostatkem agregace dat je problém měnitelné plošné jednotky, kdy průběh hranic cílových jednotek výrazně ovlivňuje výsledek agregace. Lze tedy tvrdit, že následně zjištěné prostorové vztahy jsou platné pouze pro daný způsob agregace dat.

Opačným přístupem je analytická metoda disagregace, jejímž cílem je z agregovaných dat do vyšších územních jednotek získat data za nižší územní jednotky. Z důvodu ztráty informace o prostorové heterogenitě jevu je metoda disagregace procesně náročnější. Pro modelování variability jevu v území jsou do výpočtu zahrnuta doplňková pomocná data. Největším problémem metody disagregace je výběr vhodných a dostupných pomocných dat v dostatečné úrovni detailu. Za účelem získání co největší přesnosti přepočtu je důležité, aby vrstva pomocných dat co nejvíce souvisela s modelovaným jevem. Další slabou stránkou metody disagregace je absence její standardizace, tato metoda je stále ještě součástí výzkumu. Při provádění disagregace je nutné počítat s tím, že výsledné hodnoty vždy budou pouhým odhadem lišícím se od reality.

Kombinací těchto dvou metod pak vzniká metoda areálové interpolace umožňující transformovat hodnotu jevu ze zdrojových do jiných, neskladebných cílových jednotek. Při areálové interpolaci nejprve dochází k disagregaci hodnot do detailních dasymetrických zón, z nich poté dochází ke zpětné agregaci jevu do nových územních jednotek.

1 CÍLE PRÁCE

Cílem diplomové práce je vytvořit pro prostředí ArcGIS Pro nástroj umožňující automatizaci vybraných agregačních a disagregačních metod. Na základě rešerše budou pro metody agregace a disagregace vybrány vhodné metody zpracování, které budou následně implementovány do vytvořeného nástroje. V rámci práce budou rovněž využity poznatky dosažené v bakalářské práci „Testování disagregačních metod v GIS“ (Zapletal, 2019). Za účelem testování funkčnosti vytvořeného nástroje bude proveden přepočít vybraných statistických ukazatelů v rámci zvolených testovacích území – Olomouckého kraje a celé České republiky. Jako testované ukazatele bude vybrán počet obyvatel, podíl nezaměstnaných osob, index stáří, průměrný věk a index kriminality. V rámci testovacích území bude s využitím vytvořeného nástroje proveden přepočít vybraných témat do územně správních jednotek i gridových sítí. Výsledky přepočtů budou v souladu s kartografickými pravidly vizualizovány a interpretovány, přepočty vybraných témat budou rovněž zhodnoceny na referenčních datech. K vytvořenému nástroji bude na závěr sestaven manuál, umožňující jeho praktické použití.

Výsledný nástroj bude určen pro univerzální přepočít statistických ukazatelů do administrativních jednotek nebo gridových sítí v prostředí ArcGIS Pro. Pro jeho možné nasazení v praxi bude rovněž vytvořen manuál.

2 METODY A POSTUP ZPRACOVÁNÍ

V rámci řešení diplomové práce byly v prostředí ArcGIS Pro testovány dvě analytické metody na několika statistických ukazatelích. Metoda agregace byla testována na bodových a polygonových zdrojových datech, jež byla následně seskupována do územích jednotek vyššího řádu, případně do hexagonální sítě. Metodou disagregace pak byla přepočítávána polygonová data z územních jednotek vyššího řádu do územních jednotek nižšího řádu a do hexagonální sítě. Výsledné přepočtené statistické ukazatele pak byly vizualizovány prostřednictvím konvenčních vizualizačních metod tematické kartografie.

2.1 Použité metody

Metoda agregace

Metoda agregace je analytická metoda určená pro převod dat do vyšší úrovně územních jednotek. Agregace je založená na tvorbě statistického souhrnu z bodových nebo polygonových zdrojových dat. Zdrojová data jsou zpravidla přesně lokalizována, případně již agregována do detailnějších územních jednotek. V agregačním procesu dochází k výpočtu vybraných statistických charakteristik (např. suma, průměr) sledovaných údajů v rámci zvoleného cílového území, jedná se tedy o přesnou metodu nevyužívající žádné odhady ani pomocná data. Nejčastější využití nachází metoda agregace při generalizaci, anonymizaci nebo vizualizaci statistických údajů, případně také při převodu získaných údajů z bodů na polygony.

Metoda agregace lze použít pro přepočet statistických dat do vyšších územních jednotek, případně do gridových sítí (Klauda, 2016). Agregace je také zpravidla posledním krokem metody disagregace, kdy nejprve dochází k disagregaci vybraného údaje do co nejdětalnějších segmentů zájmového území podle dostupnosti pomocných dat. Z těchto segmentů je pak údaj zpětně agregován do cílových územních jednotek.

Metoda disagregace

Metoda disagregace je analytická metoda určená pro přepočet dat z vyšších územních celků do nižších, případně jiných územních celků. Prostorová disagregace umožňuje odhadnout neznámé hodnoty statistického ukazatele pro dílčí část územní jednotky na základě známé hodnoty z vyššího územního celku (Starý, 2012). Samotný přepočet je velmi složitý, a to zejména kvůli nerovnoměrnému rozložení výskytu daného jevu ve sledovaném území a také z důvodu neshodných hranic zdrojových a cílových oblastí (Li a kol., 2007). Protože hodnota jevu ve sledovaném území není konstantní, je vhodné použít pro disagregaci pomocná data znázorňující rozložení jevu v daném území (Zapletal, 2019). Výsledkem disagregace je pak polygonová nebo rastrová vrstva s nově odvozenými hodnotami. Vzhledem k velmi hrubé prostorové lokalizaci vstupních dat a neznámé distribuci jevu v území nebude výsledek disagregace nikdy shodný s realitou, podle Rosiny a Hurbánka (2012) je však možné dospět k přijatelným výsledkům.

Areálová metoda (metoda plošných znaků)

Podle Voženilka a Kaňoka (2011) je metoda plošných znaků jednou z nejčastěji využívaných metod tematické kartografie. Využívá se ke znázornění plošných prostorových jevů buď jako samostatný vyjadřovací prostředek v areálové metodě nebo jako součást složitějších metod (kartogram, metoda teček, kartodiagram, dasymetrická metoda, aj.). Areálová metoda bývá využita pro vyjádření kvalitativních i kvantitativních jevů.

Podle autorů využívá metoda plošných znaků pro vyjádření jevů pouze dva parametry plošných znaků – výplň a obrys. Výplň areálu slouží k vyjádření kvantitativní nebo kvalitativní vlastnosti znázorňovaného jevu, je tedy hlavním vyjadřovacím prostředkem areálové metody. Podle autorů rovněž platí, že podobné jevy se mají v této metodě znázorňovat podobnými barvami a odlišné jevy odlišnými barvami. Dále platí, že čím vyšší intenzita jevu, tím intenzivnější barva/rastr. V případě nezachování vztahu mezi intenzitou jevu a intenzitou výplně pak dochází k tzv. propadání barev, což může vést k mylné interpretaci znázorňovaného jevu. Obrys areálu je díky menšímu vizuálnímu projevu méně výrazným parametrem pro vizualizaci a používá se pro vyjádření sekundárních vlastností daného jevu (Voženílek a Kaňok, 2011).

Metoda kartogramu

Metoda kartogramu je jedním z dílčích typů areálové metody. Podle Voženíleka a Kaňoka (2011) je podstatou metody kartogramu znázornění kvantitativního jevu vyjádřeného relativními hodnotami. Kvantitativní data musí být přepočtena (normalizována) na jednotku plochy dílčího územního celku, aby byly dílčí územní celky vzájemně srovnatelné. V případě znázorňování jevu pomocí relativních hodnot nepřepočtených na jednotku plochy celé dílčí jednotky pak autoři označují danou metodu za pseudokartogram. Při využití metody pseudokartogramu je velmi důležité, aby dílčí územní celky měly přibližně stejnou velikost, jinak dochází ke špatné interpretaci mapy čtenářem. Z mnoha existujících druhů kartogramu se v praxi nejčastěji používá homogenní jednoduchý kartogram, kdy je celek pokryt odstíny barvy podle předem sestavené objektivní stupnice (Voženílek a Kaňok, 2011).

Metoda gridu

Metoda gridu je dalším z dílčích typů areálové metody, slouží k vizualizaci kvalitativních i kvantitativních jevů. Podle Klaudy (2016) je v této metodě území rozděleno na stejně velké územní celky (buňky). Buňky mají nejčastěji čtvercový, hexagonální nebo trojúhelníkový tvar a každá buňka má svůj jednoznačný identifikátor, pomocí něhož jsou jí poté přiřazeny statistické údaje (Klauda, 2016). Podle Huanga a kol. (2007) je možné statistické údaje do odlišných územních jednotek přiřazovat metodou agregace, disagregace nebo interpolačními metodami. V případě vizualizace dat v pravidelné gridové síti pomocí kartogramu pak hovoříme o metodě síťového kartogramu (Zajícová, 2020).

Mezi hlavní výhody gridových sítí patří podle Klaudy (2016) zejména nezávislost na územně a časově proměnlivé struktuře administrativních jednotek, což umožňuje vzájemné srovnání statistických údajů v různých časových obdobích. Stejná velikost jednotlivých buněk gridu pak zajišťuje jejich vzájemnou srovnatelnost. Mezi další výhody pak lze zařadit možnost kombinace gridu s dalšími datovými sadami pomocí mapové algebry nebo geostatistických metod, v neposlední řadě pak zajištění ochrany osobních údajů, kdy jsou pomocí agregace veškeré údaje v gridu anonymizovány (Klauda, 2016).

Jednou z nevýhod metody gridu je nemožnost přesného kopírování hranic zájmového území gridovou sítí. Aby bylo zájmové území pokryto gridovou sítí beze zbytku, musí dojít k přesahům buněk přes hranice tohoto území. Tyto přesahy pak v případě zanedbání okolních hodnot zkreslují vypočtené hodnoty v hraničních buňkách. Při agregaci nebo disagregaci dat do gridové sítě se uplatňuje také Modifiable Areal Unit Problem (MAUP), zejména pak tzv. agregační problém, kdy podle Openshawa (1983) mají zvolené hranice areálů vliv na výsledek agregace. Gridová síť se sice skládá z pravidelné sítě shodných buněk, počátek a natočení gridové sítě však ovlivní výslednou agregaci.

2.2 Použitý software a data

Příprava geografických dat, nastavení parametrů vytvořených nástrojů, samotné agregace a disagregace statistických dat a následné vizualizace proběhly v prostředí ArcGIS Pro ve verzi 2.5.0. Pro tvorbu samotných skriptů byl využit jazyk Python ve verzi 3.6.9, skripty byly napsány v programu Atom, verze 1.55.0. Pro vygenerování hexagonové mřížky byl využit nástroj Generate Tesselation, který se nachází v toolboxu Data Management jako výchozí součást instalace ArcGIS Pro. Tento nástroj umožňuje tvorbu pravidelného gridu v zadaném extentu, na výběr nabízí pět tvarů buněk. V rámci dosažení cílů práce byl v prostředí ArcGIS Pro vytvořen toolbox, jehož součástí jsou vytvořené skripty pro automatizaci agregačních a disagregačních metod. Pro samotný přepočítání dat do cílových zón pak byly využity tyto vytvořené skripty, jejichž princip fungování je podrobně popsán v kapitole 4. Pro přípravu statistických dat, dále také pro tvorbu tabulek a grafů byl použit Microsoft Excel 2016. Pro stahování vrstvy stavebních objektů z databáze RÚIAN za celou ČR byl použit nástroj QGIS RUIAN Plugin z laboratoře GeoForAll při Fakultě stavební ČVUT v Praze. Nástroj je nově dostupný ve verzi 1.1 a funguje jako plugin do programu QGIS, verze 2 a 3. V rámci řešení práce byl však použit tento nástroj ve verzi 1.0 pro QGIS 2.18, protože bylo zjištěno, že tato verze je stabilnější, což bylo důležité při stahování velkého množství dat. Tvorba webových stránek proběhla v programu Atom, verze 1.55.0, tvorba posteru poté v prostředí Adobe Illustrator CC 2019.

Vrstva administrativních hranic byla získána z veřejně dostupné datové sady ČÚZK (ČÚZK, 2020). Tato datová sada je aktualizovaná jednou týdně a je exportem z databáze RÚIAN ve formátu SHP. Pro agregaci počtu obyvatel byla poskytnuta vedoucím práce vrstva adresních bodů „Statistické budovy (vchody k bytům)“, platná k 1. dubnu 2016. Tato data obsahují informace o počtu obyvatel s trvalým a obvyklým pobytem ze SLDB 2011, dále informace o způsobu a typu využití budovy, o počtu podlaží a počtu bytů. Jako pomocná data pro disagregaci statistických ukazatelů byla využita vrstva stavebníobjekty z databáze RÚIAN, platná ke dni 30. března 2021. Z této vrstvy byly v rámci práce využity informace o půdorysu budovy, typu a způsobu využití budovy, o počtu bytů a počtu podlaží. Data pro zvolené statistické indikátory byla získána z Veřejné databáze ČSÚ za obce ČR a jsou platná ke dni 31. prosince 2019 (ČSÚ, 2021). Jako pomocná data pro přepočítání počtu obyvatel byla také použita vrstva zastavěného území z ÚAP Olomouckého kraje platná k roku 2018, vrstva CORINE Land Cover z roku 2018 (Copernicus, 2021) a vrstva landuse vytvořená z dat OpenStreetMap podle metodiky OSMLanduse (2020). Data o kriminalitě byla stažena z webové mapové aplikace Mapa kriminality, platná ke dni 30. března 2021. Prostorová lokalizace jednotlivých deliktů byla před publikací zkruslena pomocí speciálního algoritmu do středů clusterů vytvořených ze všech deliktů v ČR (Policie ČR, 2021).

2.3 Postup zpracování

Samotné tvorbě nástrojů a jejich testování předcházelo studium řešené problematiky. Rešeršní část práce byla zaměřena především na teorii a praktické využití agregačních a disagregačních metod, rovněž byly nastudovány různé metody přepočtů statistických dat. Na základě studia byly vybrány vhodné metody agregace a disagregace, byla zhodnocena jejich složitost, výsledky a možnost implementace do vytvářeného nástroje. Z bakalářské práce byla využita znalost tématu pomocných dat, jejich dostupnosti a vhodnosti použití. Schéma postupu práce (Obr. 2.3.1) znázorňuje samotný průběh přepočtu vybraných ukazatelů.

Hlavním cílem práce bylo vytvořit sadu nástrojů pro automatizaci agregačních a disagregačních metod v prostředí ArcGIS Pro. Na základě studia literatury byla vybrána vhodná řešení, ta byla následně implementována do vytvořených nástrojů. Vzhledem k povaze a vzájemné odlišnosti řešených úloh nebylo možné obě analytické metody zakomponovat do jediného nástroje, z tohoto důvodu vzniklo nástrojů několik. Pro tvorbu nástrojů byl využit programovací jazyk Python. Ten byl upřednostněn před tvorbou nástrojů v rozhraní ModelBuilder z důvodů možnosti implementace složitější logiky, lepších možností parametrizace a možnosti práce s parametrem ValueTable definovaným v parametrech nástroje. Samotná logika nástrojů i jednotlivé vstupní parametry jsou podrobně popsány v kapitole 4. Pro vytvořený nástroj byl následně sestaven i manuál pro jeho použití.

V rámci testování vytvořených nástrojů bylo v rámci testovacího území České republiky a Olomouckého kraje vybráno pět statistických ukazatelů, na nichž pak byly nástroje vyzkoušeny. Pro přepočítání každého ukazatele byl na základě konzultace s vedoucím práce stanoven vhodný postup přepočtu a byla vybrána vhodná pomocná data pro samotný přepočet. Výběr ukazatelů a postup přepočtů je popsán v kapitole 5. V rámci testovacího území Olomouckého kraje bylo na závěr srovnáno několik různých metod areálové interpolace a různých pomocných dat pro přepočet počtu obyvatel do administrativních jednotek a gridových sítí. Srovnání technik areálové interpolace je popsáno v kapitole 6.

Hlavním výstupem práce je toolbox do prostředí ArcGIS Pro obsahující šest nástrojů umožňujících automatizaci agregačních a disagregačních metod. Pro vytvořené nástroje byl zhotoven i manuál jako součást webové stránky a ve formátu PDF. Toolbox i manuál jsou ke stažení na webových stránkách práce a jsou rovněž dostupné jako součást Přílohy 7 (DVD) této práce. Dalším výstupem práce je šest tištěných mapových výstupů obsahujících vizualizace přepočtených údajů. Dále pak 26 vizualizací přepočtených dat, pět tabulek ve formátu Microsoft Excel obsahujících hodnocení přepočtů a samotné skripty v jazyce Python.

Na závěr byla sepsána textová část práce, byl zhotoven informační poster a byly vytvořeny webové stránky prezentující výsledky práce. Součástí práce jsou také tabulky, obrázky a grafy.



Obr. 2.3.1 Grafické schéma postupu práce

3 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

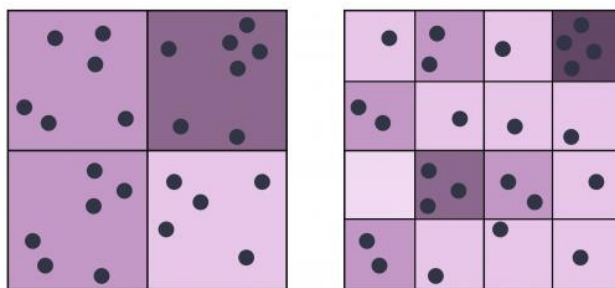
Samotnému řešení práce předcházelo podrobné studium existujících prací zaměřených na téma agregace a disagregace. Byly nastudovány obě metody, jejich teoretický základ a také aplikace v praxi. Na základě rešerše pak byly vybrány konkrétní metody implementované do vytvářených nástrojů.

3.1 Metoda agregace

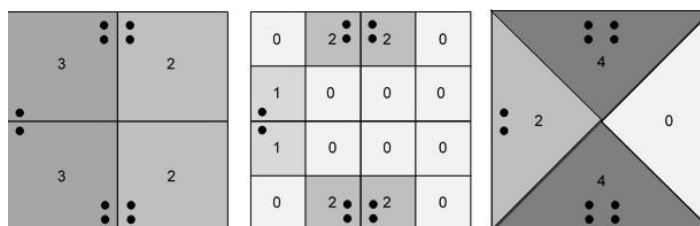
V dnešní době je pro populační analýzy v GIS k dispozici velké množství statistických dat. Data o obyvatelstvu jsou získávána převážně pomocí státem prováděných cenů, různých výběrových dotazníkových šetření nebo na základě sběru dat soukromými či státními institucemi. Výsledkem těchto populačních šetření jsou pak detailně prostorově lokalizované údaje o jednotlivcích. Tato data však není možné zejména kvůli ochraně osobních údajů respondentů dále distribuovat. Zveřejněné výsledky šetření jsou proto nejprve agregovány do územních jednotek, čímž jsou odstraněny údaje o jednotlivcích a data jsou tak anonymizována. I v případě, že již vstupní data jsou již agregována do územních jednotek, je možné tato data agregovat znovu a získat tak souhrnné informace o daném jevu za vyšší územní celky. Agregace také snižuje rozsah řešeného problému a množství požadovaných vstupních dat do analýz. Dále také snižuje výpočetní čas úloh a využití paměti počítače potřebné k řešení dané analýzy (Swink a Speier, 1999).

Podle Wilsona (2019) je agregace tzv. big data do administrativních jednotek jedním z přístupů prostorového výzkumu. Pro účely agregace se pak nejčastěji používají zóny sčítání lidu, zóny dojížděky nebo hranice oblastí PSČ. Využití agregace do větších územních jednotek přináší výhody v podobě lepší dostupnosti ostatních dat vstupujících do analýz a zajišťuje vyšší míru shody v případě využití procesu geokódování. Nevýhodou agregace však je, že získaná data mají pouze informativní charakter a během procesu agregace dochází ke ztrátě polohové přesnosti. Vzhledem ke značnému zjednodušení analyzovaného prostoru je ztížena identifikace a následné zkoumání prostorových procesů. V případě využití velkých územních jednotek pak dochází ke vzniku zbytečné nejistoty v datech, objevuje se zde vliv problému měnitelné plošné jednotky (MAUP) a ekologického klamu. Z těchto důvodů pak lze zpochybňovat správnost prostorových analýz prováděných v příliš hrubém prostorovém rozlišení.

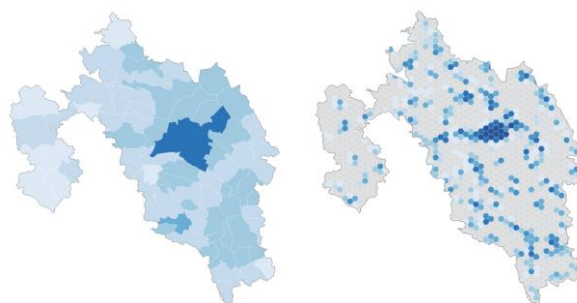
Openshaw (1983) se zabýval problémy vzniklými agregováním dat do prostorových jednotek a v rámci své práce detailně popsal problém měnitelné plošné jednotky (MAUP). Vzhledem k absenci pravidel, standardů a konvencí pro prostorovou agregaci jsou územní jednotky použité v mnoha studiích měnitelné a podléhají představám toho, kdo agregaci prováděl. Existuje mnoho územních jednotek, ale z nich jen málo není měnitelných (např. adresní body, půdorys budov). Data z cenů jsou pak nejčastěji sbírána za neměnitelné plošné jednotky, zveřejňována jsou však za administrativní jednotky, které jsou měnitelné. Žádná z administrativních jednotek nemá geografický význam, přitom jakákoliv následná analýza dat velmi závisí na její definici. MAUP se podle Openshawa (1983) skládá ze dvou částí. První z nich je problém měřítka (viz Obr. 3.1.1), kdy dochází k rozdílným výsledkům při agregaci do větších územních jednotek. Agregací problém (Obr 3.1.2) je pak definován jako jakýkoliv rozdíl ve výsledku analýzy, vzniklý z důvodu agregování do odlišných územních jednotek ve stejném nebo podobném měřítku. V neposlední řadě s MAUP souvisí i problém ekologického klamu (viz Obr. 3.1.3), který vzniká při prostorové agregaci. K ekologickému klamu dochází ve chvíli, kdy se závěry o povaze jevu v části územní jednotky odvodí z povahy jevu agregovaného v rámci celé územní jednotky.



Obr. 3.1.1 Problém měřítka – scale problem (GISGeography, 2020)



Obr. 3.1.2 Agregáčnı́ problém – aggregation problem (Mennis, 2019)

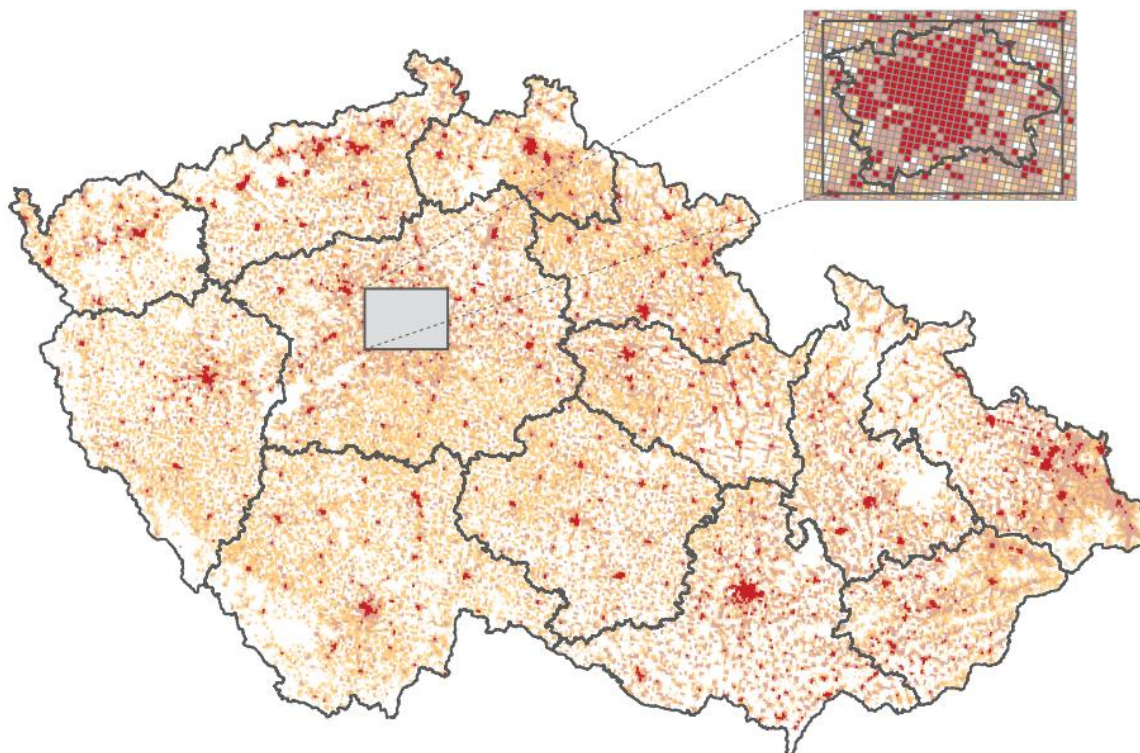


Obr. 3.1.3 Ekologický klam – hustota zalidněnı́ SO ORP Prostějov v roce 2011

Rosina a Hurbánek (2012) se ve své práci zabývali kritikou agregovaných dat a jejich vizualizace pomocí kartogramu. Data o rozmı́stění obyvatelstva jsou zpravidla poskytována pouze za úroveň statistických územních jednotek a vizualizována v podobě kartogramu, což je pro analytické i vizualizační účely nedostatečný způsob reprezentace statistického jevu. Jako hlavní problém uvádı́ autoři MAUP, popsany Openshawem (1983), kdy úprava hranic a míry agregace dat významně ovlivní výsledek prostorové agregace. Vzniklé prostorové struktury jsou pak významně ovlivněny použitými hranicemi areálů a zjištěné vztahy tedy mohou být platné pouze pro zvolený způsob agregace dat. Dalším problémem agregace statistických dat je ekologický klam (Obr. 3.1.3). Znázornění agregovaných statistických dat pomocí kartogramu způsobuje mylnou interpretaci výsledné vizualizace tím, že svádı́ čtenáře k domněnce, že hodnota jevu je v každém místě areálu stejná, což zpravidla není pravda. Vliv rozdílných velikostí územních jednotek navíc znemožňuje jejich vzájemné porovnání a hranice územních jednotek nejsou optimalizovány pro znázornění agregovaných statistických dat, ale vznikaly na základě mnoha jiných vlivů (Rosina a Hurbánek, 2012).

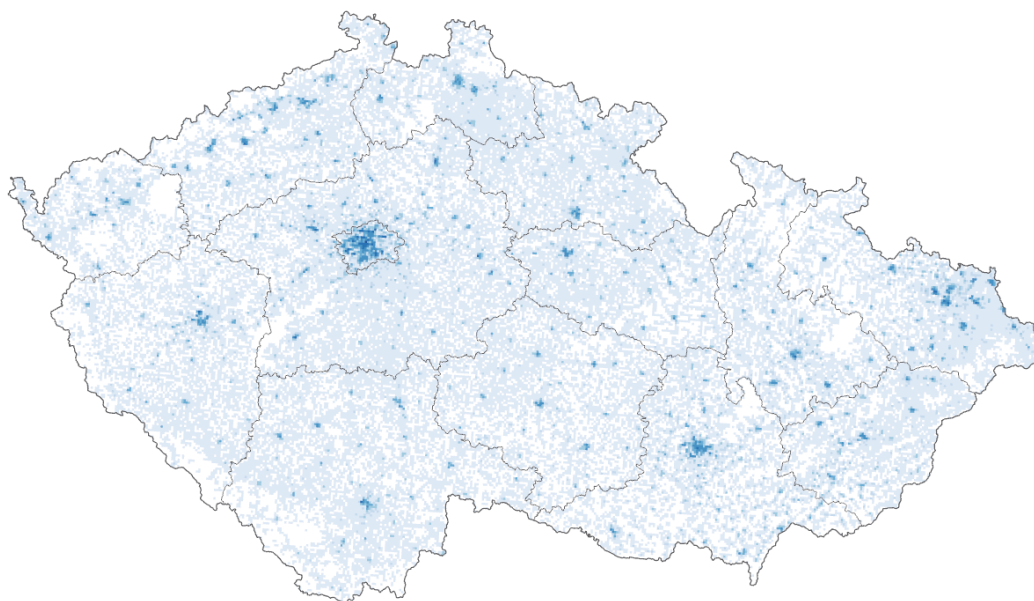
Podle Klaudy (2016) je metoda agregace nejjednodušší a zároveň i nejpřesnější metoda přiřazování statistických informací do buněk gridových sítı́. Během agregace dochází ke spojování detailnějších statistických dat do vyšších územních celků. Výhodou agregace statistických dat do gridových sítı́ je pak jejich nezávislost na administrativním členění, pravidelnost, vzájemná srovnatelnost a anonymizování dat (Klauda, 2016). Autor také vytvořil z dat SLDB 2011 grid počtu obvykle bydlících osob v ČR s rozlišením 1 km (Obr.

3.1.4). Podle Rosiny a Hurbánka (2012) již některé státy přistoupily k poskytování populačních dat agregovaných do pravidelných gridových sítí, tato snaha má však stále daleko od celoevropského rozsahu.



Obr. 3.1.4 Počet obvykle bydlících osob v ČR v gridové síti 1 km² (Klauda, 2016)

ČSÚ se v letech 2013–2014 podílel na celoevropském projektu GEOSTAT 1B, jehož výsledkem je celoevropská populační gridová síť a rozpracované agregační i disagregační metody (Klauda, 2016). Projekt GEOSTAT 1B vycházel z projektu GEOSTAT 1A, v němž byly specifikovány potřeby pro vizualizaci populačních dat v gridových sítích a také metodiky pro agregaci a disagregaci (EFGS, 2014). V rámci projektu GEOSTAT 1B vznikla první verze celoevropské gridové datové sady pro sčítací rok 2011, včetně nástroje pro tvorbu gridové sítě různých velikostí buněk pro ArcGIS 10.0 a detailního manuálu tvorby populačního gridu ze statistických dat použitím agregačních a disagregačních metod (EFGS, 2014). Data byla do buněk gridu přiřazována metodou agregace (bottom-up přístup), metodou disagregace (top-down přístup) nebo jejich kombinací (hybridní přístup). Jako zdrojová data pro agregaci byly využity adresní body, centroidy budov a centroidy parcel, včetně informací o jejich obyvatelstvu ze sčítání lidu v roce 2011 (EFGS, 2014). Disagregace počtu obyvatel pak probíhala zejména v Rakousku a Bulharsku, kdy byla jako pomocná data využita mimo jiné data z Urban Atlas, CORINE Land Cover 2006, OpenStreetMap nebo data o nepropustnosti půdy (EFGS, 2014). V roce 2021 Eurostat (2021) na svých webových stránkách zpřístupnil aktualizaci populačního gridu GEOSTAT s přepočtenými daty za rok 2018. Výsledný grid je k dispozici ke stažení za celou Evropu ve formátech SHP a TIFF (Eurostat, 2021). Data pak obsahují informaci o počtu obyvatel v dané buňce, datu platnosti daného údaje a o metodě přiřazení počtu obyvatel do dané buňky. Převažující metodou tvorby tohoto datasetu však byla disagregace změn počtu obyvatel mezi lety 2011 a 2018 (61,3 % buněk), což je i případ všech buněk v rámci ČR.



Obr. 3.1.5 Hustota zalidnění v gridové síti ČR z dat GEOSTAT 2018

Zajícová (2017) ve své bakalářské práci srovnávala prostorovou prezentaci adresních bodů pomocí administrativních jednotek a gridové sítě. Pro přiřazování informací o adresních bodech do územních jednotek a gridu využila metodu agregace. Za tímto účelem byly otestovány a zhodnoceny různé nástroje pro agregaci bodových dat v prostředí ArcGIS for Desktop a ArcGIS Pro. Rovněž byly testovány různé nástroje pro tvorbu pravidelné gridové sítě v prostředí GIS. Nejprve bylo provedeno vzájemné srovnání čtvercové a hexagonové gridové sítě pro pět různých velikostí buňky. Jako srovnávaná kritéria byla stanovena velikost přesahu buněk přes zájmové území, počet hraničních buněk a počet buněk obsahujících body. Na základě těchto kritérií byl stanoven jako vhodnější pro vizualizace hexagonový tvar gridu. Také byly stanoveny vhodné velikosti buněk pro vizualizace agregovaných dat v různých měřících. Gridová struktura také byla ohodnocena jako vhodnější pro vizualizaci statistických dat, protože vystihuje rozložení tématiky věrohodněji než většina administrativních jednotek (Zajícová, 2017).

Esri v rámci své platformy ArcGIS Insights nabízí v rámci skupiny nástrojů pro prostorovou analýzu také nástroj Spatial Aggregation (Esri, 2021a). Tento nástroj běží ve webovém prostředí a pro spuštění vyžaduje dvě vstupní vrstvy – vrstvu cílových zón obsahujících hranice územních jednotek, do nichž agregace proběhne, a agregovanou vrstvu. Tento nástroj umožňuje agregovat do cílových zón bodová, liniová i polygonová data. Nabízí rovněž parametr Style By, pomocí něhož je možné změnit vypočtenou statistickou charakteristiku zvoleného ukazatele. Při agregaci bodových dat dochází pouze k sumarizaci bodů v rámci cílových zón, žádné výpočty nejsou váženy. Při agregaci liniových a plošných dat se u statistiky Average využívá výpočtu váženého průměru podle vzorce (Obr. 3.1.6), kde N je počet záznamů (prvků), x_i je hodnota daného jevu pro daný prvek a w_i je váha daného prvku. Nástroj umožňuje rovněž agregovat atributy datového typu text. Agregace textu je prováděna spojováním hodnot atributů z jednotlivých zdrojových prvků do jednoho řetězce s oddělovači.

$$\bar{x}_w = \frac{\sum_{i=1}^N w_i \cdot x_i}{\sum_{i=1}^N w_i}$$

Obr. 3.1.6 Vzorec pro výpočet váženého průměru (Esri, 2021a)

QGIS sám o sobě žádný hotový nástroj pro prostorovou agregaci neobsahuje, jeho Expressions engine však obsahuje funkci pro prostorovou agregaci a několik funkcí pro atributovou agregaci (QGIS, 2021). Tyto funkce je možné využít v dialogovém okně Výrazy při výpočtu hodnoty atributu. Pro prostorovou agregaci byla vytvořena funkce aggregate. Tato funkce má šest argumentů, z nichž tři jsou povinné a tři volitelné. Její obecná syntaxe je *aggregate(layer, aggregate, expression, [filter], [concatenator='], [order_by])*, kde parametr layer je agregovaná vrstva, parametr aggregate slouží ke zvolení prostorové statistiky při agregaci, zde je možné použít mimo matematických funkcí i spojování řetězců pomocí oddělovače. Parametr expressions slouží pro definici pole ve zdrojové zóně, z něhož se statistika bude počítat. Parametr filter poté slouží k definici prvků, které budou agregovány. V případě prostorové agregace se zde pro dotazy na geometrii zdrojových a cílových prvků uplatňují prostorové analytické funkce (QGIS, 2021). Tento proces je možné i automatizovat, kdy lze automaticky přepočítat hodnotu atributů v případě aktualizace vrstvy. Tento postup se nastavuje přes Formulář atributů, kdy je vytvořený výraz stanoven jako výchozí hodnota editovatelného pole (SpatialThoughts, 2019).

3.2 Metoda disagregace

Metoda disagregace je analytická metoda sloužící k odhadnutí neznámé hodnoty statistického jevu v dílčích územních celcích ze známých hodnot tohoto jevu ve vyšších územních celcích. Metoda disagregace je podle Huanga a kol. (2007) speciální případ areálové interpolace, kdy cílové zóny jsou menší než zdrojové zóny a nedochází mezi nimi k žádnému průniku. Flowerdew a kol. (1991) ve své práci rozdělili přepočítávané statistické ukazatele do tří kategorií – extenzivní, intenzivní a kategorické. U **extenzivní proměnné** se předpokládá, že poměr mezi hodnotou jevu v části a celku je shodný jako poměr rozloh jejich areálů. Nejčastěji se jedná o jev vyjádřený počtem nebo sumou, např. počet obyvatel. **Intenzivní proměnná** předpokládá, že hodnota jevu v části je shodná s hodnotou jevu v celku. Hodnota proměnné v cílové zóně je poté rovna váženému průměru hodnot ze všech částí zdrojových zón, které cílovou zónu protínají. Zpravidla jde o jev vyjádřený mírou nebo podílem (Flowerdew a kol, 1991).

Metoda disagregace bývá také někdy označována jako areálová interpolace. Flowerdew a kol. (1991) definují areálovou interpolaci jako transformaci dat z jedné sady územních jednotek do jiné. Jedná se tedy o proces transformace dat ze zdrojových do cílových zón, kdy se hranice těchto zón mohou překrývat (Flowerdew a kol., 1991). Tato prostorová redistribuce socioekonomických dat vyžaduje dva kroky – prvním z nich je dasymetrické mapování, druhým je poté zpětná agregace do nových cílových zón. Areálová interpolace bez využití pomocných dat může být založena na plošných (areálové vážení) metodách a na bodových (kriging, kernel-based metody) metodách (Holt a Lu, 2011). Podle Huanga a kol. (2007) má areálová interpolace dvě varianty. První z nich, non-volume preserving (nezachovávající hodnoty), je založena na přiřazení hodnoty centroidům územních jednotek a následné interpolaci těchto bodů. Součet hodnot jevu v dílčích zónách pak nebude odpovídat hodnotě jevu v celku. Tato metoda je vhodná, pokud se předpokládá rovnoměrné rozložení jevu v území. Druhá metoda, volume preserving (zachovávající hodnoty), je založena na metodách polygonového překryvu a zachovává pycnophylaktický charakter (Huang a kol., 2007). Dasymetrická mapa splňuje pycnophylaktický charakter, pokud je celková populace každé zdrojové zóny zachována při přenosu do dasymetrické mapy, tedy pokud suma předpokládané populace dasymetrických (cílových) zón, které dohromady vytváří jednu zdrojovou zónu, se rovná počtu obyvatel ve zdrojové zóně (Mennis, 2009).

Dalším možným označením této metody je dasymetrické mapování. Dasymetrické mapování znamená proces přenosu libovolně agregovaných dat do dasymetrické mapy, která dokáže mnohem přesněji zobrazit rozložení jevu v území (Holt a Lu, 2011). Dasymetrické mapy znázorňují kvantitativní data s využitím nově vytvořených hranic, které rozdělují území na relativně homogenní celky z hlediska sledovaného jevu (Eicher a Brewer, 2001). Hranice zón jsou vytvořeny na základě ostrých změn ve statistickém povrchu, každá zóna je vytvořena tak, aby byla vnitřně co možná nejvíce homogenní (Eicher a Brewer, 2001). Podle Holta a Lu (2011) však dasymetrické mapování neústí ve skutečné dasymetrické mapy, neboť dasymetrické mapování zobrazuje rozložení jevu v území odhadnuté z dat agregovaných do administrativních jednotek, zatímco dasymetrické mapy zobrazují přesné rozložení jevu v území.

Proces dasymetrického mapování obvykle zahrnuje disagregaci populačních dat, agregovaných do administrativních jednotek, pomocí doplňkové sady prostorových (pomocných) dat. Vrstva pomocných dat a zdrojové zóny jsou typicky kombinovány použitím operace polygonových překryvů, což vytvoří sadu odvozených (dasymetrických) zón. Data jsou do dasymetrických zón redistribuována ze zdrojových zón pomocí funkčního vztahu mezi kategoriemi pomocných dat a statistickým povrchem, tento funkční vztah může výzkumník předem specifikovat nebo odhadnout pomocí různých statistických technik (Mennis, 2009). Pomocná data jsou nezbytnou komponentou areálové interpolace i dasymetrického mapování, protože jsou proměnnou, podle které se celý výpočet řídí (Holt a Lu, 2011). Ačkoliv pomocná data pomáhají zpřesnit výpočet disagregovaného jevu, vzhledem ke složitosti modelovaných jevů bude výsledná disagregace vždy pouze odhad, lišící se od reality. Batista e Silva a kol. (2013) pak popsali obecný vzorec dasymetrického mapování s využitím pomocných dat (Obr. 3.2.1), kdy P_i je počet obyvatel v dasymetrické zóně i , P_s je počet obyvatel ve zdrojové zóně s , A_i je rozloha areálu dasymetrické zóny i a W_i je váha pomocných dat v dasymetrické zóně i .

$$P'_i = P_s \times \frac{A_i \times W_i}{\sum_i (A_i \times W_i)}$$

Obr. 3.2.1 Obecný vzorec disagregace s využitím pomocných dat (Batista e Silva a kol., 2013)

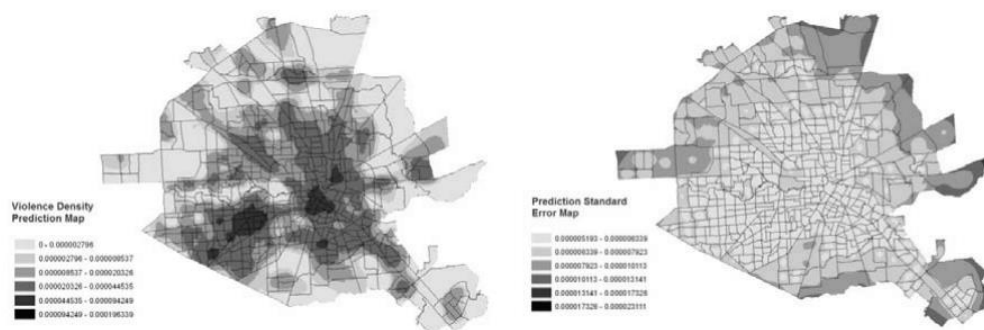
Za dobu studia problematiky disagregace bylo vytvořeno, popsáno a otestováno několik různých metod disagregace. Většina vzniklých metod se liší zejména způsobem práce s vrstvou pomocných dat. Podle Gallega (2010) má ale na přesnost disagregace daleko větší vliv volba vhodných a kvalitních pomocných dat než volba disagregačního algoritmu. Zandbergen (2011) ve své práci zjistil, že přesnější datové sady použité jako pomocná data vykazují přesnější výsledky přepočtů, jejich prostorové pokrytí je však limitováno na velmi malé oblasti výzkumu.

Využití interpolačních metod

Huang a kol. (2007) ve své práci zmiňují využití bodově založených interpolačních metod v areálové interpolaci. Tato metoda je využitelná pro generování prostorové distribuce daného jevu v rastrové podobě, nebere v potaz hranice územních celků. Metoda je vhodná spíše pokud je socioekonomický jev rovnoměrně rozmístěn v prostoru (Huang a kol., 2007).

Protože jsou standardní interpolační metody použitelné pouze nad bodovými daty, Krivoruchko a kol. (2011) se zabývali možností využití interpolačních metod nad polygonovými daty. Nejprve byly stanoveny tři různé typy vstupních dat podle typu jejich statistického rozdělení – Gaussovo, binomické a Poissonovo. Gaussovo rozdělení dat vzniká, když se spojitá bodová měření zprůměrují do polygonů. Výstupem je pak povrch

předpovídající hodnotu jevu v každém jednotlivém místě. U dat s Poissonovým rozdělením je výstupem povrch předpovídající hustotu počtů v dané lokaci, vytvořený model také umožňuje porovnávat jev i pro různé časy pozorování napříč polygony. Výstupem pro binomická data je povrch předpovídající pro každé místo riziko výskytu jedince s určitým rysem. Pro každý typ vstupních dat pak byl vytvořen areálový krigingový model, jeho výstupem je rastrová vrstva interpolovaných hodnot a vrstva chyb odhadu. Jakmile jsou tyto vrstvy vytvořeny, je možné data zpětně agregovat do cílových zón (Krivoruchko a kol., 2011). Vytvořené krigingové modely byly implementovány do vytvořeného nástroje Areal Interpolation. Tento nástroj je nyní součástí balíčku Geostatistical Wizard v ArcGIS for Desktop i ArcGIS Pro.

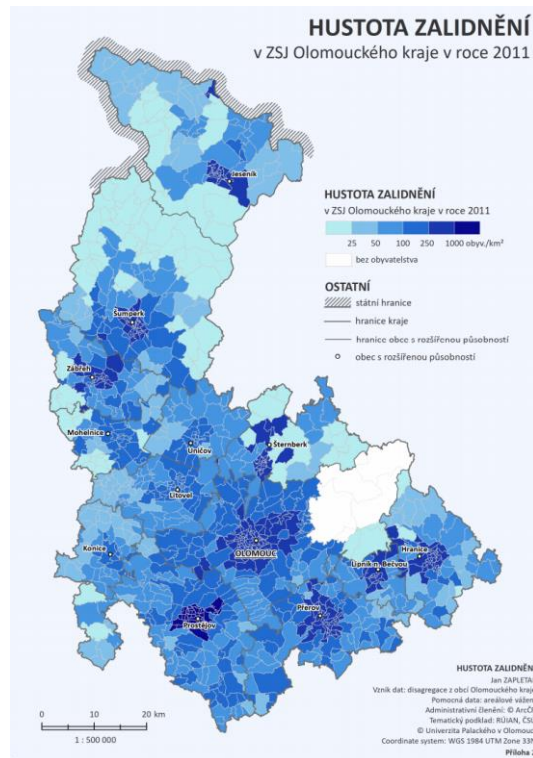


Obr. 3.2.1 Výstup nástroje Areal Interpolation (Krivoruchko a kol., 2011)

Areálové vážení

Areálové vážení je metoda areálové interpolace nevyužívající žádná externí pomocná data. Tato metoda je založena na principu využití rozlohy areálů jako váhy pro přepočítání, předpokládá tedy, že sledovaný jev je v celém území rozložen rovnoměrně (Flowerdew a kol., 1991). Tuto metodu lze využít v případě, že nejsou k dispozici žádné detailnější doplňující informace o území, nebo za předpokladu, že je sledovaný jev v daném území homogenní. Pokud však sledovaný jev není v daném území rozmístěn rovnoměrně, bude při přepočtech docházet k výrazným nepřesnostem (Zapletal, 2019).

Zapletal (2019) v rámci své bakalářské práce porovnával metodu areálového vážení s využitím dalších pomocných dat v rámci disagregace počtu obyvatel z obcí do ZSJ Olomouckého kraje. Na základě vyhodnocení procenta špatně umístěných obyvatel bylo zjištěno, že velké množství ZSJ vykazuje chybu do 5 %, toto je však způsobeno existencí obcí s jedinou ZSJ. Areálové vážení nedokáže dobře odhadnout hodnotu jevu v cílové zóně z důvodu existence malých územních jednotek s vysokou hodnotou jevu a velkých územních jednotek s nízkou hodnotou jevu (Zapletal, 2019). Z důvodu předpokladu interní homogenity jevu bude v případě přepočtení disagregovaného jevu na jednotku plochy výsledná hodnota v rámci celé zdrojové zóny shodná (viz Obr. 3.2.2). Podle Zandbergena (2011) by všechny metody používající pomocná data měly překonat areálové vážení.

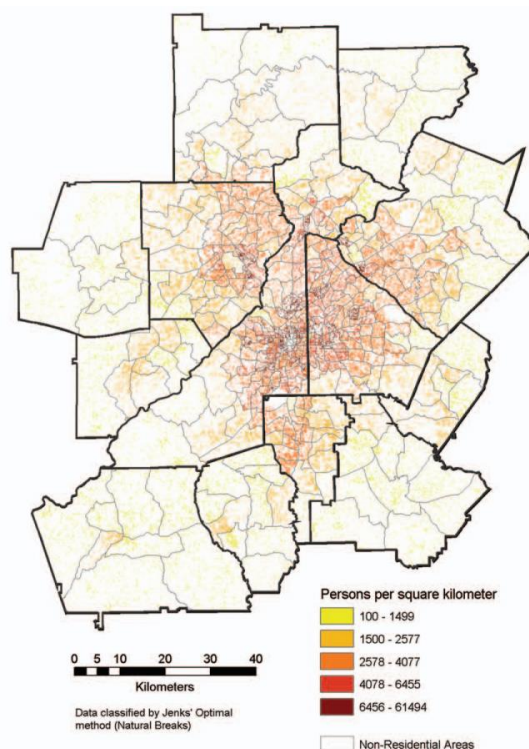


Obr. 3.2.2 Areálové vážení – znázornění interní homogenity jevu (Zapletal, 2019)

Dasymetrické mapování – binární metoda

Binární metoda je nejjednodušší metodou dasymetrického mapování využívající pomocná data pro zpřesnění výpočtu. Principem binární metody je rozdělení sledovaného území na místa s výskytem jevu a místa bez výskytu jevu. Hodnota jevu je pak rozdělena areálovým vážením pouze do částí území s výskytem jevu, v rámci těchto částí je však stále uvažována homogenní hustota. Výhodou binární metody je její jednoduchost, nevýhodou však subjektivita reklasifikace dat do dvou tříd (Bielecka, 2005).

Holt a kol. (2004) testovali použití binární dasymetrické metody pro disagregaci počtu obyvatel v okolí Atlanty za roky 1980, 1990 a 2000. Autoři nejprve z důvodu změn hranic sčítacích okrsků převedli pomocí binární dasymetrické metody data ze sčítacích obvodů z let 1980 a 2000 do sčítacích obvodů z roku 1990. Na základě těchto přepočtů pak byly porovnány změny počtu obyvatel v rámci sčítacích obvodů. Jako pomocná data byly použity družicové snímky Landsat 5, senzoru Thematic Mapper. Tyto snímky byly klasifikovány neřízenou klasifikací metodou ISODATA do dvou kategorií – osídlené a neosídlené oblasti. Závěrem byly srovnány vizualizace hustoty zalidnění ve formě dasymetrické mapy a kartogramu (Obr. 3.2.3). Na základě zhodnocení přesnosti bylo zjištěno, že vlivem nezahrnutí neobydlených oblastí do výpočtu hustoty zalidnění poskytuje dasymetrické mapování výrazně přesnější hodnoty hustoty zalidnění v osídlených oblastech (Holt a kol., 2004).



Obr. 3.2.3 Hustota zalidnění v oblasti Atlanty v roce 1990 (Holt a kol., 2004)

Dasymetrické mapování – klasifikované vážení

Metoda klasifikovaného vážení je jednou z nejčastěji testovaných metod dasymetrického mapování. Metoda je založena na rozdělení studovaného území do kategorií. Pro tento účel jsou nejčastěji využity tři třídy (3-class dasymetric), území je do nich rozděleno podle předpokládané homogenity hustoty zalidnění. Nejjednodušší verze této metody předpokládá, že pro všechny územní jednotky je procentuální rozložení počtu obyvatel v jednotlivých třídách shodné. Největší slabinou tohoto přístupu je, že nebere v potaz rozlohy jednotlivých tříd v daném území. I když daná kategorie zabírá v daném území pouze nepatrnou rozlohu, stále dostane přiřazený stanovený počet obyvatel, což může vést k neobvykle vysokým hustotám zalidnění (Eicher a Brewer, 2001).

Dle Mennise (2003) jsou sčítací obvody v hustě osídlených oblastech relativně malé a mají téměř homogenní distribuci populace, zatímco ve venkovských oblastech jsou sčítací obvody velké a velmi heterogenní. Z tohoto důvodu vytvořená predikce touto metodou bude výrazně přesnější pro hustě osídlené, městské oblasti. Mennis (2003) nahradil subjektivní určování procentuálního rozložení objektivním, kdy na základě známých hodnot počtu obyvatel a rozlohy vypočítal podíl hustoty zalidnění v každé třídě pro pět testovacích území.

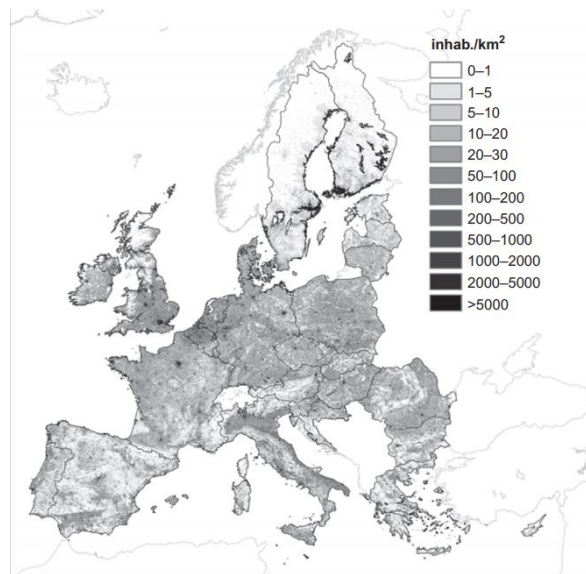
Li a kol. (2007) také využili podíl hustoty zalidnění ve 3-class dasymetrické metodě. Výpočet koeficientů podílu hustoty zalidnění však vypočítali pomocí regresního modelu. Tato metoda pak pro dané testovací území předčila binární i regresní metody (Li a kol., 2007).

Li a Corcoran (2011) předpokládali, že v rámci zdrojových zón bude daleko větší heterogenita jevu, než můžou osvětlit tři použité třídy, snažili se tedy rozdělit sledované území do více tříd pro získání větší variability jevu. Bylo zjištěno, že 3-class metoda poskytuje lepší výsledky než binární a regresní metody, ale nedokáže dostatečně osvětlit heterogenitu jevu. Přidání čtvrté třídy dokáže heterogenitu lépe osvětlit, přidávání dalších tříd už ale výsledky nezlepšuje (Li a Corcoran, 2011).

Dasymetrické mapování s limitující proměnnou

Metoda limitující proměnné je tradiční metodou dasymetrického mapování. Tato metoda je založena na podobném principu jako klasifikované vážení, její princip ve své práci popsali Eicher a Brewer (2001). Území je rozděleno do několika kategorií, nejčastěji podle typu využití území. Každé z těchto kategorií je výzkumníkem nastavena maximální prahová hodnota hustoty zalidnění, tedy limitující proměnná. Neobydleným oblastem je nastavena hodnota nula. Kategorie jsou pak seřazeny podle hustoty zalidnění od nejmenší po největší. Všem osídleným polygonům je areálovým vážením přiřazen počet obyvatel. Pokud v první kategorii přesáhne hustota zalidnění stanovený práh, je jí přiřazena hustota zalidnění podle stanoveného prahu a přebývající počet obyvatel se rozmístí areálovým vážením do zbývajících kategorií využití území. Tato metoda poté funguje v iteracích, dokud nejsou tímto způsobem opraveny všechny kategorie (Eicher a Brewer, 2001).

Eicher a Brewer (2001) zjistili, že metoda limitující proměnné v porovnání s klasifikovaným vážením a binární metodou dosahuje nejnižší hodnoty chyby pro všechny sledované ukazatele. Hodnota maximální hustoty zalidnění v jednotlivých kategoriích využití území se však může v různě osídlených územních jednotkách lišit, proto Gallego a kol. (2011) modifikovali tuto metodu využitím logistické regrese pro výpočet prahových hodnot. Tato modifikovaná varianta však přinesla pouze mírné zpřesnění výsledků disagregace.

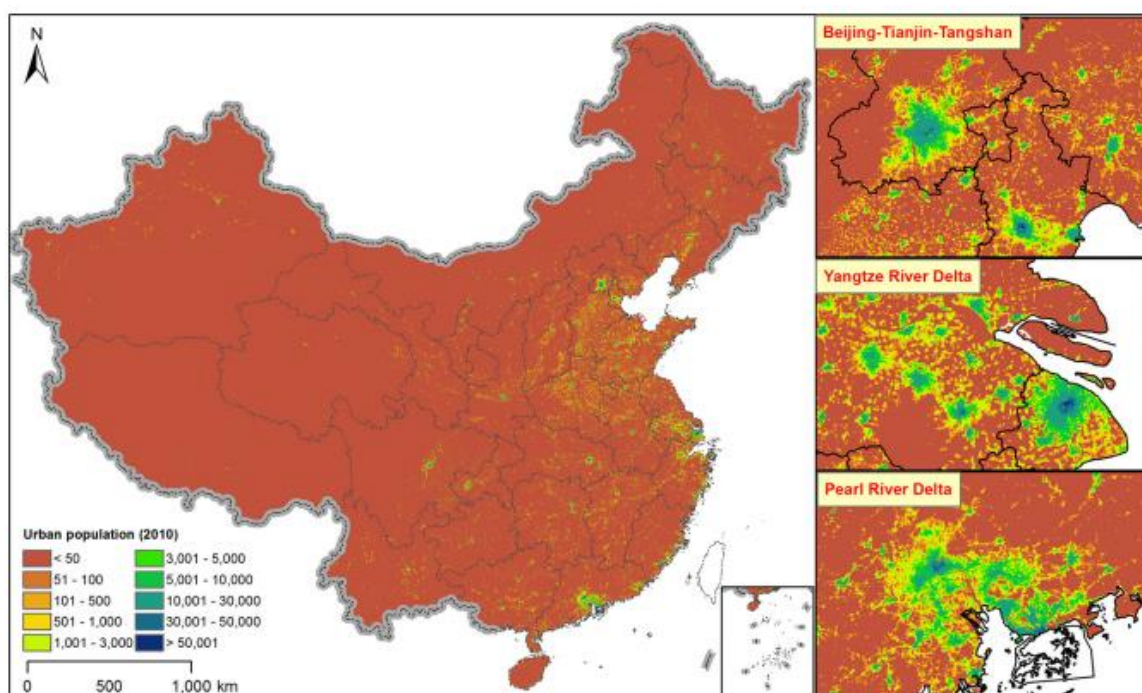


Obr. 3.2.4 Disagregovaná hustota zalidnění metodou limitující proměnné (Gallego a kol., 2011)

Dasymetrické mapování - spojitě vážení

Metoda spojitěho vážení je podle Šimbery (2019) společně s metodou klasifikovaného vážení jeden ze dvou základních typů dasymetrického mapování. Podle Li a Zhou (2018) je jedním z nedostatků klasifikovaného vážení neschopnost vysvětlit rozložení populace v rámci jednotlivých tříd. Díky vzrůstající dostupnosti detailnějších pomocných dat a zvyšujícímu se výpočetnímu výkonu poskytuje spojitě vážení detailnější informace o rozložení jevu v území a tím odstraňuje zmíněnou nevýhodu klasifikovaného vážení. Spojitě vážení může být také považováno za klasifikované vážení s velkým množstvím tříd (Li a Zhou, 2018). Pomocná data jsou převedena do jednoho nebo více prvků v rámci cílové zóny a číselná váha je pak pro tyto prvky vypočítána například pomocí regresního modelu (Šimbera, 2020).

Li a Zhou (2018) ve své práci využili metodu spojitého vážení pro disagregaci počtu obyvatel v Číně. Jako pomocná data byla použita kombinace radiometricky kalibrované intenzity nočního osvětlení (RcNTL) a zastavěného území. Zastavěné oblasti sloužily jako maska rastru nočního osvětlení, čímž došlo k odstranění efektu přesvitu (overglow). Na základě zhodnocení výsledků disagregace počtu obyvatel na referenčních datech bylo zjištěno, že ve studovaném území rastr RcNTL oříznutý zastavěným územím dokázal vysvětlit 80 % rozptylu populace. Z důvodu prostorového rozlišení rastru RcNTL 1 km byl výsledkem práce čtvercový populační grid o straně 1 km (Obr. 3.2.5). Bylo zjištěno, že převzorkování rastru RcNTL podle prostorového rozlišení vrstvy zastavěného území (30 m) mírně zvýší přesnost výpočtu, dramaticky však naroste čas výpočtu pro území velké rozlohy (Li a Zhou, 2018). Šimbera (2020) ve své práci uvádí, že využití intenzity nočního osvětlení jako pomocných dat pro disagregaci počtu obyvatel je vhodnější spíše na nižší úrovni podrobnosti. Největší nevýhodou těchto pomocných dat je pak přesvit z hustě zastavěných oblastí do blízké sousedících neobydlených oblastí (Šimbera, 2020).



Obr. 3.2.5 Populační grid vzniklý dasymetrickým mapováním se spojitou vahou (Li a Zhou, 2018)

4 TVORBA NÁSTROJŮ

Hlavním cílem diplomové práce bylo vytvořit pro prostředí ArcGIS Pro sadu nástrojů pro automatizaci agregačních a disagregačních metod. Všechny nástroje byly vytvořeny v jazyce Python ve verzi 3.6.9 a jsou součástí toolboxu *Dasymetric.tbx*. Vytvořený toolbox je k dispozici ke stažení na webových stránkách práce a jako součást Přílohy 7 (DVD).

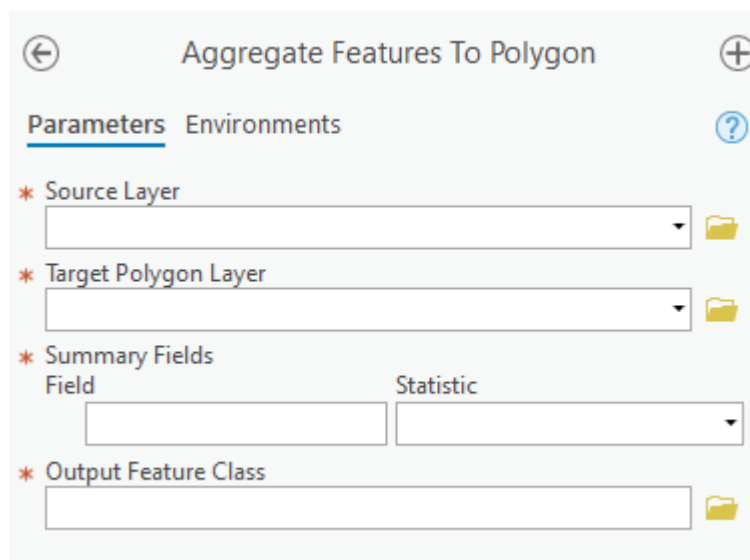
Problematika agregace a disagregace je navzájem velmi rozdílná. Obě analytické metody vyžadují odlišný postup zpracování i jiné vstupní parametry, proto nebylo možné obě metody zahrnout do jednoho nástroje. V rámci práce vzniklo celkem šest nástrojů, z toho dva pro agregaci, dva pro disagregaci a dva pro tvorbu vrstvy pomocných dat. Vytvořené nástroje jsou včetně jejich parametrů a principu fungování detailně popsány v následujících podkapitolách. Závěrem byl k vytvořeným nástrojům zhotoven i manuál pro jejich použití. Manuál je k dispozici jako Příloha 9 a 10 diplomové práce, zároveň je k dispozici ke stažení na webových stránkách práce.

4.1 Aggregate Features To Polygon

Nástroj *Aggregate Features To Polygon* umožňuje agregovat vstupní vektorová data do cílových polygonových zón. Vstupní data do tohoto nástroje mohou být bodová, liniová nebo polygonová, výstupem je pak nová polygonová vrstva s agregovanými atributy.

Vstupní parametry

Nástroj *Aggregate Features To Polygon* vyžaduje čtyři vstupní parametry (viz Obr. 4.1.1), všechny z nich jsou povinné. Parametr *Source Layer* je datovým typem Feature Layer a označuje vstupní bodovou, liniovou nebo polygonovou vrstvu, jejíž atributy budou agregovány. *Target Polygon Layer* je existující polygonová vrstva, do níž budou zdrojová data agregována, datový typ je také Feature Layer. *Summary Fields* je parametr typu ValueTable (tabulka hodnot) a skládá se z částí Field a Statistic. Část *Field* označuje atribut ze zdrojové zóny, pro nějž se provede agregace. Po zvolení zdrojové zóny je zpřístupněna rozbalovací nabídka s atributy ze zdrojové vrstvy, atributy jsou omezeny na datové typy short, long, float a double. *Statistic* je typ statistiky, která bude pro dané pole spočítána. Na výběr je k dispozici pět možností – suma, průměr, minimální hodnota, maximální hodnota a směrodatná odchylka. V parametru *Output Feature Class* je zvolen název a místo uložení nové polygonové vrstvy s agregovanými údaji.

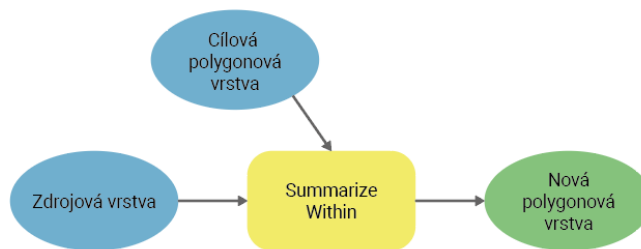


Obr. 4.1.1 Aggregate Features To Polygon – vstupní parametry

Funkcionalita

Funkcionalita nástroje je založena zejména na nástroji *Summarize Within*. Tento nástroj je součástí základní licence ArcGIS Pro jako součást toolboxu Analysis. Nástroj *Summarize Within* byl vybrán na základě doporučení Zajicové (2017) a z důvodu schopnosti pracovat se všemi základními typy geometrie. Jeho výhodou je možnost provést agregaci více atributů zároveň a pro každý z nich zvlášť zvolit počítanou statistiku. Podporuje však pouze číselné atributy, ostatní datové typy nejsou podporovány (Zajicová, 2017).

Bodová, liniová nebo polygonová data vstupují do nástroje *Summarize Within* společně s cílovou polygonovou vrstvou. Pro zvolené atributy jsou v rámci cílových zón vypočítány zvolené statistiky a výstup je uložen do nové polygonové vrstvy. Princip funkcionality nástroje je naznačen na Obr. 4.1.2.



Obr. 4.1.2 Workflow nástroje Aggregate Features To Polygon

4.2 Aggregate Polygon To Polygon

Nástroj *Aggregate Polygon To Polygon* slouží pro agregaci vstupních polygonových dat do cílových polygonových zón. Na rozdíl od předchozího nástroje pracuje pouze s polygonovými vstupními daty, neumožňuje agregovat více atributů najednou a podporuje výpočet pouze statistik SUM a MEAN. Výhodou je však možnost volby typu vstupní proměnné a náročnost výstupu na úložiště.

Vstupní parametry

Nástroj *Aggregate Polygon To Polygon* má na vstupu šest parametrů (Obr. 4.2.1), prvních pět je povinných a poslední je volitelný. *Source Zone* je zdrojová polygonová zóna obsahující agregovaný atribut. *Target Zone* je cílová polygonová zóna, do níž agregace proběhne. Datový typ obou dvou parametrů je Feature Layer. *Aggregation Field* je agregovaný atribut ze zdrojové zóny, po zvolení zdrojové zóny se zpřístupní seznam povolených polí, podobně jako u předchozího nástroje. *Variable Type* je typ vstupní proměnné podle Flowerdewa a kol. (1991). Lze zvolit mezi extenzivní a intenzivní proměnnou, kdy extenzivní zpravidla označuje počet nebo sumu (např. počet obyvatel), intenzivní značí míru nebo podíl (např. míra nezaměstnanosti). V případě, že je zvolena extenzivní proměnná, je uživateli nabídnut ještě parametr *Aggregation Type*. Typ agregace nabývá hodnot SUM a MEAN, uplatňuje se pouze pro extenzivní proměnnou a určuje, jaká statistika bude pro zvolený atribut počítána.

Obr. 4.2.1 Aggregate Polygon To Polygon – vstupní parametry

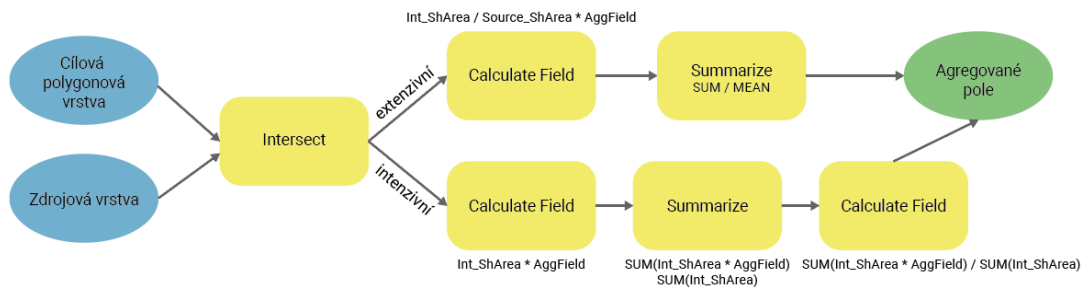
Funkcionalita

Princip nástroje je založen na polygonovém překryvu a metodě areálového vážení popsané v kapitole 3.2, protože v některých případech agregačních úloh může docházet k překryvu hranic zdrojových a cílových zón (např. agregace z budov do hexagonů). Zdrojové zóny by při polygonovém překryvu byly rozděleny do dvou cílových zón a atributy by zůstaly beze změny, tím pádem by byly započítány vícekrát. Pro polygonový překryv byl využit nástroj *Intersect*, protože do dalších výpočtů zachovává pouze místa s výskytem obou zón. Toto je využitelné zejména v případě, že zdrojové zóny kompletně nepokrývají cílové zóny. Na závěr jsou všechny dočasné vrstvy odstraněny a výstupem nástroje je pouze jeden nový atribut v cílové zóně. Princip funkcionality nástroje je naznačen na Obr. 4.2.2.

Pro výpočet areálového vážení je nejprve potřeba do atributu zdrojové i cílové zóny vypočítat jejich rozlohu. Aby nedošlo k nechtěným modifikacím obou vstupních vrstev, byly jejich zkopírováním vytvořeny dočasné vrstvy. Do nich pak byly vypočítány jejich rozlohy v km². Takto upravené vrstvy vstupovaly do polygonového překryvu a do výsledné vrstvy byla vypočítána rozloha překryvné oblasti.

Pro extenzivní proměnnou pak byla vydělena rozloha překryvné oblasti (Int_ShArea) rozlohou zdrojové oblasti (Source_ShArea) a tato hodnota byla vynásobena hodnotou agregovaného pole (AggField). S využitím nástroje *Statistics* byly výsledné hodnoty v rámci cílových zón sečteny a vytvořené pole bylo pomocí nástroje *Join Field* připojeno k cílové zóně.

Pro intenzivní proměnnou byly kurzorovými metodami odstraněny všechny prvky, u nichž agregované pole (AggField) nabývalo hodnoty Null. Rozloha překryvné oblasti pak byla vynásobena hodnotou agregovaného pole. V rámci cílových zón byla tato hodnota spolu s rozlohou zbylých překryvných zón nástrojem *Statistics* sečtena. Závěrem byl vypočítán podíl obou polí a výsledek byl připojen do cílové zóny.



Obr. 4.2.2 Workflow nástroje Aggregate Polygon To Polygon

4.3 Create Ancillary Layer

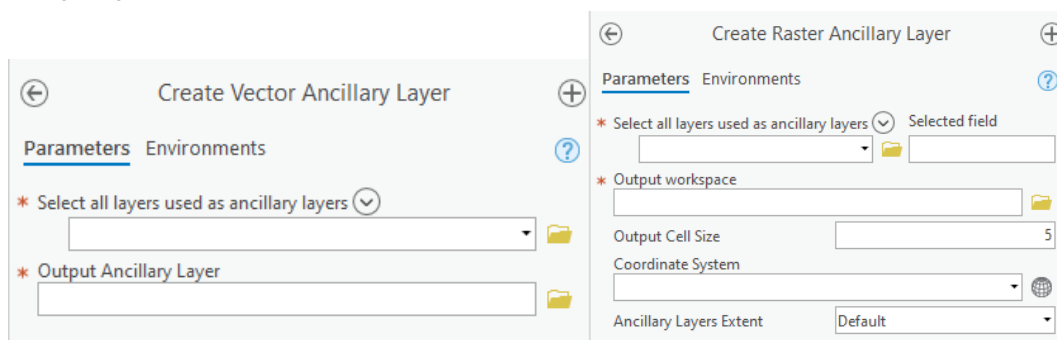
Nástroj *Create Ancillary Layer* je určen pro tvorbu vrstvy pomocných dat vstupující do procesu disagregace. Vrstvu pomocných dat je vhodné použít při disagregaci pro vysvětlení prostorového rozložení jevu. Použití samotných disagregačních nástrojů (kapitola 4.4) není podmíněno využitím tohoto nástroje, v případě využití složitějších pomocných dat nebo rasterizace vrstev pro použití v rastrové verzi disagregačního nástroje je však vhodné vrstvu pomocných dat tímto nástrojem vytvořit.

Nástroj byl vytvořen v rastrové a vektorové variantě. Rastrová verze (*Create Ancillary Raster Layer*) podporuje bodová, liniová i polygonová vstupní data, vektorová verze (*Create Ancillary Vector Layer*) na vstupu povoluje pouze polygonová vstupní data. Obě verze nástroje se od sebe liší vstupními parametry a jejich výstupem. Výstupem z rastrové varianty je jeden nebo více rastrů podle počtu vstupních atributů, zatímco vektorová verze vytváří pouze jednu polygonovou vrstvu.

Vstupní parametry

Vektorová verze nástroje vyžaduje pouze dva vstupní parametry (Obr. 4.3.1). Parametr *Participating layers* umožňuje zvolit libovolné množství vstupních polygonových vrstev a parametr *Output Ancillary Layer* slouží k volbě místa uložení a názvu vytvořené vrstvy pomocných dat. Před spuštěním nástroje je vhodné se ujistit, že všechny vstupní polygonové vrstvy obsahují atribut s váhou jednotlivých prvků.

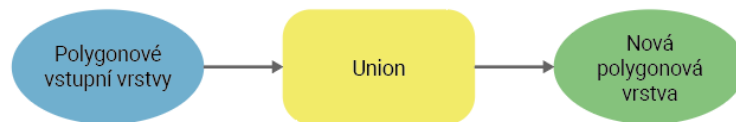
Rastrová verze nástroje má pět vstupních parametrů (Obr. 4.3.1), z toho dva povinné a tři volitelné. Parametr *Participating layers* datového typu ValueTable umožňuje vybrat více vektorových vrstev a z každé z nich zvolit jedno pole, podle něhož bude převod do rastru proveden, tedy pole obsahující váhy jednotlivých prvků. *Output workspace* slouží ke zvolení pracovního adresáře, do něž budou vytvořené rastrové vrstvy uloženy. *Output Cell Size* je nastavení velikosti buňky vytvořeného rastru, výchozí hodnota je nastavena na 5. Použitá délková jednotka je však ovlivněna vybraným souřadnicovým systémem, který je nastaven v parametru *Coordinate System*. Posledním parametrem nástroje je *Ancillary Layers Extent*, ten umožňuje vybrat prostorový rozsah vytvořené rastrové vrstvy.



Obr. 4.3.1 Create Ancillary Layer – vstupní parametry

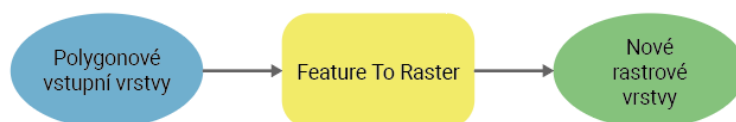
Funkcionalita

Vektorová verze nástroje funguje na principu polygonového překryvu (Obr. 4.3.2). Za účelem vytvoření unikátní kombinace všech vstupních vrstev byl použit nástroj *Union* zachovávající všechny vstupní prvky a jejich atributy. Výsledná vrstva je poté uložena pod zvoleným názvem do vybraného adresáře.



Obr. 4.3.2 Workflow nástroje Create Ancillary Vector Layer

Rastrová verze nástroje funguje na principu převodu vektorových dat do rastru (Obr. 4.3.3). Nejprve je pomocí uživatelsky definovaných parametrů nastaven pracovní adresář, extent, souřadnicový systém a velikost buňky rastru. V případě, že není vyplněn parametr *Output Coordinate System*, skript použije souřadnicový systém aktivní mapy. Všechny zvolené vrstvy jsou poté převedeny do rastru a jako hodnota rastru jsou stanoveny hodnoty zvoleného atributu.



Obr. 4.3.3 Workflow nástroje Create Ancillary Raster Layer

4.4 Disaggregate Polygon To Polygon

Nástroj *Disaggregate Polygon To Polygon* byl vytvořen jako multifunkční nástroj umožňující automatizovat metodu areálové interpolace, disagregace i dasymetrického mapování. Funkcionalita nástroje byla sestavena na základě metodiky Holta a Lu (2011). Ti rozdělují proces areálové interpolace (a tedy i disagregace) na dva kroky. Prvním krokem je provedení dasymetrického mapování, tedy přepočtu zvoleného jevu ze zdrojových zón do detailnějších dasymetrických zón. Druhým krokem je poté zpětná agregace z dasymetrických zón do cílových zón (Holt a Lu, 2011). Tato metoda je tedy využitelná pro areálovou interpolaci i disagregaci.

Z důvodu prostorové heterogenity většiny jevů je proces disagregace složitý. Aby bylo možné získat akceptovatelné výsledky přepočtů, je potřeba do analýzy doplnit další data vysvětlující tuto heterogenitu. Takovým datům se říká pomocná data (ancillary data). Pomocná data by měla být k dispozici ve větší úrovni podrobnosti než disagregovaný jev a měla by s ním co nejvíce souviset. Pomocná data jsou pak využita společně s rozlohou areálu jako váhy jednotlivých dílčích územních celků.

Nástroj byl vytvořen v rastrové i vektorové variantě, umožňuje tedy pracovat s vektorovými i rastrovými pomocnými daty. Pro tento účel je možné použít jakoukoli vektorovou vrstvu obsahující číselnou váhu v atributu nebo rastrovou vrstvu obsahující číselnou váhu jako hodnotu rastru. Vektorová verze tohoto nástroje podporuje výběr více atributů z maximálně jedné vektorové vrstvy pomocných dat. V případě potřeby použít více vrstev je vhodné nejprve použít nástroj *Create Vector Ancillary Layer* (kapitola 4.3). Rastrová verze umožňuje výběr více pomocných rastrů a stanovení váhy každého z nich. V případě potřeby tvorby těchto vstupních pomocných rastrů z vektorových prvků je vhodné nejprve využít nástroj *Create Raster Ancillary Layer* (kapitola 4.3).

4.4.1 Vektorová verze

Vektorová verze nástroje *Disaggregate Polygon To Polygon* byla vytvořena pro disagregaci hodnoty jevu z vektorových zdrojových zón do vektorových cílových zón s možností použití vektorové vrstvy pomocných dat. Nástroj umožňuje provádět přepoččet pouze pro jeden atribut současně. Podle zvoleného typu disagregované proměnné (extenzivní/intenzivní) se pak liší samotný výpočet. Výstupem nástroje je pouze jeden sloupec v atributové tabulce cílové zóny, ostatní dočasná data jsou na závěr smazána.

Vstupní parametry

Do vektorové verze nástroje vstupuje celkem sedm proměnných (viz Obr. 4.4.2), z nichž je pět povinných a dvě volitelné. Proměnná *Source Zone* datového typu Feature Layer je zdrojová polygonová vrstva obsahující disagregovaný atribut. *Target Zone* je cílová polygonová vrstva, do níž disagregace proběhne, datový typ proměnné je opět Feature Layer. *Disaggregated Field* je pole ze zdrojové zóny obsahující hodnoty disagregovaného jevu. *Output Field Name* umožňuje uživateli stanovit název nově vytvořeného pole s přepočtenými hodnotami. Proměnná *Variable Type* umožňuje stanovit typ přepočítávané proměnné. *Ancillary Layer* je bodová, liniová nebo polygonová vrstva pomocných dat. Tato proměnná je volitelná, není tedy potřeba pomocná data specifikovat. *Ancillary Data Fields* umožňuje vybrat jedno nebo více polí z pomocných dat a stanovit jeho váhu ve výpočtu.

Obr. 4.4.2 Disaggregate Polygon To Polygon (Vector version) – vstupní parametry

Funkcionalita

Princip fungování nástroje je založen na polygonovém překryvu mezi zdrojovou zónou, cílovou zónou a případně vrstvou pomocných dat. Pro polygonový překryv byl využit nástroj *Intersect*. Výstupem nástroje *Intersect* je nová polygonová vrstva s geometrií pouze v místech, kde se překrývají všechny vrstvy zároveň. Výpočet se tedy provádí pouze pro lokality, v nichž se daný jev vyskytuje. Oproti nástroji *Union* je snížena datová a výpočetní náročnost. Zároveň jsou do nové vrstvy přeneseny i atributy ze všech původních vrstev. Po doběhnutí výpočtů jsou všechny dočasné vrstvy odstraněny a výstupem nástroje je pouze jeden nový atribut v cílové zóně. Princip funkcionality nástroje je naznačen na Obr. 4.4.9.

Prvním krokem nástroje je vytvoření kopie všech vstupních vrstev, aby byla zajištěna bezpečná manipulace s jejich atributy. Do atributů zdrojových a cílových zón je vypočítána jejich rozloha v km². Nad vrstvou zdrojových zón, cílových zón a pomocných

dat je vypočítán polygonový překryv. V případě, že pomocná data nejsou definována, probíhá polygonový překryv jen nad zdrojovou a cílovou vrstvou. Pokud je výsledná dasymetrická vrstva polygonová, je jí do nového atributu vypočítána rozloha. Pro liniovou vrstvu je pak do nového atributu vypočítána délka linií a pro bodovou vrstvu je přiřazena hodnota 1. Váha dasymetrické zóny je pak spočítána podle Obr. 4.4.3, kde W_i je váha zóny i , a_1 až a_n je atribut zvolený pro výpočet a w_1 až w_n je jeho váha.

$$W_i = a_1 \times w_1 + \dots + a_n \times w_n$$

Obr. 4.4.3 Vzorec výpočtu váhy dasymetrické zóny

Pro **extenzivní proměnnou** pak byla váha dasymetrické zóny vynásobena její rozlohou. Pomocí nástroje *Statistics* byl výsledek v rámci zdrojových zón sečten. Pro každou dasymetrickou zónu pak byla vypočítána hodnota jevu podle Obr. 4.4.4, kde P_i je hodnota jevu v dasymetrické zóně, P_s je hodnota jevu ve zdrojové zóně, A_i je rozloha dasymetrické zóny a W_i je váha dasymetrické zóny.

$$P_i = P_s \times \frac{A_i \times W_i}{\sum_i (A_i \times W_i)}$$

Obr. 4.4.4 Výpočet hodnoty extenzivní proměnné v dasymetrické zóně s pomocnými daty

V případě, že součástí analýzy nejsou pomocná data, a tedy není definována žádná váha, probíhá stejný výpočet, akorát bez váhy pomocných dat (Obr. 4.4.5). Jako váha jsou nyní použity pouze rozlohy jednotlivých dasymetrických zón, jedná se tedy o metodu areálového vážení.

$$P_i = P_s \times \frac{A_i}{\sum_i A_i}$$

Obr. 4.4.5 Výpočet hodnoty extenzivní proměnné v dasymetrické zóně bez pomocných dat

Vypočtené hodnoty jsou pak pomocí nástroje *Statistics* agregovány (sečteny) v rámci cílové zóny. Vzniklý atribut byl připojen do cílové zóny a všechny dočasné vrstvy byly smazány.

Pro **intenzivní proměnnou** byla váha dasymetrické zóny vynásobena její rozlohou. Pomocí nástroje *Statistics* byl z tohoto součinu vypočítán průměr. Pro každou dasymetrickou zónu pak byla vypočítána hodnota jevu podle Obr. 4.4.6, kde P_i je hodnota jevu v dasymetrické zóně, P_s je hodnota jevu ve zdrojové zóně, A_i je rozloha dasymetrické zóny, W_i je váha dasymetrické zóny a n je počet dasymetrických zón ve zdrojové zóně.

$$P_i = P_s \times \frac{A_i \times W_i}{\frac{\sum_{i=1}^n (A_i \times W_i)}{n}}$$

Obr. 4.4.6 Výpočet hodnoty intenzivní proměnné v dasymetrické zóně s pomocnými daty

V případě absence pomocných dat je ve výpočtu, stejně jako pro extenzivní proměnnou, vynechána váha dasymetrické zóny (Obr. 4.4.7). Jako váha dasymetrických zón jsou použity pouze jejich rozlohy, jde tedy o metodu areálového vážení.

$$P_i = P_s \times \frac{A_i}{\frac{\sum_{i=1}^n A_i}{n}}$$

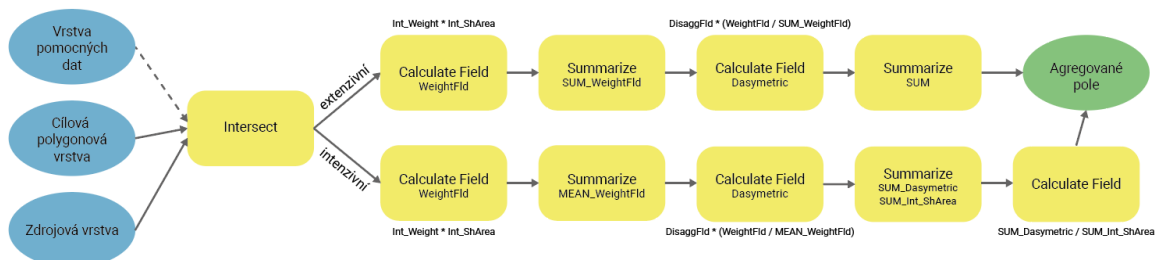
Obr. 4.4.7 Výpočet hodnoty intenzivní proměnné v dasymetrické zóně bez pomocných dat

Pomocí nástroje *Statistics* jsou pak v rámci cílových zón sečteny hodnoty jevu jednotlivých dasymetrických zón a jejich rozlohy. Výsledný výpočet hodnoty jevu v cílové zóně je pak naznačen na Obr. 4.4.8. P_t je hodnota jevu v cílové zóně, $\sum_i P_i$ je suma hodnot

jevu v dasymetrických zónách v rámci cílové zóny a $\sum_i A_i$ je suma rozloh dasymetrických zón v rámci cílové zóny. Výsledné pole je poté připojeno do cílové zóny jako nový atribut a všechny dočasné vrstvy jsou smazány.

$$P_t = \frac{\sum_i P_i}{\sum_i A_i}$$

Obr. 4.4.8 Zpětná agregace hodnot intenzivní proměnné do cílové zóny



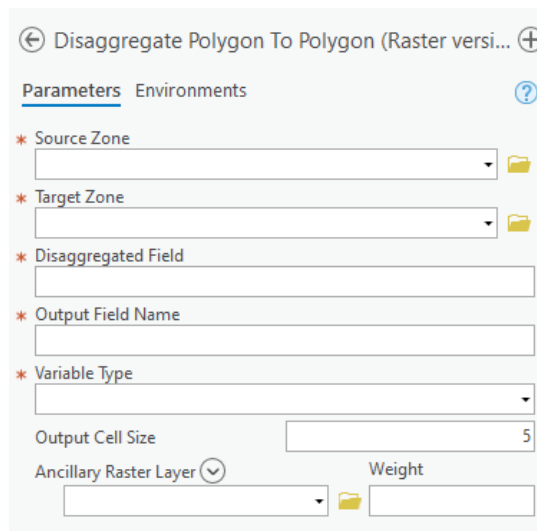
Obr. 4.4.9 Workflow nástroje Disaggregate Polygon To Polygon (Vector version)

4.4.2 Rastrová verze

Rastrová verze nástroje *Disaggregate Polygon To Polygon* byla vytvořena pro disagregaci hodnoty jevu z vektorových zdrojových zón do vektorových cílových zón s možností použití rastrové vrstvy pomocných dat. Nástroj umožňuje zvolit typ počítané proměnné a umožňuje provádět přepočítání pouze pro jeden atribut současně. Výstupem nástroje je jeden sloupec v atributové tabulce cílové zóny, ostatní vytvořená data jsou po dobehnutí nástroje smazána.

Vstupní parametry

Rastrová verze nástroje má na vstupu celkem sedm parametrů (Obr. 4.4.10). Prvních pět parametrů je povinných a zbylé dva volitelné. *Source Zone* je zdrojová zóna obsahující agregovaný atribut, *Target Zone* je pak cílová zóna, do níž bude zvolený atribut agregován. *Disaggregated Field* je pole ze zdrojové zóny obsahující hodnoty disagregovaného jevu, *Output Field Name* umožňuje uživateli stanovit název výsledného pole s agregovanými hodnotami. Proměnná *Variable Type* umožňuje stanovit typ přepočítávané proměnné, na výběr je mezi možnostmi intensive a extensive. *Output Cell Size* je velikost buňky vytvářených rastrů, nad nimiž analýza proběhne. Čím větší velikost buňky, tím větší přesnost výpočtu. Se snižující se velikostí buňky však roste výpočetní náročnost analýzy. Velikost buňky je v jednotkách aktuálně používaného souřadnicového systému. *Ancillary Raster Layer* umožňuje vybrat více rastrových vrstev pomocných dat a stanovit jejich váhy.



Obr. 4.4.10 Disaggregate Polygon To Polygon (Raster version) – vstupní parametry

Funkcionalita

Princip fungování nástroje je založen na jazyku mapové algebry, pro výpočet hodnot rastrů využívá sadu funkcí *Map Algebra* pro arccpy. *Map Algebra* je součástí extenze Spatial Analyst a je k dispozici pod licencí Spatial Analyst nebo Image Analyst. Jedná se o sadu funkcí, nástrojů a operátorů pro geografické analýzy s pomocí rastrů (Esri, 2021b). Po dokončení výpočtu jsou všechny dočasné vrstvy i rastry smazány a výsledkem je pouze nové pole s přepočítanými hodnotami v cílové zóně. Princip funkcionality nástroje je naznačen na Obr. 4.4.14.

Z důvodu manipulace s atributy zdrojových a cílových zón jsou nejprve tyto zóny zkopírovány a uloženy do dočasné vrstvy. Zdrojová zóna je převedena na rastr a jako hodnota rastru je nastaven disagregovaný atribut. Poté je pomocí jazyka mapové algebry vytvořen váhový rastr jako součet jednotlivých pomocných rastrů vynásobených jejich vahou (viz Obr. 4.4.3). V případě, že pomocné rastry nejsou stanoveny, vznikne pomocný váhový rastr s konstantní vahou 1. Extent tohoto rastru je stanoven polygonovým překryvem mezi zdrojovou a cílovou zónou pomocí nástroje *Union* a velikost buňky je nastavena pomocí vstupního parametru *Cell Size*.

Pro **extenzivní proměnnou** v každém pixelu pomocí mapové algebry spočítána předpokládaná hodnota jevu tak, že váhový rastr je vydělen sumou vah v rámci zdrojové zóny a vynásoben disagregovanou hodnotou (Obr. 4.4.11). P_i je hodnota jevu v pixelu, P_s je hodnota jevu ve zdrojové zóně a W_i je váha jevu v pixelu.

$$P_i = P_s \times \frac{W_i}{\sum_i W_i}$$

Obr. 4.4.11 Výpočet hodnoty extenzivní proměnné v dasymetrické zóně

Pomocí nástroje *Zonal Statistics as Table* jsou pak všechny pixely v rámci cílové zóny sečteny a výsledné pole je připojeno do cílové zóny.

Pro **intenzivní proměnnou** je nejprve pomocí nástroje *Zonal Statistics* vypočítána průměrná váha rastru v rámci zdrojové zóny. Poté je v každém pixelu spočítána hodnota jevu tak, že váhový rastr je vydělen průměrnou vahou ve zdrojové zóně a vynásoben disagregovanou hodnotou (Obr. 4.4.12). P_i je hodnota jevu v pixelu, P_s je hodnota jevu ve zdrojové zóně, W_i je váha pixelu a n je počet pixelů ve zdrojové zóně.

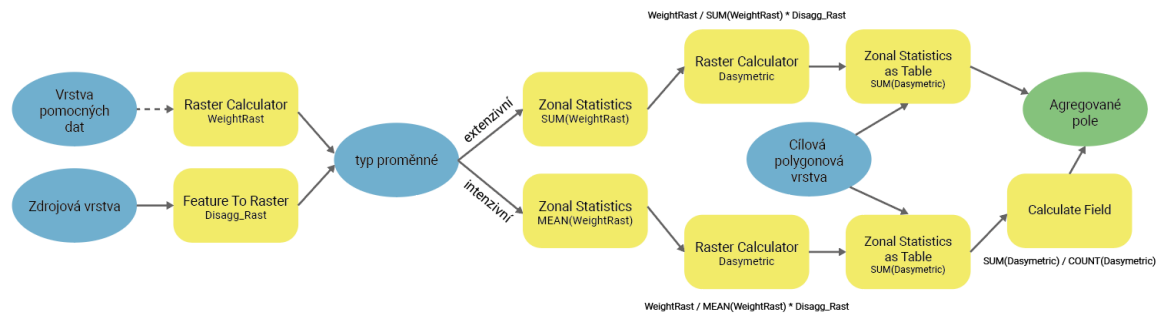
$$P_i = P_s \times \frac{W_i}{\frac{\sum_{i=1}^n W_i}{n}}$$

Obr. 4.4.12 Výpočet hodnoty intenzivní proměnné v pixelu

Pomocí nástroje *Zonal Statistics as Table* jsou poté v rámci cílové zóny sečteny všechny pixely v rámci cílové zóny. Součet hodnot jevu v cílové zóně je nakonec vydělen počtem pixelů v cílové zóně, viz Obr. 4.4.13. P_t je hodnota jevu v cílové zóně, $\sum_{i=1}^n P_i$ je suma pixelů v rámci cílové zóny a n je počet pixelů v rámci cílové zóny. Výsledné pole je připojeno do cílové zóny jako nový atribut, všechny dočasné vrstvy a rastry jsou smazány.

$$P_t = \frac{\sum_{i=1}^n P_i}{n}$$

Obr. 4.4.13 Zpětná agregace hodnot intenzivní proměnné do cílové zóny



Obr. 4.4.14 Workflow nástroje Disaggregate Polygon To Polygon (Raster version)

5 TESTOVÁNÍ NÁSTROJŮ

Dilčím cílem diplomové práce mimo tvorby nástrojů bylo testování jejich funkčnosti na referenčních datech. Testování nástrojů sloužilo zejména ke zkontrolování funkcionality a odladění chyb v kódu nebo v nastavení parametrů. Testován byl však i samotný způsob přepočtů, zda zvolená logika přepočtů pro oba typy proměnné vykazuje akceptovatelné výsledky a zda byla tato logika správně implementována do vytvořených nástrojů.

Prvním krokem práce byla tvorba hexagonového gridu pro celé území ČR v různých měřících, aby došlo ke snížení vlivu agregačního problému (viz kapitola 3.1). Shodnost buněk gridu umožňuje zajistit vzájemnou srovnatelnost různých analýz. Pro generování hexagonových sítí byl využit nástroj *Generate Tessellation*. Vygenerované gridy, počet jejich buněk a dobu trvání generování znázorňuje Tab. 5.1. Jednou z nevýhod gridové sítě je podle Zajícové (2017) neschopnost dokonale kopírovat hranice zájmového území, ve většině případů tak dochází k překryvům buněk přes hranice zájmového území. V případě zachování překryvů všech buněk je nutné do hodnoty jevu v hraničních buňkách zahrnout i hodnoty jevu mimo zájmové území. Výhodou tohoto způsobu je zachování vzájemné srovnatelnosti všech buněk a zachování pravidelnosti gridu. Nevýhodou je však nutnost zahrnout do analýzy hodnoty jevu a případně také pomocná data mimo zájmové území. Pokud by nebyly do hodnot hraničních buněk zahrnuty i hodnoty jevu mimo sledované území, neodpovídaly by hodnoty jevu skutečnosti a při jejich přepočtu na jednotku plochy by docházelo ke zkreslení. Druhou možností řešení je ořezat vygenerovanou gridovou síť hranicí zájmového území. Tento způsob řeší problém s přeshraničními hodnotami, nevýhodou je však ztráta pravidelnosti a vzájemné srovnatelnosti. Z důvodu dostupnosti dat byla v rámci práce použita metoda ořezu příhraničních buněk. Výsledné gridové sítě tedy byly ořezány státní hranicí ČR z dat RÚIAN pomocí nástroje Clip.

Tab. 5.1 Vzájemné srovnání vygenerovaných hexagonových sítí

Rozloha buňky	0,25 km²	0,5 km²	1 km²	2,5 km²	5 km²	25 km²
Doba generování sítě	9min 56s	4min 57s	2min 48s	1min 6s	34s	14s
Počet buněk	711 770	356 202	178 215	71 415	36 015	7 260
Počet buněk v ČR	317 832	159 464	80 046	32 299	16 279	3 363
Podíl buněk v ČR	44,6 %	44,76 %	44,92 %	45,23 %	45,20 %	46,32 %

Na základě rešerše a získaných poznatků v rámci bakalářské práce byly vybrány vhodné statistické ukazatele různého prostorového rozsahu a vhodná pomocná data pro jejich přepočet. Vybrané ukazatele byly s pomocí zvoleného vytvořeného nástroje přepočteny do cílových územních jednotek nebo gridových sítí. Vybrané výsledky těchto přepočtů pak byly vizualizovány v souladu s kartografickými pravidly a jsou součástí příloh této diplomové práce. Volba testovaných ukazatelů a pomocných dat pro každý nástroj je pak uvedena v podkapitolách kapitoly 5. Podkapitoly jsou rozčleněny podle testovaných nástrojů. Níže uvedené doby trvání analýz byly testovány na zařízení s konfigurací CPU AMD Ryzen 7 1700 @ 3,50 GHz, GPU NVIDIA GeForce GTX 1060 6GB, RAM 16 GB.

5.1 Aggregate Features To Polygon

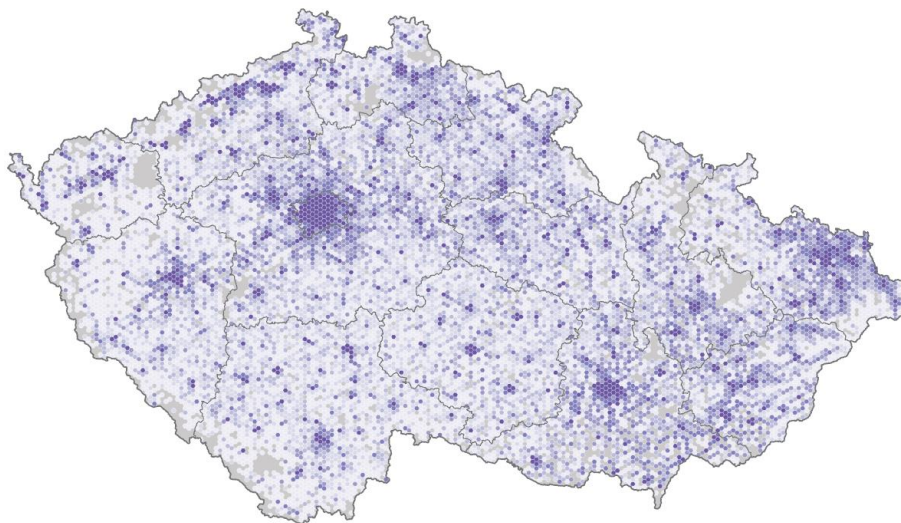
Nástroj *Aggregate Features To Polygon* umožňuje prostorově agregovat bodová, liniová nebo polygonová data do polygonových územních jednotek. Vychází zejména z nástroje *Summarize Within*. V rámci testování bylo ověřováno a opravováno zejména správné nastavení vstupních parametrů ve vlastnostech nástroje. Funkční nástroj byl poté prakticky otestován na několika bodových, liniových i polygonových datových sadách. Vzhledem ke své univerzálnosti byl poté tento nástroj použit pro agregaci statistických údajů ve většině dalších analýz.

Počet obyvatel

Nástroj byl testován zejména na agregaci adresních bodů a budov do územních jednotek nebo gridových sítí. Vrstva adresních bodů „Statistické budovy – vchody k bytům“ od ČSÚ byla nejprve agregována do vrstvy ZSJ ČR. V rámci přepočtu byly agregovány atributy

- BUDOBYTSL – počet obyvatel s trvalým pobytem,
- BUDOBYOSL – počet obyvatel s obvyklým pobytem,
- COUNT – počet adresních bodů,
- SUM_BYT – počet bytů,
- POCPODBUD – počet podlaží budov.

Tato agregace trvala na výše specifikovaném zařízení 8 minut a 55 vteřin. Stejná vrstva adresních bodů poté byla agregována do hexagonů o rozloze 5 km² v rámci celé ČR a o rozloze 0,5 km² v rámci Olomouckého kraje. Agregace do hexagonů ČR zabrala 8 minut a 17 vteřin, do hexagonů Olomouckého kraje pak 55 vteřin. Z agregovaných dat pak byla vizualizován údaj o počtu bytů na km² v hexagonech ČR v roce 2016 (Obr. 5.1.1).

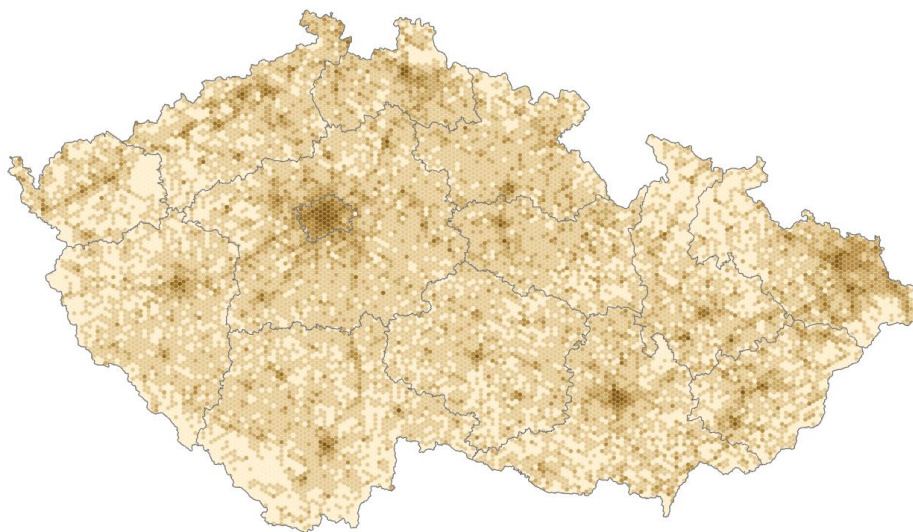


Obr. 5.1.1 Počet bytů na km² v hexagonech ČR v roce 2016 – agregací z adresních bodů

V rámci přepočtu počtu obyvatel byl tento nástroj testován ještě pro agregaci dat z polygonů do polygonů. Z veřejné databáze ČSÚ byla pro všechny okresy stažena data o počtu obyvatel podle pohlaví a hlavních věkových skupin za obce ČR platná k datu 31. 12. 2019 (ČSÚ, 2021). Tato vrstva byla poté v rámci testování agregována do krajů ČR. Byly agregovány atributy POCOB (počet obyvatel), MUZI a ZENY (počet mužů a žen), PREDPRODUKTIVNI, PRODUKTIVNI a POSTPRODUKTIVNI (počet obyvatel v hlavních věkových skupinách). Tato agregace trvala 1 minutu a 59 vteřin.

Hustota silniční sítě

Dalším testovaným ukazatelem byla hustota silniční sítě. Za tímto účelem byla použita data OSM platná k datu 7. července 2020. Z kompletní datové sady byly vybrány pouze zpevněné cesty (living street, motorway, primary, secondary, tertiary, trunk a unclassified). Výběr byl agregován nejprve do hexagonů ČR o velikosti 5km² a poté do hexagonů Olomouckého kraje o velikosti 0,5km². V rámci obou přepočtů byl agregován atribut Shape_Length označující délku daného úseku silnice v metrech. Výsledná suma délky silniční sítě byla poté podělena rozlohou hexagonu a zvizualizována jako hustota silniční sítě v km na km² v hexagonech ČR v roce 2020 (Obr. 5.1.2). Agregace silniční sítě do hexagonů ČR trvala 24 minut a 29 vteřin, do hexagonů Olomouckého kraje pak 2 minuty a 4 vteřiny.



Obr. 5.1.2 Hustota silniční sítě v hexagonech ČR v roce 2020 – agregací z linií

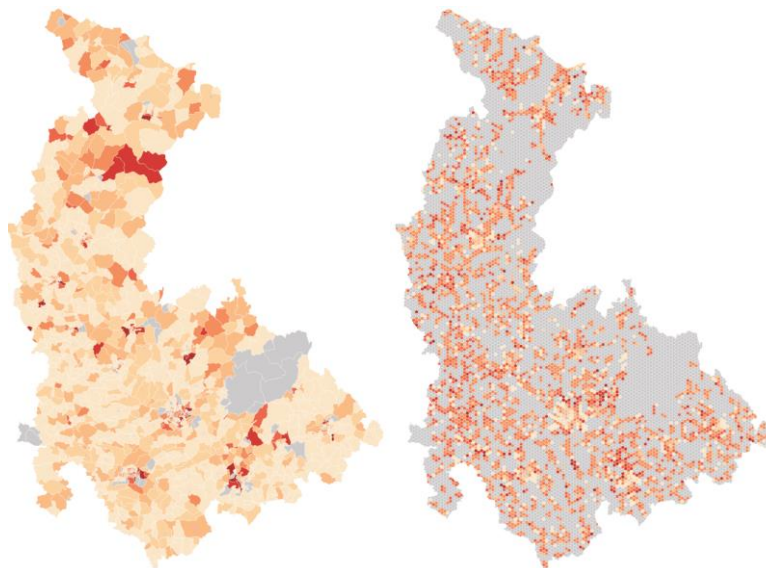
Index kriminality

Posledním testovaným ukazatelem byl index kriminality. Data pro tento účel byla stažena z mapové aplikace *Mapa kriminality* Policie ČR. Data jsou z důvodu ochrany osobních údajů polohově zkreslena s využitím Voroného tessellace a clusterovací funkce K-Means (Policie ČR, 2021). Z tohoto důvodu byl zkreslen i výsledný index kriminality pro data agregovaná do menších územních jednotek než do obcí. V rámci dat jsou rovněž zobrazovány jen ty delikty, u nichž může mít jejich zveřejnění preventivní charakter a jejich zveřejnění nebrání taktické hledisko (Policie ČR, 2021). Zveřejněná data tedy nejsou zcela kompletní, v přílohách mapové aplikace *Mapa kriminality* lze zjistit zahrnuté a vynechané typy trestných činů a přestupků. Z důvodu citlivosti dat byly vynechány zejména kategorie mravnostní trestná činnost, trestná činnost na mládeži nebo nelegální migrace. V rámci zahrnutých kategorií jdou pak vynechány trestné činy jako například násilné přerušování těhotenství nebo týrání osoby, dítěte či zvířete (Policie ČR, 2021).

Data obsahují 13 atributů definujících zahrnutý druh deliktu. Atributy mohou nabývat hodnoty 1 (tento delikt se vyskytuje) nebo Null (tento delikt se v daném bodě nevyskytuje). Jeden záznam pak může obsahovat více druhů deliktů, například dopravní nehoda s řidičem pod vlivem drog. Podle Otevřené společnosti (2021) je pak index kriminality definován jako počet zjištěných skutků za zvolené období přepočtený na 10 000 obyvatel.

Ze stažených datových sad byly pro výpočet indexu kriminality vybrány pouze delikty za rok 2020, dále byly vynechány dopravní nehody a přestupky bez dalších souběžných deliktů a delikty bez kategorie. Výsledná bodová vrstva byla pomocí tohoto nástroje

agregována do hexagonů Olomouckého kraje o rozloze 0,5 km² a do ZSJ Olomouckého kraje, obě analýzy trvaly 20 vteřin. Pro výpočet indexu kriminality bylo potřeba z obcí Olomouckého kraje disagregovat do zvolených územních jednotek počet obyvatel ke dni 31. 12. 2019. Tento přepočítání proběhl pomocí vektorové verze nástroje *Disaggregate Polygon To Polygon* s využitím počtu podlaží a počtu bytů budov jako pomocných dat. Index kriminality byl poté vizualizován za hexagony a ZSJ Olomouckého kraje (Obr. 5.1.3).



Obr. 5.1.3 Index kriminality v ZSJ a hexagonech Olomouckého kraje v roce 2020

5.2 Aggregate Polygon To Polygon

Nástroj *Aggregate Polygon To Polygon* umožňuje prostorově agregovat polygonová data do polygonových územních jednotek. Je založen na kombinaci polygonového překryvu a metody areálového vážení. V rámci testování nástroje byla opravována hlavně logika přepočtu pro intenzivní proměnnou a logika mazání dočasných vrstev. Funkční nástroj byl poté otestován na několika datových sadách.

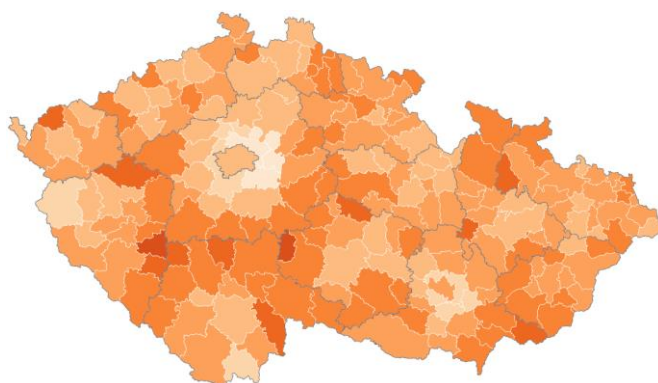
Počet obyvatel

V rámci přepočtu počtu obyvatel byl tento nástroj testován pro agregaci z obcí do SO ORP ČR. Za tímto účelem byla využita data z Veřejné databáze ČSÚ platná ke dni 31. 12. 2019. Byly agregovány atributy POCOB (počet obyvatel), MUZI a ZENY (počet mužů a žen), PREDPRODUKTIVNI, PRODUKTIVNI a POSTPRODUKTIVNI (počet obyvatel v hlavních věkových skupinách). Agregace každého z těchto atributů trvala 58 vteřin. Tento nástroj tedy dokáže zpracovat jeden přepočítávaný atribut rychleji než předchozí nástroj, v případě přepočtu tří a více atributů je již vhodnější použít nástroj *Aggregate Features To Polygon* (kapitola 5.1). Výhodou tohoto nástroje oproti předchozímu nástroji je však agregace přepočtených dat do stávající vrstvy.

Průměrný věk

Dalším zvoleným ukazatelem byl průměrný věk. Informace o průměrném věku v obcích ČR byla stažena z Veřejné databáze ČSÚ společně s daty o počtu obyvatel a jsou platná ke dni 31. 12. 2019 (ČSÚ, 2021). V rámci přepočtu průměrného věku byl tento nástroj testován pro agregaci z obcí do SO ORP ČR. Přepočítávaný atribut PRUMERNY_VEK jako intenzivní proměnná, přepočítání trvalo 1 minutu a 5 vteřin. Výsledná

vrstva byla poté vizualizována jako průměrný věk v SO ORP ČR ke dni 31. 12. 2019 (Obr. 5.2.1).

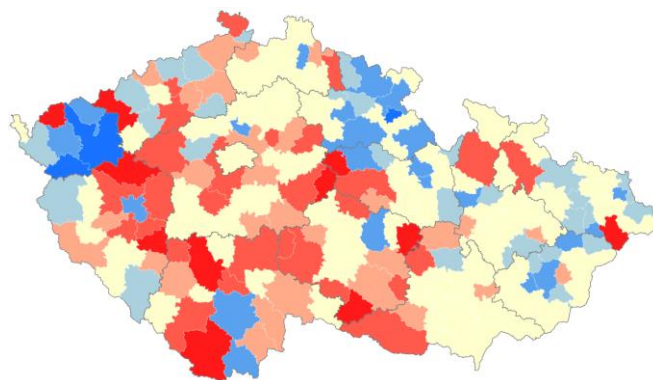


Obr. 5.2.1 Průměrný věk v SO ORP ČR ke dni 31. 12. 2019 – agregaci z obcí

Index stáří

Dalším ukazatelem využitým pro testování tohoto nástroje byl index stáří. Index stáří byl agregován z obcí do SO ORP dvěma způsoby. Prvním z nich byla agregace hodnoty indexu jako intenzivní proměnné. Za tímto účelem byl nejprve v rámci jednotlivých obcí vypočítán index stáří jako podíl obyvatel v postproduktivním věku a v předproduktivním věku, vynásobený 100. Informace o počtu obyvatel v předproduktivním a postproduktivním věku za obce ČR byly získány z Veřejné databáze ČSÚ ke dni 31. 12. 2019 (ČSÚ, 2021). Výsledný index byl poté agregován do SO ORP, agregace trvala 81 sekund.

Druhým a přesnějším způsobem agregace indexu stáří byla agregace počtu obyvatel v předproduktivním a postproduktivním věku do SO ORP samostatně jako extenzivní proměnné. Index stáří pak byl vypočten z těchto agregovaných hodnot. Oba přístupy byly od sebe následně odečteny a výsledný rozdíl byl vizualizován s využitím areálové metody s divergentní stupnicí, kdy modré hodnoty značí podhodnocenou hodnotu indexu, červené hodnoty nadhodnocenou hodnotu indexu a žluté hodnoty značí agregaci s odchylkou indexu do 5 bodů (Obr. 5.2.2).



Obr. 5.2.2 Rozdíl mezi agregací indexu stáří jako intenzivní a extenzivní hodnoty

Podíl nezaměstnaných osob

Posledním testovaným ukazatelem byl podíl nezaměstnaných osob. Pro agregaci podílu nezaměstnaných osob byla použita data Uchazeči o zaměstnání v evidenci úřadu práce ke dni 31. 12. 2019 (ČSÚ, 2021). Data byla stažena za obce ČR z Veřejné databáze ČSÚ. Podle MPSV (2021) podíl nezaměstnaných osob vyjadřuje podíl dosažitelných uchazečů o zaměstnání ve věku 15–64 let ze všech obyvatel ve stejném věku. Výsledná

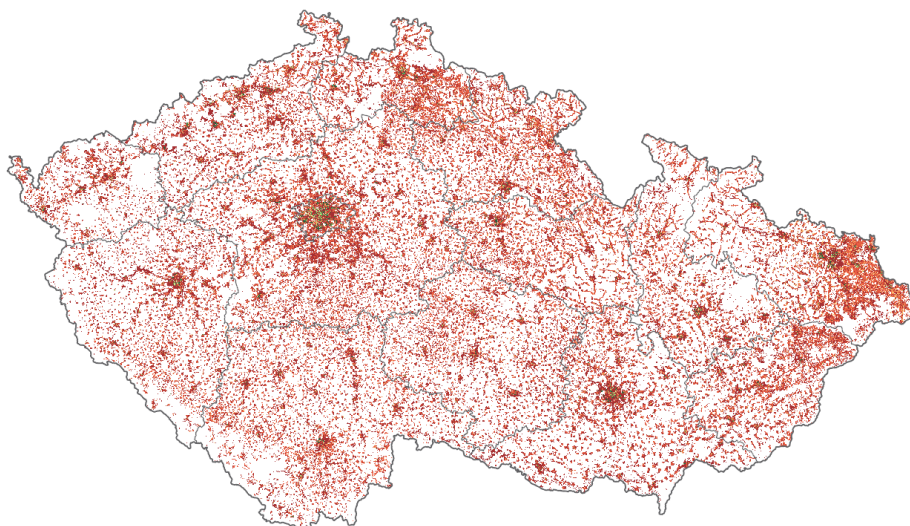
hodnota je poté vynásobením 100 převedena na procenta. Vypočítaná hodnota indexu pak byla agregována do SO ORP ČR jako intenzivní proměnná, výpočet trval 68 vteřin.

5.3 Create Ancillary Layer

Nástroj *Create Ancillary Layer* usnadňuje tvorbu vrstvy pomocných dat vstupující do nástrojů pro disagregaci. Nástroj byl vytvořen ve vektorové a rastrové variantě. Vektorová varianta funguje na bázi polygonového překryvu vstupních vrstev a lze ji využít, pokud do disagregace vstupuje více polygonových vrstev pomocných dat. Výstupem je jedna polygonová vrstva obsahující geometrie a atributy ze všech vstupních vrstev. Rastrová varianta funguje na bázi převodu polygonů na rastr o zvolené velikosti buňky. Jako hodnota rastru je pak stanovena hodnota pole obsahujícího pomocná data. Výstupem je jeden nebo více rastrů, kdy hodnota rastru je určena hodnotou jednoho atributu.

Nástroj byl nejprve vyvíjen jako jeden nástroj pro převod vrstev pomocných dat do vektoru nebo do rastru, z důvodu velké odlišnosti vstupních parametrů obou variant byl však rozdělen na dva nástroje. V rámci testování byly odstraňovány zejména chyby způsobené různými způsoby definice vstupních dat a chyby způsobené mezerami v adresářových strukturách.

Rastrová varianta nástroje byla využita pro tvorbu pomocných dat vstupujících do disagregace počtu obyvatel. Pro přepočítání počtu obyvatel byly využity na základě doporučení z bakalářské práce (Zapletal, 2019) stavební objekty z databáze RÚIAN a jejich atributy – počet bytů a počet podlaží. Za tímto účelem byly z polygonové vrstvy stavebních objektů vygenerovány rastry s informacemi o počtu bytů a počtu podlaží o hraně buňky 5 metrů v rámci celé České republiky a Olomouckého kraje. V rámci Olomouckého kraje mají vytvořené rastry 17500×25850 buněk, nekomprimovaná velikost jednoho rastru je 431,67 MB a tvorba obou rastrů proběhla za 1 minutu a 5 vteřin. Pro celou ČR má vytvořený rastr 101290×70090 buněk, nekomprimovaná velikost jednoho rastru je 13,22 GB a tvorba obou rastrů proběhla za 14 minut a 49 vteřin. Se snižující se velikostí buňky se zvyšuje velikost rastru i doba přepočtu, v případě využití buněk větší velikosti však mohou být následné disagregace méně přesné.



Obr. 5.3.1 Počet bytů ve stavebních objektech České republiky – rastr o velikosti 5 m

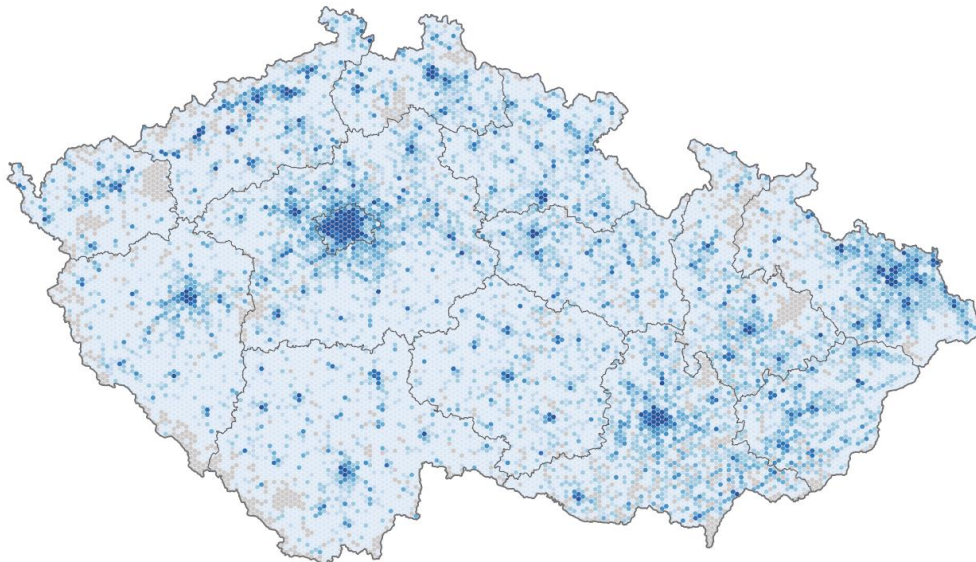
5.4 Disaggregate Polygon To Polygon

Nástroj *Disaggregate Polygon To Polygon* slouží pro automatizaci přepočtu agregovaného statistického jevu do nižších nebo jiných územních jednotek, umožňuje tedy provádět jak disagregaci, tak i areálovou interpolaci. Nástroj byl vytvořen ve vektorové a rastrové variantě pro oba dva typy pomocných dat. Obě varianty nástroje fungují na principu dasymetrického mapování, kdy nejprve proběhne přepočet do detailních dasymetrických zón a poté dochází ke zpětné agregaci do cílových zón. Kompletní popis funkcionality nástrojů je k dispozici v kapitole 4.4. Nástroje rovněž umožňují použít pro zpřesnění výpočtu pomocná data, jejich použití však není povinné.

V rámci testování nástrojů byla opravována zejména logika přepočtů pro oba typy proměnné. Dále byl upraven způsob ukládání dočasných dat. Místo pracovního adresáře *in_memory* používajícího paměť RAM byl použit výchozí pracovní adresář (výchozí geodatabáze), do nějž jsou dočasné vrstvy uloženy. Po doběhnutí nástroje jsou z něj všechny vytvořené dočasné vrstvy odstraněny. Funkční nástroje byly poté otestovány na několika statistických ukazatelích v rámci Olomouckého kraje i České republiky.

Počet obyvatel

Počet obyvatel byl nejčastěji využitým statistickým jevem pro testování vytvořených disagregačních nástrojů. V rámci celé ČR byla rastrová verze nástroje testována na areálové interpolaci počtu obyvatel z obcí do hexagonů o ploše 5 km². Pro tento přepočet byla použita data o počtu obyvatel za obce České republiky z Veřejné databáze ČSÚ platná ke dni 31. 12. 2019 (ČSÚ, 2021). Jako vrstva pomocných dat byly využity pomocné rastry počet bytů a počet podlaží vygenerované rastrovou verzí nástroje *Create Ancillary Layer* ze stavebních objektů z databáze RÚIAN. Výsledný počet obyvatel pak byl vizualizován jako hustota zalidnění v hexagonové síti ČR v roce 2019 (Obr. 5.4.1). Areálová interpolace počtu obyvatel do hexagonů ČR trvala 1 hodinu 1 minutu a 33 vteřin.



Obr. 5.4.1 Hustota zalidnění v hexagonové síti ČR v roce 2019 – areálová interpolace z obcí

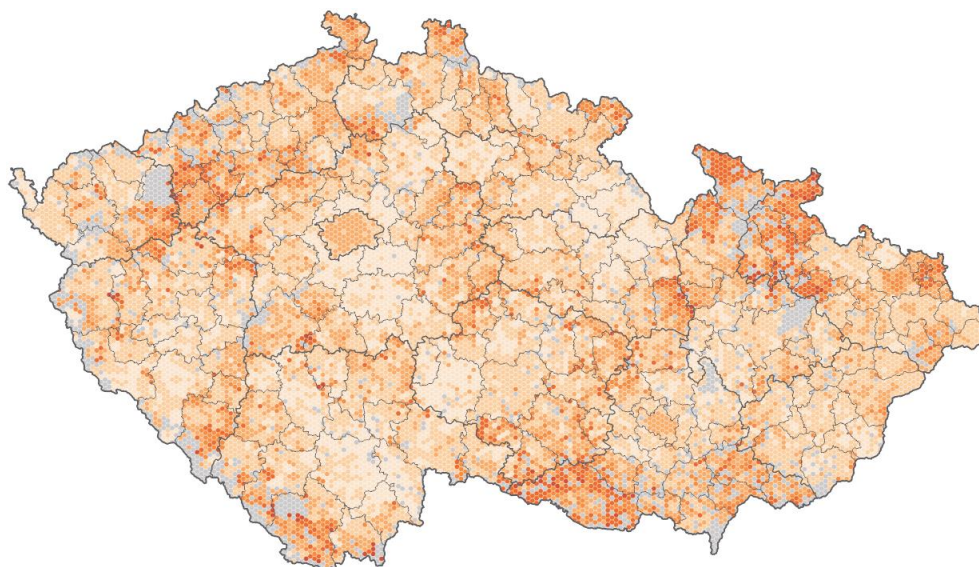
Počet obyvatel byl přepočítáván také vektorovou verzí nástroje. Z důvodu rozsáhlého zájmového území a z důvodu detailu použitých pomocných dat byl přepočet vektorovou variantou nástroje pro celé území ČR výpočetně a časově náročný. Testovaný přepočet počtu obyvatel z obcí ČR do hexagonů o rozloze 5 km² s využitím stavebních objektů jako pomocných dat trval celkem 2 hodiny a 21 minut. Z tohoto důvodu byla vektorová

varianta testována výhradně v rámci testovacího území Olomouckého kraje. V rámci testování přepočtu počtu obyvatel v Olomouckém kraji bylo srovnáno několik různých disagregačních metod a druhů pomocných dat vstupujících do přepočtu. Výsledky těchto přepočtů byly následně zvalidovány na referenčních datech. Detailní popis těchto přepočtů, jejich validace a zhodnocení je k dispozici v kapitole 6.

Podíl nezaměstnaných osob

V rámci testovacího území České republiky byl proveden přepočet podílu nezaměstnaných osob z obcí do hexagonů o rozloze 5 km². Jako zdrojová data o uchazečích o zaměstnání za obce ČR byla použita data z Veřejné databáze ČSÚ platná ke dni 31. 12. 2019 (ČSÚ, 2021). Podíl nezaměstnaných osob pak byl přepočten dvěma způsoby. Prvním způsobem byl přepočet dosažitelných uchazečů a počtu obyvatel v produktivním věku do hexagonů odděleně a následný výpočet podílu nezaměstnaných osob přímo v hexagonech. Oba dva přepočty trvaly dohromady 2 hodiny a 45 minut. Pro přepočet obou jevů byla použita jako pomocná data vrstva stavebních objektů, váha objektu byla stanovena jako $2 \times \text{počet podlaží} + \text{počet bytů}$. Z důvodu využití stejných pomocných dat a jejich vah pro přepočet obou jevů byl v každé obydlené části jedné zdrojové jednotky uvažován stejný poměr mezi oběma jevy, proto je výsledek v rámci jedné obce konstantní.

Druhým z nich byl přepočet samotného indexu jako intenzivní proměnné s využitím pomocných dat ve formě stavebních objektů z databáze RÚIAN. Tento přepočet trval 1 hodinu 23 minut 28 sekund. Snahou tohoto přepočtu bylo eliminovat vnitřní homogenitu výsledku v rámci jednotlivých obcí. Výsledek tohoto přepočtu byl následně vizualizován jako podíl nezaměstnaných osob v hexagonech ČR v roce 2019 (Obr. 5.4.2).

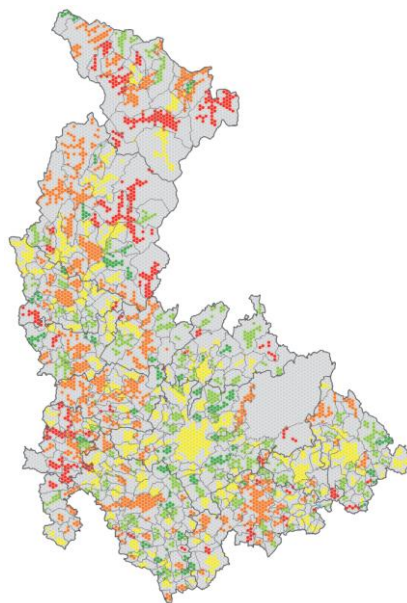


Obr. 5.4.2 Podíl nezaměstnaných osob v hexagonech ČR v roce 2019 – areálová interpolace z obcí

Index stáří

Pro testovací území Olomouckého kraje byla provedena areálová interpolace indexu stáří z obcí do hexagonů o rozloze 0,5 km². Přepočet proběhl jako dva oddělené přepočty počtu obyvatel v předproduktivním a postproduktivním věku z obcí do hexagonů. Jako pomocná data byly využity stavební objekty, váha stavebního objektu byla experimentálně stanovena jako $2 \times \text{počet podlaží} + \text{počet bytů}$. Výsledek v rámci jedné obce je opět konstantní z důvodu použití stejných pomocných dat pro oba dva

disagregované jevy. Výsledkem přepočtů je vizualizace indexu stáří v Olomouckém kraji v hexagonové síti o rozloze 0,5 km² (Obr. 5.4.3).



Obr. 5.4.3 Index stáří v hexagonech Olomouckého kraje v roce 2019 – areálová interpolace z obcí

6 SROVNÁNÍ METOD

Doplňujícím cílem práce bylo s využitím vytvořených nástrojů a dostupných pomocných dat provést přepočítání statistického jevu do nižších územních jednotek a provést jeho validaci s pomocí referenčních dat. Na základě konzultace s vedoucím práce a z důvodu dobré dostupnosti dat pro validaci byl pro tento účel vybrán počet obyvatel s obvyklým pobytem. Počet obyvatel byl v rámci testovacího území přepočítán z obcí do ZSJ a hexagonů v rámci testovacího území Olomouckého kraje. Výsledné přepočty byly poté zvalidovány na referenčních datech pro obě cílové územní jednotky.

6.1 Volba metod a vstupních dat

Za účelem přepočtu počtu obyvatel z území obcí do hexagonů a ZSJ Olomouckého kraje byla využita metoda areálové interpolace. Areálová interpolace má na svém vstupu hodnoty statistického jevu agregované do územních jednotek, z nich je pak proveden přepočítání do odlišných územních jednotek. Protože jev ve sledovaném území není konstantní, je vhodné použít pro disagregaci pomocná data znázorňující rozložení jevu v daném území.

Podle Holta a Lu (2011) je prvním krokem areálové interpolace dasymetrické mapování. Během dasymetrického mapování dochází k přepočtu hodnoty jevu ze zdrojových zón do detailnějších dasymetrických zón. Dasymetrické zóny byly v případě areálové interpolace s využitím vektorových pomocných dat definovány jako průnik zdrojové zóny, cílové zóny a vrstvy pomocných dat. Pro rastrovou areálovou interpolaci jsou pak dasymetrické zóny definovány buňkou rastru pomocných dat. Pro každou dasymetrickou zónu byla metodou dasymetrického mapování vypočítána hodnota sledovaného jevu. Druhým krokem areálové interpolace je poté zpětná agregace hodnot z dasymetrických zón do cílových zón. Postup použitého přepočtu je detailně popsán v kapitole 4.4. Metoda disagregace je pak zvláštním případem areálové interpolace, kdy cílové zóny jsou podmnožinou zdrojových zón a mezi jejich hranicemi nedochází k žádnému průniku (Huang a kol., 2007).

6.1.1 Zdrojová a referenční data

Během agregace statistických dat do územních jednotek dochází ke ztrátě prostorové heterogenity sledovaného jevu v území. Tento jev popsal Openshaw (1983) jako problém měřítka a zařadil ho jako součást problému měnitelné plošné jednotky (MAUP). Čím větší je rozdíl mezi dvěma územními jednotkami, tím větší je efekt problému měřítka a tím obtížnější je poté zpětné modelování prostorové heterogenity pomocnými daty v rámci areálové interpolace. Z důvodu získání co nejpřesnějších výsledků přepočtů byly pro srovnání metod vybrány administrativní jednotky tak, aby byl efekt problému měřítka co nejmenší.

Testování areálové interpolace počtu obyvatel proběhlo v rámci testovacího území Olomouckého kraje. Jako zdrojové zóny byly vybrány obce Olomouckého kraje. Území obcí je nejdětalnější prostorová jednotka, za níž ČSÚ publikuje každoroční data o počtu obyvatel ve Veřejné databázi (ČSÚ, 2021). Lze tedy předpokládat, že právě území obcí bude pro území ČR jedna z nejčastěji využitelných prostorových jednotek pro disagregaci. Jako cílové zóny byly vybrány ZSJ Olomouckého kraje. Některé obce jsou však složeny pouze z jedné ZSJ. Při disagregaci u těchto obcí nedochází k rozpočítání počtu obyvatel do několika dílčích zón, ale pro cílovou ZSJ je převzat počet obyvatel z obce. Výsledek takové disagregace pak bude vykazovat nulovou chybu, protože počet obyvatel v ZSJ je

shodný s počtem obyvatel obce. Ve skutečnosti však tato nulová chyba neznamená správnost zvolené metody, ale absenci potřeby modelování heterogenity jevu v území. Dochází tedy k nechtěnému zkreslení hodnocení přesnosti těmito správnými hodnotami a testovaná metoda se jeví přesnější, než ve skutečnosti je. Z tohoto důvodu byl pro validaci jako druhý typ cílové zóny použit hexagonový grid o ploše buňky 0,5 km². Tato velikost buňky je vhodná pro vizualizaci statistických jevů v měřítku 1 : 500 000. S ohledem na prezentaci výsledků na formátu A4 je tato velikost buňky dostatečně detailní, ale stále dobře rozlišitelná.

Jako referenční data byla použita data „Statistické budovy – vchody k bytům“ poskytnutá vedoucím práce. Data obsahují mimo jiné informace o počtu obyvatel v rámci adresních bodů platný k 26. 3. 2011. Tato detailní bodová vrstva byla poté s využitím nástroje *Aggregate Features To Polygon* agregována do zdrojové zóny a pro validaci přepočtů také do cílových zón.

6.1.2 Pomocná data

V rámci testování přesnosti areálové interpolace bylo použito několik druhů pomocných dat vstupujících do přepočtu. Pomocná data pro přepočet byla vybrána na základě rešerše domácích i zahraničních studií, testování v rámci bakalářské práce (Zapletal, 2019) a na základě jejich dostupnosti pro testovací území.

Data o krajinném pokryvu

Data o krajinném pokryvu jsou jedním z nejčastěji používaných druhů pomocných dat pro areálovou interpolaci, a to z důvodu jejich široké dostupnosti a schopnosti odlišit obydlené území od neobydleného. V rámci práce byly použity datové sady CORINE Land Cover a OpenStreetMap.

CORINE Land Cover (CLC) je evropská databáze o krajinném pokryvu vytvářená službou Copernicus Land Monitoring Service. Její první kompletní verze vyšla v roce 1990 a od roku 200 dochází k její pravidelné aktualizace každých šest let. Minimální mapovací jednotka CLC je stanovena jako území o rozloze 25 ha pro plošné jevy a 100 m pro liniové jevy, minimální mapovací jednotka změn je pak stanovena na 5 ha (Copernicus, 2021). Pro potřeby diplomové práce byla stažena nejnovější vektorová datová sada z roku 2018. Tato data byla vytvořena za 1,5 roku s využitím částečné automatické klasifikace satelitních snímků z družic Sentinel-2 a Landsat-8 (Copernicus, 2021). Při přepočtech byly využity zejména plochy s kódem využití území 111 (Continuous Urban Fabric) a 112 (Discontinuous Urban Fabric).

Vektorová datová sada OpenStreetMap (OSM) byla stažena ze serveru Geofabrik.de dne 31. 10. 2020. S využitím metodiky OSM Landuse (OSMLanduse, 2021) byla tato vektorová vrstva reklasifikována pomocí atributu *landuse* do tříd CLC. Snahou vytvořených dat je poskytnout podrobnější informace o využití území s pomocí detailnější zdrojové datové sady. Při přepočtech pak byly použity zejména plochy s kódem využití území 11 (Urban Fabric).

Hustota silniční sítě

Jako zdroj dat o silniční síti v testovacím území byla využita datová sada OpenStreetMap (OSM) dostupná pod licencí Open Database License (ODbL). Data jsou vytvářena komunitou přispěvatelů OpenStreetMap a jsou dostupná mimo jiné v upravené verzi na serveru Geofabrik.de za celé území ČR ve formátu SHP. Takto upravená datová sada obsahuje 18 vektorových vrstev, pro přepočet byla využita pouze liniová vrstva *gis_osm_roads_free_1* obsahující všechny zakreslené cesty a informace o jejich typu. Na

základě atributu *fclass* byly z dat vybrány pouze dálnice, silnice pro motorová vozidla, silnice první až třetí třídy, obytné ulice a neklasifikované silnice. S využitím nástroje *Aggregate Features to Polygon* byly vybrané prvky agregovány do cílových zón a byla vypočítána délka silnic v km pro každou cílovou zónu. Výsledná délka silnic pak byla použita jako váha pro přepočet.

Zastavěné území

Jako zdroj dat o zastavěném území v rámci Olomouckého kraje byla použita z dat ÚAP Olomouckého kraje vrstva *ft_01_URB_zastavene_uzemi_p* poskytnutá vedoucím práce. Vrstva byla stažena dne 22. dubna 2021 a obsahuje například informace o pořizovateli a správci dat, o zdroji dat, datu změny i názvu obce. Pro potřeby práce byla využita pouze geometrie vrstvy. Po důkladném zkoumání vrstvy bylo zjištěno, že vrstva zastavěného území není kompletní, tři obce Olomouckého kraje vůbec neobsahují informaci o zastavěném území a v několika dalších obcích nejsou zahrnuty všechny zastavěné plochy.

Adresní body a budovy

Adresní body a půdorysy budov jsou jedny z nejdetailnějších dostupných informací o území použitelných jako pomocná data pro areálovou interpolaci. V rámci těchto pomocných dat byly použity dva různé zdroje.

Prvním datovým zdrojem je vrstva „Statistické budovy – vchody k bytům“ vytvořená a spravovaná Českým statistickým úřadem. Tato data byla za účelem diplomové práce poskytnuta vedoucím práce a jsou platná k 1. dubnu 2016. Prostorová lokalizace adresních bodů pochází z databáze RÚIAN nebo z Registru sčítacích obvodů a budov. Součástí této vrstvy jsou mimo jiné i informace o typu a způsobu využití budovy, počtu podlaží, počtu bytů, počtu obyvatel s trvalým a obvyklým pobytem. Jako pomocná data byly v případě této vrstvy využity informace o počtu bytů a počtu podlaží.

Druhým datovým zdrojem o adresních bodech a budovách je databáze RÚIAN. Data byla stažena z databáze RÚIAN 28. března 2021 pro celé území ČR pomocí zásuvného modulu QGIS RUIAN Plugin v 1.0.0. Pro práci byly využity stažené vrstvy *adresnimista* a *stavebniobjekty*. Vrstva *adresnimista* obsahuje adresní body ČR včetně jejich PSČ, domovního a orientačního čísla. Nedostatkem této vrstvy je absence dalších informací o adresním bodu, pro přepočet bylo tedy možné použít pouze geometrii prvku. Vrstva *stavebniobjekty* obsahuje informace o typu stavebního objektu, způsobu využití, o počtu bytů a počtu podlaží. Pro přepočet pak byly využity zejména o atributy počet bytů a počet podlaží. Jak již bylo zjištěno v bakalářské práci (Zapletal, 2019), data RÚIAN nepokrývají kompletně celé zájmové území a chybí zejména v místech, kde nebyla ještě digitalizována katastrální mapa. V těchto místech se vyskytuje pouze vrstva adresních bodů, úplně chybí vrstva parcel nebo stavebních objektů. Vzhledem k tomu, že se adresní body vždy nachází uvnitř budovy, je vhodné tyto dvě vrstvy zkombinovat, čímž dojde k odstranění nedostatků obou z nich (Zapletal, 2019).

6.2 Přepočet dat

Dalším krokem srovnání jednotlivých technik areálové interpolace a disagregace byl samotný přepočet dat. Přepočet dat probíhal za použití vytvořených nástrojů a jeho výsledkem je celkem 64 atributů obsahujících přepočtený počet obyvatel s obvyklým pobytem. Přepočet probíhal v rámci studovaného území Olomouckého kraje na dvou úrovních územních jednotek – z obcí do ZSJ a z obcí do hexagonů. Pro obě úrovně pak byl proveden přepočet vektorovou i rastrovou variantou nástroje *Disaggregate Polygon To*

Polygon. Pro každou úroveň a variantu nástroje bylo provedeno 16 přepočtů s využitím pomocných dat uvedených v kapitole 6.1.

První testovanou variantou areálové interpolace bylo areálové vážení. Tato technika nevyžaduje žádná pomocná data, vahou dasymetrické zóny je její rozloha. Pro obě varianty přepočtů nebyla do nástrojů dodána žádná pomocná data. Ve vektorovém nástroji byl proveden polygonový překryv pouze nad zdrojovými a cílovými zónami. V rámci vzniklých dasymetrických zón vznikl váhový atribut s hodnotou 1. V rastrovém nástroji byl proveden polygonový překryv zdrojových a cílových zón. Výsledek byl poté převeden na rastr s hodnotami atributu ID. Všechny nenulové hodnoty poté byly pomocí mapové algebry nahrazeny hodnotou 1, čímž vznikl konstantní váhový rastr pro celé zájmové území. Následně byl proveden pro oba atributy přepočet, jehož výsledkem je atribut *POCOB_AW*.

Dalších pět variant metody areálové interpolace zahrnovalo bodová pomocná data v podobě adresních bodů. Nejprve byly otestovány adresní body z databáze RÚIAN, na něž byla aplikována binární metoda. Nevýhodou binární metody je, že všechny adresní body (nebo jiné objekty) v rámci zdrojové zóny mají stejnou váhu. Výsledkem tohoto přepočtu je atribut *POCOB_AdrBody_RUIAN_BIN*. Dále byly testovány adresní body ČSÚ poskytnuté vedoucím práce. Pro ně byl proveden nejprve přepočet binární metodou, výsledným atributem je *POCOB_AdrBody_CSU_BIN*. Na tyto adresní body byla poté aplikována metoda spojitého vážení, kdy byly adresní body váženy jejich počtem podlaží (*POCOB_AdrBody_CSU_PocPodl*) a počtem bytů (*POCOB_AdrBody_CSU_PocByt*). Poslední testovanou variantou spojitého vážení pro adresní body ČSÚ byl součet dvojnásobku počtu bytů a počtu podlaží (*POCOB_AdrBody_CSU_PodlByt*). Vektorová varianta nástroje umožňuje pracovat s bodovou vrstvou pomocných dat přímo, pro použití rastrové verze bylo nejprve nutné převést všechny atributy použité jako pomocná data do rastru pomocí nástroje *Create Ancillary Layer*.

Následující variantou přepočtu byla binární metoda s využitím zastavěného území. Nevýhodou vrstvy zastavěného území je zahrnutí všech zastavěných neobytných částí obce, např. průmyslové zóny. Vektorová varianta nástroje umožňuje s touto vrstvou pracovat přímo, pro použití rastrové varianty bylo potřeba nejprve převést zastavěné území na rastr s konstantní hodnotou buňky. Výsledkem přepočtů byl atribut *POCOB_ZastUz_UAP_BIN*.

Dále byly pro přepočet počtu obyvatel testovány stavební objekty z databáze RÚIAN. Pro ně byl proveden přepočet binární metodou, kdy všechny stavební objekty mají stejnou váhu (*POCOB_budovy_RUIAN_BIN*). Dále byla využita metoda spojitého vážení, kdy vahou jednotlivých stavebních objektů byl počet podlaží (*POCOB_budovy_RUIAN_PocPodl*), počet bytů (*POCOB_budovy_RUIAN_PocByt*) a součet dvojnásobku počtu bytů a počtu podlaží (*POCOB_budovy_RUIAN_PodlByt*). Z důvodu chybějících dat RÚIAN v některých částech studovaného území byla využita kombinace stavebních objektů a zastavěného území. Pro oblasti pokryté daty RÚIAN bylo zachováno využití stavebních objektů, pro zbylé území byla do vrstvy pomocných dat přidána ještě zastavěná území. Pro přepočet byla všem takto přidaným zastavěným územím stanovena hodnota počtu bytů na 1. Z této vrstvy byl vypočítán součet dvojnásobku počtu bytů a počtu podlaží a výsledek byl uložen do atributu *POCOB_PodlByt_ZastUz_Fix*. Z důvodu velmi dobrého výsledku přepočtu adresních bodů ČSÚ vážených počtem bytů byly ještě stavební objekty RÚIAN převedeny na body a jejich vahou byl stanoven počet bytů. Výsledek přepočtu je uložen v atributu *POCOB_PocBytu_RUIAN_bod*.

Závěrem byla testována binární metoda na data o krajinném pokryvu z CLC (*POCOB_CLC_UrbanFabric*) a data OSM (*POCOB_CLC_ENVIG_UrbanFabric*). Metoda spojitého vážení byla otestována na hustotě silniční sítě (*POCOB_OSM_roaddensity_**).

6.3 Validace

Pro přepočítání počtu obyvatel s obvyklým pobytem z obcí do hexagonů a ZSJ Olomouckého kraje bylo otestováno celkem 16 variant pomocných dat. Na základě nich vzniklo 16 atributů s přepočtenými hodnotami:

- *POCOB_AW* – bez využití pomocných dat (areálové vážení),
- *POCOB_AdrBody_RUIAN_BIN* – adresní body z RÚIAN, binární metoda,
- *POCOB_AdrBody_CSU_BIN* – adresní body z ČSÚ, binární metoda,
- *POCOB_AdrBody_CSU_PocPodl* – počet podlaží adresních bodů ČSÚ,
- *POCOB_AdrBody_CSU_PocByt* – počet bytů adresních bodů ČSÚ,
- *POCOB_AdrBody_CSU_PodlByt* – počet podlaží + 2* počet bytů adresních bodů ČSÚ,
- *POCOB_ZastUz_UAP_BIN* – zastavěné území z ÚAP, binární metoda,
- *POCOB_budovy_RUIAN_BIN* – stavební objekty RÚIAN, binární metoda,
- *POCOB_budovy_RUIAN_PocPodl* – počet podlaží stavebních objektů RÚIAN,
- *POCOB_budovy_RUIAN_PocByt* – počet bytů stavebních objektů RÚIAN,
- *POCOB_budovy_RUIAN_PodlByt* – počet podlaží + 2 * počet bytů stavebních objektů RÚIAN,
- *POCOB_PodlByt_ZastUz_Fix* – počet podlaží + 2 * počet bytů stavebních objektů RÚIAN, doplněno o zastavěné území,
- *POCOB_PocBytu_RUIAN_bod* – počet bytů bodových stavebních objektů RÚIAN
- *POCOB_CLC_UrbanFabric* – obydlené oblasti CLC, binární metoda,
- *POCOB_CLC_ENVIG_UrbanFabric* – obydlené oblasti OSM, binární metoda,
- *POCOB_OSM_roaddensity_hexa/ZSJ* – délka silniční sítě v cílové zóně.

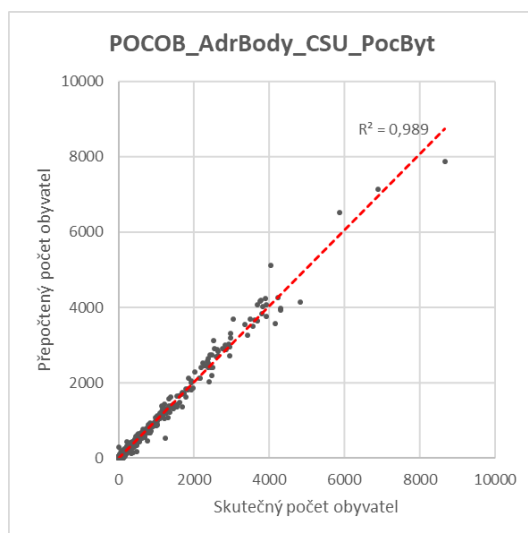
Výsledné atributy byly poté v rámci obou cílových zón pro oba dva nástroje validovány na referenčních datech a výsledná zjištění byla interpretována. Pro validaci přepočtů byly použity dva různé způsoby. Prvním způsobem validace je zjištění celkové velikosti chyby výpočtem Relative Total Absolute Error (RTAE). Tuto metodu využil ve své práci Šimbera (2020) pro hodnocení přesnosti disagregačních technik. RTAE (Obr. 6.3.1) se vypočítá jako podíl sumy absolutních chyb ku celkovému počtu obyvatel v rámci zvoleného území, kde m_i je skutečný počet obyvatel v územní jednotce a m_i' je odvozený (přepočtený) počet obyvatel v územní jednotce.

$$RTAE = \frac{\sum_i |m_i' - m_i|}{\sum_i m_i}$$

Obr. 6.3.1 Výpočet RTAE

RTAE nabývá hodnot od 0 % do 200 %, kde 0 % značí shodnost odhadnutého a skutečného počtu obyvatel ve všech cílových zónách a 200 % značí, že veškerý počet obyvatel byl přiřazen do neosídlených lokalit (Šimbera, 2020). Hodnota RTAE byla vypočtena v prostředí Microsoft Excel pro všech 16 atributů přepočtených pomocí dvou nástrojů v obou cílových zónách, celkem tedy 64 hodnot. Výsledné hodnoty RTAE včetně jejich výpočtu jsou dostupné jako tabulky Microsoft Excel jako součást Přílohy 7 (DVD).

Výsledky areálových interpolací byly rovněž zhodnoceny za pomoci korelogramů. Korelogramy byly vytvořeny v prostředí Microsoft Excel jako XY bodový graf, kdy se na ose x korelogramu nachází skutečný počet obyvatel v jednotlivých obcích a na ose y přepočtený počet obyvatel. Výsledný korelogram byl ještě proložen lineární spojnicí trendu a byla do něj přidána hodnota spolehlivosti R^2 (viz Obr. 6.3.2). Tímto způsobem byly vytvořeny korelogramy pro všech 64 přepočtených atributů. Výsledné korelogramy včetně hodnoty spolehlivosti R^2 jsou součástí Přílohy 7 (DVD) jako tabulky Microsoft Excel a ve formátu PDF.



Obr. 6.3.2 Korelogram areálové interpolace do hexagonů s využitím počtu bytů adresních bodů

Vypočtené hodnoty R^2 a RTAE u všech testovaných metod pro všechny kombinace použitých nástrojů a cílových zón byly převedeny do tabulky (Obr. 6.3.3), kde první čtyři sloupce značí hodnotu RTAE a další čtyři sloupce hodnotu R^2 . Pro každou metodu bylo vypočítáno její pořadí mezi ostatními metodami v rámci sledované statistiky. U výpočtu RTAE platí, že čím nižší hodnota, tím lepší výsledek přepočtu. U hodnoty spolehlivosti naopak platí, že vyšší hodnota R^2 značí lepší výsledek. Vyšší spolehlivost metody značí větší pravděpodobnost, že přepočet počtu obyvatel touto metodou bude vycházet podobně i pro jiná než testovací území. U každé metody pak byl vypočítán součet jejich pořadí v rámci všech osmi hodnotících technik. Z výsledku součtu ve sloupci BODY na Obr. 6.3.3 pak bylo vytvořeno pořadí jednotlivých metod a metody byly pro testovací území Olomouckého kraje seřazeny od nejvhodnější po nejméně vhodnou.

Způsob disagregace	Hexa V	Hexa R	ZSJ V	ZSJ R	Hexa_V_R ²	Hexa_R_R ²	ZSJ_V_R ²	ZSJ_R_R ²	BODY	POŘADÍ
POCOB_AdrBody_CSU_PocByt	9,71%	9,71%	6,72%	6,68%	0,9897	0,989	0,9884	0,9887	9	1
POCOB_AdrBody_CSU_PodlByt	13,69%	9,72%	9,94%	6,49%	0,9768	0,9883	0,9735	0,9884	16	2
POCOB_PocBytu_RUIAN_bod	14,62%	14,54%	9,16%	9,14%	0,9718	0,9732	0,9664	0,9651	23	3
POCOB_AdrBody_CSU_PocPodl	21,93%	21,91%	16,94%	16,92%	0,9277	0,925	0,9108	0,9127	36	4
POCOB_PodlByt_ZastUz_Fix	23,48%	25,21%	17,03%	18,41%	0,9371	0,9312	0,9246	0,9193	36	4
POCOB_budovy_RUIAN_PodlByt	26,19%	35,53%	28,56%	28,61%	0,9246	0,9223	0,9066	0,9011	54	6
POCOB_budovy_RUIAN_PocPodl	35,49%	30,89%	22,44%	22,31%	0,7738	0,7637	0,7169	0,7162	61	7
POCOB_budovy_RUIAN_PocByt	30,96%	60,01%	47,21%	47,25%	0,9095	0,9104	0,8839	0,8855	74	8
POCOB_AdrBody_CSU_BIN	43,78%	43,73%	37,92%	37,97%	0,6768	0,6568	0,5932	0,5977	75	9
POCOB_AdrBody_RUIAN_BIN	44,39%	44,41%	38,27%	38,32%	0,6732	0,6527	0,5891	0,5928	83	10
POCOB_CLC_UrbanFabric	53,73%	53,74%	42,04%	42,05%	0,7082	0,6823	0,5837	0,6171	84	11
POCOB_ZastUz_UAP_BIN	66,28%	27,86%	18,91%	20,25%	0,5317	0,5148	0,3293	0,3333	90	12
POCOB_CLC_ENVIG_UrbanFabric	51,24%	51,18%	44,27%	44,40%	0,6419	0,6241	0,5258	0,5346	94	13
POCOB_budovy_RUIAN_BIN	59,97%	66,15%	54,05%	53,67%	0,5471	0,525	0,3803	0,3824	109	14
POCOB_OSM_roaddensity_hexa	87,57%	87,57%	84,69%	84,69%	0,5698	0,5698	0,0543	0,0543	116	15
POCOB_AW	145,28%	145,29%	88,87%	88,87%	0,125	0,1249	0,0383	0,0383	128	16

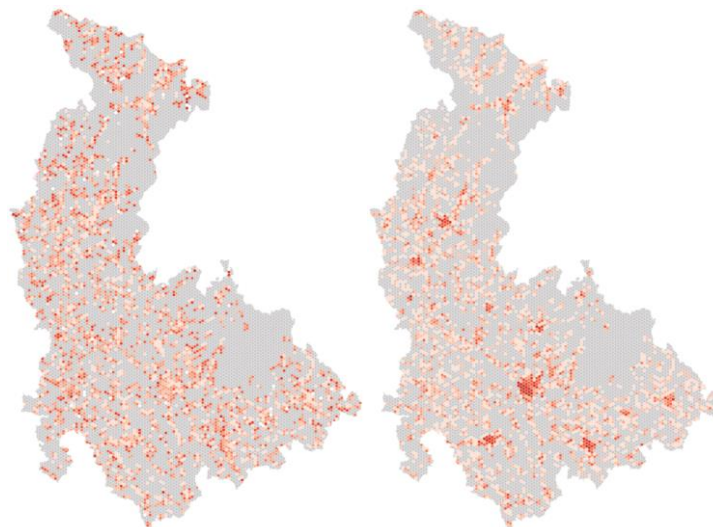
Obr. 6.3.3 Zhodnocení jednotlivých metod disagregace

Z výsledků je patrné, že v první polovině tabulky se vyskytují přepočty využívající spojitě vážení, v dolní polovině tabulky se nachází binární metody. Spojitě vážení na rozdíl od binární metody umožňuje pro jeden typ pomocných dat stanovit různé váhy v rámci sledovaného území. V případě vhodného použití pomocných dat a jejich vah pak dokáže lépe vysvětlit prostorovou variabilitu přepočítávaného jevu než klasifikované vážení nebo binární metoda.

Dále lze z výsledné tabulky zjistit, že ve velké většině případů vychází RTAE lépe u ZSJ než u hexagonů. Tento výsledek je dán především tím, že některé obce obsahují pouze jednu ZSJ, u těchto ZSJ je přepočtený počet obyvatel s využitím všech metod vždy shodný se skutečným stavem. Hexagonová síť naopak může hranice obcí různě křížit, proto přepočtený počet obyvatel v hexagonu nikdy nebude shodný se skutečným počtem obyvatel ve zdrojové obci. Použití hexagonů a následná validace v nich více prověřuje schopnost metody modelovat prostorovou heterogenitu sledovaného jevu.

V rámci testování metod bylo zjištěno, že nejlépe ve všech sledovaných indikátorech vychází adresní body, případně bodová vrstva stavebních objektů. Z porovnání budov RÚIAN reprezentovaných bodovou a polygonovou vrstvou lze usuzovat, že pro zvolené testovací území dokáží bodově lokalizované budovy lépe vystihnout prostorovou heterogenitu počtu obyvatel než polygonové budovy a že využití rozlohy budov jako součást jejich váhy (viz kapitola 4.4.1) výsledkům přepočtu spíše škodí.

Závěrečnou částí validace přepočtů byla také vizualizace chyb přepočtů pro vybrané atributy. Prvním možným způsobem vizualizace chyby je zobrazení procentuální hodnoty chyby v cílových jednotkách. Způsob výpočtu procentuální chyby v území je stejný jako výpočet RTAE pro jednu cílovou zónu. Absolutní hodnota rozdílu přepočteného a skutečného počtu obyvatel je vydělena skutečným počtem obyvatel a vynásobena 100. Čím méně je dané území osídleno, tím výraznější však sebemenší odchylka bude. Druhým způsobem je zobrazení absolutní chyby počtu obyvatel s využitím metody kartogramu. Absolutní hodnota rozdílu přepočtené a skutečné hodnoty jevu je přepočtena na plochu. V hustě zastavěných oblastech však bude i při malé RTAE nabývat tento způsob výpočtu chyby větších hodnot než v řídké zastavěných oblastech. Srovnání obou metod vizualizace chyby je znázorněno na Obr. 6.3.4. Hodnocení chyb pomocí absolutních odchylek, RTAE apod. lze využít jen v případě dostupnosti dat za cílové zóny. Pokud data k dispozici nejsou, měla by být vizualizována nejistota jiným způsobem (Mennis, 2009).



Obr. 6.3.4 Srovnání vizualizace chyby pomocí RTAE a absolutní chyby v kartogramu

7 VÝSLEDKY

Cílem diplomové práce bylo vytvořit pro prostředí ArcGIS Pro nástroj pro automatizaci agregačních a disagregačních metod umožňující pracovat jak s územně správními jednotkami, tak i s gridovými sítěmi.

Samotné tvorbě nástrojů předcházela podrobná rešerše agregačních a disagregačních metod. Metoda agregace (bottom-up) je procesně jednodušší metoda. Používá se v případě dostupnosti bodově lokalizovaných nebo detailnějších plošných dat. Vizualizace agregovaných dat do územních jednotek však může uživateli způsobit problémy při interpretaci mapy. Metoda disagregace (top-down) je procesně náročnější metoda, při níž dochází k odhadu chybějících dat z dat agregovaných do vyšších územních jednotek. Při tomto procesu dochází k modelování prostorové heterogenity sledovaného jevu, za tímto účelem je vhodné použít doplňkové vrstvy pomocných dat. Všechna specifika a úskalí obou metod jsou detailně popsána v kapitole 3.

V rámci diplomové práce vznikly tři hlavní výsledky – vytvořená sada nástrojů včetně dvou manuálů pro jejich použití, data přepočtená pomocí vytvořených nástrojů včetně hodnocení přesnosti vybraných přepočtů a vizualizace přepočtených dat jako datových náhledů, případně mapových výstupů. Výsledky práce jsou prezentovány na webových stránkách diplomové práce a na informačním posteru.

7.1 Nástroje pro agregaci a disagregaci dat

Hlavním výsledkem práce je sada vytvořených skriptů pro automatizaci agregačních a disagregačních metod pro prostředí ArcGIS Pro. Na základě rešerše a získaných poznatků v rámci bakalářské práce (Zapletal, 2019) byl vytvořen princip výpočtu umožňující provádět areálovou interpolaci ze zdrojových zón do cílových zón. Princip areálové interpolace (a tedy i disagregace) je založen na dvou krocích. Prvním krokem je dasymetrické mapování, kdy dochází k přepočtu hodnoty jevu ze zdrojových zón do dasymetrických zón. Za tímto účelem se používají váhy pro jednotlivé dasymetrické zóny definované vrstvou pomocných dat a rozlohou dasymetrické zóny. Druhým krokem je pak zpětná agregace z dasymetrických zón do cílových zón.

Vzhledem k velkým odlišnostem metody agregace a disagregace, zejména v logice přepočtu a vstupních parametrech, nemohly být obě metody zakomponovány do jednoho nástroje. Výsledkem práce je tedy šest skriptů v jazyce Python. Dva z nich jsou pro agregaci, dva pro disagregaci a dva pro tvorbu vrstvy pomocných dat. Nástroj *Aggregate Features To Polygon* je založen na skriptu Summarize Within, zbytek skriptů však využívá metody polygonového překryvu nebo jazyk mapové algebry. Implementace logiky areálové interpolace do vytvářených nástrojů umožňuje data agregovat či disagregovat nejen do skladebných, ale i do naprosto nesouvisejících územních jednotek, například do gridových sítí.

V prostředí ArcGIS Pro byl vytvořen toolbox *Dasymetric.tbx*, do něj bylo vytvořeno šest nových skriptů. Skriptům byly nastaveny parametry, jejich datové typy a závislosti na ostatních parametrech. K vytvořeným skriptům byly poté importovány příslušné programové kódy. Ty byly napsány v programu Atom s využitím balíčku *arcpy* pro Python. Samotný skript kombinuje již existující nástroje v ArcGIS Pro s vytvořenou logikou posloupnosti přepočtu. Nástroje založené na principu areálové interpolace umožňují pracovat s extenzivním i intenzivním typem proměnné, pro každý typ proměnné je nastavena příslušná logika přepočtu. Pomocná data pro přepočet jsou nastavena jako volitelná, bez jejich využití fungují nástroje na principu areálového vážení. Vytvořeným nástrojům byla následně vyplněna metadata a byl sestaven manuál pro jejich použití,

včetně vzorového příkladu využití. Manuál byl sestaven v českém i anglickém jazyce a je k dispozici jako součást webové stránky práce a v rámci Přílohy 7 (DVD) je uložen ve formátu PDF.

7.2 Agregovaná a disagregovaná data

Dalším výsledkem práce jsou přepočtená data pomocí vytvořených nástrojů. Dílčím cílem práce bylo otestovat vytvořené skripty na několika datových sadách pro testovací území Olomouckého kraje a České republiky. Jako cílové zóny přepočtů byly použity administrativní jednotky i gridové sítě. Vhodná velikost gridové sítě byla vybrána na základě Zajícové (2017) a Zapletala (2019) za účelem následné prezentace přepočtených dat. Pro přepočet bylo vybráno pět statistických ukazatelů – index stáří, počet obyvatel, index kriminality, podíl nezaměstnaných osob a průměrný věk. Zvolené indikátory byly s využitím vytvořených nástrojů přepočteny do cílových zón. Samotné testování vytvořených nástrojů bylo rozděleno na dvě části.

První částí testování nástrojů bylo testování všech nástrojů na zvolených statistických ukazatelích. Tato část testování sloužila zejména k otestování funkcionality nástrojů na různých vstupních datových sadách a jejich schopnosti práce s různými typy územních jednotek. Během této části bylo nalezeno a opraveno nejvíce chyb v logice skriptů i programovém kódu, případně byla doplněna a otestována další funkcionality. V rámci této části testování bylo pro každý nástroj vybráno několik statistických ukazatelů, ty byly následně přepočteny do cílových jednotek. V rámci tohoto testování byla rovněž sledována i doba přepočtu jednotlivých ukazatelů. Kompletní postup i výsledky testování jsou popsány v kapitole 5.

Druhou částí testování nástrojů bylo srovnání různých přístupů pro disagregaci počtu obyvatel. Tato část byla z důvodu větší časové náročnosti a složitosti metody disagregace testována pouze v rámci Olomouckého kraje. Jako zdrojové zóny byly zvoleny obce Olomouckého kraje, jako cílové zóny byly využity ZSJ a hexagonová síť o rozloze buňky 0,5 km². Do zdrojových zón byl agregován údaj o počtu obyvatel s obvyklým pobytem z bodové vrstvy adresních bodů od ČSÚ. Pro validaci přepočtů byl tento údaj agregován i do obou cílových zón. V rámci této části testování bylo využito celkem sedm druhů pomocných dat. Z nich bylo sestaveno celkem 16 srovnávaných variant. Každá varianta pomocných dat poté byla testována pro území ZSJ i hexagonovou síť pomocí vektorového i rastrového nástroje. Celkem vzniklo 64 přepočtených atributů. Pro každý atribut byla vypočítána chyba přepočtu pomocí metody RTAE a dále byly vytvořeny korelogramy s vyznačenou hodnotou spolehlivosti. S využitím RTAE a hodnoty spolehlivosti byly pak všechny metody srovnány. Jako nejvhodnější pomocná data pro přepočet počtu obyvatel v rámci testovacího území byly stanoveny adresní body nebo stavební objekty s využitím počtu bytů nebo počtu podlaží jako jejich váhy. Bylo také zjištěno, že spojitě vážení v případě využití vhodné váhy pomocných dat výrazně zlepšuje přesnost a míru spolehlivosti disagregace oproti binární metodě. Dalším zjištěním bylo, že využití libovolných pomocných dat souvisejících s jevem překonává areálové vážení. Na základě srovnání disagregace dat do gridové sítě a ZSJ lze tvrdit, že obce obsahující jednu ZSJ výrazně zlepšují výsledné hodnoty RTAE. Toto „zlepšení“ je však dáno zkrácením hodnoty RTAE z důvodu shodného počtu obyvatel ve zdrojové obci a cílové ZSJ. Administrativní jednotky se tedy jeví jako méně vhodné pro validaci disagregace než gridové sítě, pokud hrozí, že alespoň jedna cílová zóna bude shodná s její zdrojovou zónou. Kompletní postup srovnání metod byl popsán v kapitole 6.

7.3 Vizualizace

Dalším výsledkem práce byla vizualizace vybraných přepočtů jako datových náhledů a mapových výstupů. V rámci mapového projektu ArcGIS Pro dostupného v Příloze 7 (DVD) vzniklo v průběhu práce celkem 10 samostatných mapových oken podle řešených témat. Součástí každého mapového okna je několik vizualizací přepočtených témat. Každému přepočítávanému jevu byly nastaveny barevné stupnice tak, aby pro stejný jev byly použity stejné barvy ve všech mapových oknech. Sekvenční barevné stupnice byly vytvořeny úpravou základního nastavení stupnic v softwaru ArcGIS Pro, kdy každé základní stupnici byla odebrána nejsvětější a nejtmaší barva tak, aby barvy výsledných vizualizací nebyly příliš výrazné nebo naopak příliš nevýrazné. Místa bez výskytu jevu pak byla jednotně vizualizována šedou barvou. Tímto způsobem vzniklo celkem 26 vizualizací přepočtených jevů, výsledné vizualizace jsou součástí mapového projektu dostupného v Příloze 7 (DVD).

Šest vybraných přepočtených témat bylo dotaženo do finální mapové vizualizace splňující kartografická pravidla a konvence. Tato témata byla doplněna o titul, měřítko, legendu a tiráž. V tiráži se nachází informace o zdroji dat, souřadnicovém systému, způsobu tvorby přepočtených dat a případně i velikosti buňky gridu. Pro lepší orientaci uživatele v mapě byly rovněž doplněny administrativní hranice. Pro mapy vytvořené v rámci celé ČR byla přidána rovněž krajská města, pro mapy Olomouckého kraje byly doplněny názvy obcí s rozšířenou působností. Tyto výsledné vizualizace byly vtištěny a vevázány do práce jako Příloha 1 až 6.

8 DISKUZE

Cílem diplomové práce bylo vytvořit nástroj do prostředí ArcGIS Pro umožňující automatizaci agregačních a disagregačních metod a vytvořený nástroj poté otestovat na několika datových sadách různého prostorového rozsahu.

V rámci studia agregačních a disagregačních metod bylo zjištěno, že obě analytické metody mají několik slabých míst. Metoda agregace je široce používaná a standardizovaná metoda zpracování dat. Během agregace dat však dochází ke ztrátě polohové přesnosti detailnějších dat a generalizaci, výsledek agregace je tedy výrazně ovlivněn zvolenými cílovými územními jednotkami. Tento problém popisuje Openshaw (1983) jako Modifiable Areal Unit Problem. Čím větší cílové jednotky pro agregaci, tím více se stírá prostorová heterogenita jevu. V případě použití odlišných hranic územních jednotek v rámci podobného měřítka bude odlišný i výsledek agregace. Lze tedy tvrdit, že výsledné analýzy nad agregovanými daty jsou platné pouze pro daný způsob agregace dat. Naproti tomu metoda disagregace je kvůli své složitosti stále zkoumána a existuje několik různých přístupů, z nichž ani jeden není standardizovaný. Metoda disagregace se nejčastěji opírá o použití vrstvy pomocných dat nebo iterační metody. Pomocná data slouží k vysvětlení heterogenity jevu v území, měla by tedy co nejvíce korespondovat s disagregovaným jevem. Nejčastěji zkoumaným ukazatelem je počet obyvatel, disagregace ostatních ukazatelů je limitována nedostatkem vhodných pomocných dat v dostatečné míře podrobnosti.

V rámci rešerše byla zkoumána i kritika prezentace dat agregovaných do administrativních jednotek. Data bývají nejčastěji vizualizována areálovou metodou nebo metodou kartogramu za administrativní jednotky, do nichž byla agregována. Největším problémem takto agregovaných dat je ekologický klam, kdy stejná hodnota jevu v rámci celého území evokuje konstantnost daného jevu v území. Alternativou prezentace jevu za administrativní jednotky jsou dasymetrické mapy. Dasymetrické mapování z hodnot agregovaných do obcí však podle Holta a Lu (2011) neústí ve skutečné dasymetrické mapy, protože se jedná pouze o odhad hodnoty jevu v dílčím územním celku z agregované hodnoty. Dasymetrické mapy naproti tomu znázorňují skutečnou distribuci jevu a jejich tvorba je velmi složitá. Druhou alternativou prezentace dat je znázornění jevu agregovaného do pravidelného gridu. Buňky gridu jsou z důvodu jejich stejné velikosti vzájemně srovnatelné a v případě použití dostatečně malé velikosti buňky dokáží lépe zobrazovat heterogenitu jevu v území. Gridová síť se ale rovněž řadí mezi měnitelné plošné jednotky a vztahují se na ni všechny části MAUP.

Při testování disagregačních metod na území celé České republiky byly jako pomocná data použity stavební objekty z dat RÚIAN. Tato vrstva musela být pro celé území České republiky stažena nástrojem QGIS RUIAN Plugin. Z důvodu problémů se stabilitou nástroje musela být data stahována maximálně po krajích, kdy stahování dat za kraje trvalo v řádech desítek minut až hodin. Některá stahování navíc byla přerušena pádem nástroje a musela být zopakována. Je otázkou, na kolik pomalý a méně stabilní internet ovlivňuje stabilitu stahování tímto nástrojem, rychlost internetu však ovlivnila výslednou dobu stahování. Vrstvy stavebních objektů za kraje pak byly nástrojem Merge spojeny v prostředí ArcGIS Pro do jedné vrstvy.

Během výsledného srovnání technik disagregace počtu obyvatel Olomouckého kraje bylo srovnáno několik různých druhů pomocných dat. Z výsledků je patrné, že detailnější pomocná data přináší přesnější výsledky přepočtů. V případě využití spojitě váhy pomocných dat dochází k ještě výraznějším zpřesněním přepočtů, nejpřesnější metodou v sedmi z osmi sledovaných kritérií bylo využití adresních bodů s vahou počet bytů.

Výsledná validace v rámci této práce je však platná pouze v rámci zvoleného území a použité agregace dat. Pro získání objektivnějšího zhodnocení jednotlivých technik přepočtů by bylo nutné tyto přepočty provést na více zájmových území a pro několik různých způsobů agregace dat.

Největším problémem validace disagregačních technik je nedostupnost detailních referenčních dat. Pro validaci disagregace počtu obyvatel je jako nejnovější vrstva pro validaci k dispozici bodová vrstva „Statistické budovy – vchody k bytům“ poskytovaná ČSÚ. Geometrická část a většina atributů je sice platná k 1. dubnu 2016, počet obyvatel s trvalým a obvyklým pobytem je však převzat ze SLDB 2011 a je tedy platný k 26. březnu 2011. Z důvodu neexistence novější detailní datové sady není možné provádět validace disagregačních technik k novějším datům pro cílové zóny detailnější, než pro území obcí. V případě použití dat z databáze RÚIAN jako vrstvy pomocných dat je nutné rovněž počítat s jejich nedostatky. Stažená datová sada nepokrývá kompletně celé území České republiky a v některých lokalitách tak mohou chybět informace o parcelách nebo o stavebních objektech. Z tohoto důvodu je vhodné při použití dat RÚIAN v přepočtech využít bodovou vrstvu adresních bodů jako geometrii a přiřadit jí informace o budově z vrstvy stavebních objektů.

Metoda disagregace je nejlépe aplikovatelná právě na počet obyvatel z důvodu existence velkého množství dat, u nichž byla prokázána souvislost s rozmístěním obyvatelstva v území. Mezi taková data lze zařadit například data o krajinném pokryvu, nepropustnosti půdy, intenzitě nočního osvětlení, zastavěném území, stavebních objektech nebo adresních bodech. V případě potřeby disagregace dalších ukazatelů, jako například podílu nezaměstnaných osob nebo indexu vzdělanosti, lze sice předpokládat souvislost některých dalších ukazatelů se sledovanými jevy, taková souvislost však zatím nebyla ověřena. Největší překážkou v testování disagregačních technik pro další statistické ukazatele je neexistence možných souvisejících datových sad na dostatečně detailní úrovni. Dalším problémem je pak absence detailních referenčních dat, na nichž by bylo možné testované přepočty validovat.

Použitý detail pomocných dat ovlivňuje kromě přesnosti přepočtů také rychlost tvorby pomocných vrstev a v případě vektorové disagregace pak také rychlost polygonového překryvu. Doba trvání disagregačních přepočtů se odvíjela od složitosti vrstvy pomocných dat a rozsahu území. Disagregace počtu obyvatel z obcí do hexagonů pro celou ČR s využitím stavebních objektů trvala 2 hodiny a 25 minut. Přepočet 32 atributů rastrovým nástrojem v rámci Olomouckého kraje zabral dohromady 1 hodinu a 37 minut, tedy průměrně tři minuty na jeden atribut. Vektorová verze byla naopak pomalejší, kdy doba přepočtu jednoho atributu trvala průměrně osm minut. K době trvání rastrové disagregace je nutné připočítat ještě tvorbu vrstvy pomocných dat, ta však trvala průměrně jednu minutu. Celkový čas přepočtu v rámci Olomouckého kraje rastrovou verzí nástroje byl oproti vektorovému nástroji poloviční. Doba přepočtů však také mohla být ovlivněna různým zatížením počítače v průběhu času.

Při testování přesnosti jednotlivých technik disagregace bylo zjištěno, že při hodnocení metody s využitím RTAE dochází ke zkreslení hodnocení přepočtu v obcích s jednou ZSJ. Přepočtené hodnoty u těchto obcí jsou automaticky správné, což snižuje hodnotu RTAE dané metody. Pro získání lepší představy o velikosti chyby je vhodnější použít takové cílové územní jednotky, u nichž nehrozí, že budou shodné se zdrojovými. Validace přepočtů v gridové síti může zpřesnit skutečnou velikost chyby.

I přesto, že je snaha dosáhnout co nejpřesnějších přepočtů sledovaných jevů, bude výstupem disagregace vždy jen odhad lišící se od reality. V případě nedostupných dat v cílových zónách však nelze srovnat přepočtenou hodnotu jevu se skutečným stavem.

Nejistotu tohoto odhadu je však rovněž nutné vizualizovat. Nadstavbou této práce tak může být například vylepšení vytvořených nástrojů o výpočet předpokládané chyby přepočtů pro cílové územní jednotky.

V práci byly splněny všechny dílčí cíle a výsledky mohou sloužit jako doporučení při agregaci nebo disagregaci statistických dat. Výsledné skripty jsou připraveny pro praktické nasazení při řešení agregačních a disagregačních úloh. Nedostatky a návrhy zmíněné v diskuzi mohou být předmětem dalšího studia.

9 ZÁVĚR

Hlavním cílem diplomové práce bylo vytvořit pro prostředí ArcGIS pro nástroj pro automatizaci agregačních a disagregačních metod tak, aby byl schopen pracovat s územně správními jednotkami i gridovými sítěmi. Za tímto účelem byly využity poznatky získané během rešerše a v rámci bakalářské práce. Dalším úkolem bylo poté otestovat funkcionalitu nástroje pro přepočítání několika datových sad různé úrovně podrobnosti a vybrané výsledky testování patřičně zhodnotit a zvizualizovat.

Během teoretické části práce byly nastudovány metody agregace a disagregace, jejich praktické řešení a využití v České republice i v zahraničí. Součástí bylo rovněž zjištění výhod a nevýhod obou analytických metod. Pro metodu disagregace byl detailněji nastudován teoretický princip přepočtu. Dílčím cílem rešerše bylo vydefinovat rozdíl mezi pojmy areálová interpolace, dasymetrické mapování a disagregace a stanovit vazby mezi nimi. Kompletní postup teoretické části práce je popsán v kapitole 3.

Na základě teoretické části byly pro metody agregace a disagregace vybrány vhodné postupy pro přepočítání dat. V jazyce Python s balíčkem *arcpy* bylo za účelem automatizace agregačních a disagregačních metod vytvořeno šest skriptů pro agregaci, disagregaci a tvorbu vrstvy pomocných dat. Skripty fungují na dvoukrokovém principu areálové interpolace, je tedy možné je použít i pro gridové sítě. V prostředí ArcGIS Pro byl vytvořen toolbox *Dasymetric.tbx*, do něž byly vytvořené skripty naimportovány. Pro vytvořenou sadu nástrojů byl vytvořen i manuál pro jejich použití jako součást webové stránky práce a ve formátu PDF.

Dílčím výsledkem práce bylo otestovat vytvořené nástroje na datových sadách různého prostorového rozsahu. Za tímto účelem byly na základě konzultace s vedoucím práce vybrány vhodné statistické ukazatele. Pro každé testovací území bylo na základě dostupnosti dat vybráno několik statistických ukazatelů a ty byly poté přepočteny do administrativních jednotek nebo gridových sítí.

Doplňkovým cílem práce bylo provést srovnání několika disagregačních technik na areálové interpolaci počtu obyvatel v rámci Olomouckého kraje. Testování proběhlo pro 16 atributů rastrovým i vektorovým způsobem disagregace. Jako cílové zóny přepočtu byly stanoveny ZSJ a hexagony. Výsledné přepočty byly zvalidovány na dostupných referenčních datech a výsledky validací byly interpretovány. Vybrané výsledky přepočtů byly zvizualizovány v digitální podobě projektu ArcGIS Pro, šest reprezentativních vizualizací bylo zpracováno dle kartografických pravidel do podoby tištěných map.

Závěrem byla sepsána textová část práce. Výsledky práce jsou prezentovány na webových stránkách diplomové práce a na informačním posteru. Celá práce (textová část, přílohy, výstupy a vytvořená data) je dostupná v Příloze 7 (DVD).

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

BATISTA E SILVA, Filipe, Javier GALLEGO a Carlo LAVALLE. A high-resolution population grid map for Europe. *Journal of Maps* [online]. 2013, 9(1), 16–28 [cit. 2021-4-30]. ISSN 1744-5647. Dostupné z: doi:10.1080/17445647.2013.764830

BIELECKA, Elzbieta, 2005. A dasymetric population density map of Poland. *Proceedings of the 22nd International Cartographic Conference*. (48 22), 9–15.

CORINE Land Cover — Copernicus Land Monitoring Service [online]. 2021 [cit. 2021-4-27]. Dostupné z: <https://land.copernicus.eu/pan-european/corine-land-cover>

ČSÚ. *Veřejná databáze VDB* [online]. 2021 [cit. 2021-3-30]. Dostupné z: <https://vdb.czso.cz/vdbvo2/>

ČÚZK - *Mapové služby* [online]. 2020 [cit. 2020-11-9]. Dostupné z: <https://services.cuzk.cz/>

EFGS. *Draft notes* [online]. 2014 [cit. 2020-4-16]. Dostupné z: <https://www.efgs.info/wp-content/uploads/geostat/1b/GEOSTAT1B-final-technical-report.pdf>

EICHER, Cory L. a Cynthia A. BREWER. Dasymetric Mapping and Areal Interpolation: Implementation and Evaluation. *Cartography and Geographic Information Science* [online]. 2001, 28(2), 125-138 [cit. 2021-4-30]. DOI: 10.1559/152304001782173727. ISSN 1523-0406. Dostupné z: <http://www.tandfonline.com/doi/abs/10.1559/152304001782173727>

Esri. *Spatial Aggregation—ArcGIS Insights | Documentation* [online]. 2021 [cit. 2020-4-13]. Dostupné z: <https://doc.arcgis.com/en/insights/latest/analyze/spatial-aggregation.htm>

Esri. *What is Map Algebra?—ArcGIS Pro | Documentation* [online]. 2021 [cit. 2020-4-20]. Dostupné z: <https://pro.arcgis.com/en/pro-app/latest/help/analysis/spatial-analyst/mapalgebra/what-is-map-algebra.htm>

Eurostat. *GEOSTAT - GISCO* [online]. 2021 [cit. 2020-4-16]. Dostupné z: <https://pro.arcgis.com/en/pro-app/latest/help/analysis/spatial-analyst/mapalgebra/what-is-map-algebra.htm>

FLOWERDEW, Robin, Mick GREEN a Evangelos KEHRIS. Using areal interpolation methods in geographic information systems. *Papers in Regional Science* [online]. 1991, 70(3), 303-315 [cit. 2021-4-30]. DOI: 10.1007/BF01434424. ISSN 1056-8190. Dostupné z: <http://doi.wiley.com/10.1007/BF01434424>

GALLEGO, Francisco Javier. A population density grid of the European Union. *Population and Environment* [online]. 2010, **31**(6), 460-473 [cit. 2021-4-30]. ISSN 0199-0039. Dostupné z: doi:10.1007/s11111-010-0108-y

GALLEGO, F.J., F. BATISTA, C. ROCHA a S. MUBAREKA. Disaggregating population density of the European Union with CORINE land cover. *International Journal of Geographical Information Science* [online]. 2011, **25**(12), 2051-2069 [cit. 2021-4-30]. ISSN 1365-8816. Dostupné z: doi:10.1080/13658816.2011.583653

GISGeography. *MAUP - Modifiable Areal Unit Problem* [online]. 2020 [cit. 2021-4-30]. Dostupné z: <https://gisgeography.com/maup-modifiable-areal-unit-problem/>

HOLT, James B., C.P. LO a Thomas W. HODLER. Dasymetric Estimation of Population Density and Areal Interpolation of Census Data. *Cartography and Geographic Information Science* [online]. 2004, **31**(2), 103-121 [cit. 2021-4-30]. ISSN 1523-0406. Dostupné z: doi:10.1559/1523040041649407

HOLT, James B. a Hua LU. Dasymetric Mapping for Population and Sociodemographic Data Redistribution. YANG, Xiaojun, ed. *Urban Remote Sensing* [online]. Chichester, UK: John Wiley & Sons, 2011, 2011-04-13, s. 195-210 [cit. 2021-4-30]. ISBN 9780470979563. Dostupné z: doi:10.1002/9780470979563.ch14

HUANG, Zhengdong, Henk OTTENS a Ian MASSER. A Doubly Weighted Approach to Urban Data Disaggregation in GIS: A Case Study of Wuhan, China. *Transactions in GIS* [online]. 2007, **11**(2), 197-211 [cit. 2021-4-30]. ISSN 1361-1682. Dostupné z: doi:10.1111/j.1467-9671.2007.01041.x

KLAUDA, Petr. Prostorově určená statistická data. *Statistika&My*. Praha: ČSÚ, 2016, 6(5), 18 - 19. ISSN 1804-7149.

KRIVORUCHKO, Konstantin, Alexander GRIBOV a Eric KRAUSE. Multivariate Areal Interpolation for Continuous and Count Data. *Procedia Environmental Sciences* [online]. 2011, **3**, 14-19 [cit. 2021-4-30]. ISSN 18780296. Dostupné z: doi:10.1016/j.proenv.2011.02.004

LI, Tiebei, David PULLAR, Jonathan CORCORAN a Robert STIMSON. *A comparison of spatial disaggregation techniques as applied to population estimation for South East Queensland (SEQ), Australia* [online]. 2007 [cit. 2021-4-30]. DOI: 10.4225/03/57E9AECEBA789. Dostupné z: https://figshare.com/articles/A_comparison_of_spatial_disaggregation_techniques_as_applied_to_population_estimation_for_South_East_Queensland_SEQ_Australia/385980

LI, Tiebei a Jonathan CORCORAN. Testing dasymetric techniques to spatially disaggregate the regional population forecasts for South East Queensland. *Journal of Spatial Science* [online]. 2011, **56**(2), 203-221 [cit. 2021-4-30]. ISSN 1449-8596. Dostupné z: doi:10.1080/14498596.2011.623343

LI, Xiaoma a Weiqi ZHOU. Dasymetric mapping of urban population in China based on radiance corrected DMSP-OLS nighttime light and land cover data. *Science of The Total Environment* [online]. 2018, **643**, 1248-1256 [cit. 2021-4-30]. ISSN 00489697. Dostupné z: doi:10.1016/j.scitotenv.2018.06.244

MENNIS, Jeremy. Generating Surface Models of Population Using Dasymetric Mapping. *The Professional Geographer*. 2003, **55**(1), 31-42. Dostupné z: doi:10.1111/0033-0124.10042

MENNIS, Jeremy. Dasymetric Mapping for Estimating Population in Small Areas. *Geography Compass* [online]. 2009, **3**(2), 727-745 [cit. 2021-4-30]. ISSN 17498198. Dostupné z: doi:10.1111/j.1749-8198.2009.00220.x

MENNIS, Jeremy. Problems of Scale and Zoning. *Geographic Information Science & Technology Body of Knowledge* [online]. 2019, **2019**(Q1) [cit. 2021-4-30]. ISSN 25772848. Dostupné z: doi:10.22224/gistbok/2019.1.2

MPSV. *Statistiky* [online]. 2021 [cit. 2021-4-24]. Dostupné z: <https://www.mpsv.cz/statistiky-2>

OPENSHAW, Stan. *The Modifiable Areal Unit Problem*. Norwich: Geo Books, 1983. ISBN 0-86094-134-5.

OSM - Landuse [online]. 2021 [cit. 2021-4-27]. Dostupné z: <https://osmlanduse.org/>

Otevřená společnost. *MAPAKRIMINALITY.CZ* [online]. 2021 [cit. 2021-4-24]. Dostupné z: <https://www.mapakriminality.cz/>

Policie ČR. *Mapa kriminality* [online]. 2021 [cit. 2021-3-30]. Dostupné z: <https://kriminalita.policie.cz/>

QGIS Documentation. *List of functions* [online]. 2021 [cit. 2021-4-15]. Dostupné z: https://docs.qgis.org/3.16/en/docs/user_manual/working_with_vector/functions_list.html

ROSINA, Konstantin a Pavol HURBÁNEK. Na báze európskych služieb pre monitoring krajiny [online]. *Kartografické listy*, 2012, **20**(2), 56-65. [cit. 2021-4-23] Dostupné z: <https://gis.fns.uniba.sk/kartografickelisty/archiv/KL20/14.pdf>

Spatial Thoughts. *Summary Aggregate and Spatial Filters in QGIS* [online]. 2021 [cit. 2021-4-15]. Dostupné z: <https://spatialthoughts.com/2019/04/12/summary-aggregation-qgis/>

STARÝ, Vlastimil. Prostorová disagregace s využitím dat LU/LC [online]. VŠB – TU Ostrava, 2012 [cit. 2021-04-28]. Dostupné z WWW: http://gisak.vsb.cz/GISacek/GISacek_2012/papers/stary.pdf

SWINK, Morgan a Cheri SPEIER. Presenting Geographic Information: Effects of Data Aggregation, Dispersion, and Users' Spatial Orientation. *Decision Sciences* [online]. 1999, **30**(1), 169-195 [cit. 2021-4-30]. ISSN 0011-7315. Dostupné z: doi:10.1111/j.1540-5915.1999.tb01605.x

ŠIMBERA, J. Neighborhood features in geospatial machine learning: the case of population disaggregation. *Cartography and Geographic Information Science* [online]. 2020, **47**(1), 79-94 [cit. 2021-4-30]. ISSN 1523-0406. Dostupné z: doi:10.1080/15230406.2019.1618201

VOŽENÍLEK, Vít a Jaromír KAŇOK. *Metody tematické kartografie: vizualizace prostorových jevů*. Olomouc: Univerzita Palackého v Olomouci pro katedru geoinformatiky, 2011. ISBN 9788024427904.

WILSON, Benjamin, Neal WILSON a Sierra MARTIN. Using GIS to Advance Social Economics Research: Geocoding, Aggregation, and Spatial Thinking. *Forum for Social Economics* [online]. 2019, 1-25 [cit. 2021-4-30]. ISSN 0736-0932. Dostupné z: doi:10.1080/07360932.2018.1509798

ZAJÍCOVÁ, Hana. Bakalářská práce: Srovnání prostorové reprezentace adresních bodů – grid/areály [online]. 2017. [cit. 2021-4-12] Dostupné z WWW. http://www.geoinformatics.upol.cz/dprace/bakalarske/zajicova17/files/Zajicova_BP.pdf

ZAJÍCOVÁ, Hana. Diplomová práce: Zpracování rozboru udržitelného rozvoje území pomocí gridové analýzy [online]. 2020. [cit. 2021-4-19] Dostupné z WWW. http://www.geoinformatics.upol.cz/dprace/bakalarske/zajicova17/files/Zajicova_BP.pdf

ZANDBERGEN, Paul A. Dasymetric Mapping Using High Resolution Address Point Datasets. *Transactions in GIS* [online]. 2011, **15**, 5-27 [cit. 2021-4-30]. ISSN 13611682. Dostupné z: doi:10.1111/j.1467-9671.2011.01270.x

ZAPLETAL, Jan. Bakalářská práce: Testování disagregačních metod v GIS [online]. 2019 [cit. 2021-4-22]. Dostupné z WWW. http://www.geoinformatics.upol.cz/dprace/magisterske/zajicova20/files/Zajicova_DP.pdf

PŘÍLOHY

SEZNAM PŘÍLOH

Vevázané přílohy

Příloha 1	Počet bytů v České republice v roce 2016
Příloha 2	Průměrný věk v SO ORP ČR v roce 2019
Příloha 3	Nezaměstnanost v České republice v roce 2019
Příloha 4	Index stáří v Olomouckém kraji v roce 2019
Příloha 5	Hustota zalidnění v ZSJ Olomouckého kraje v roce 2011
Příloha 6	Hustota zalidnění v Olomouckém kraji v roce 2011

Volné přílohy

Příloha 7	DVD
Příloha 8	Poster
Příloha 9	Manuál k vytvořeným nástrojům v češtině
Příloha 10	Manuál k vytvořeným nástrojům v angličtině

Popis struktury DVD

Adresáře:

DP_Projekt
DP_Projekt.gdb
DP_Projekt.tbx
DP_Projekt.aprx
Dasymetric.tbx

Prilohy

Skripty

Validace

Korelogramy

Tabulky





Web

Zapletal_DP.pdf

POČET BYTŮ

v České republice v roce 2016


ADMINISTRATIVNÍ ČLENĚNÍ


-  státní hranice
-  krajská hranice
-  hlavní město
-  krajské město

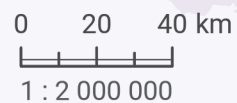
POČET BYTŮ NA KM²

v České republice v roce 2016



 bez výskytu jevu

 5 km²



POČET BYTŮ NA KM²

Autor: Bc. Jan ZAPLETAL

Vznik dat: agregace z bodové vrstvy

Zdrojová data: Statistické budovy - vchody k bytům

Rozloha buňky gridu: 5 km²

Platnost dat ke dni: 1. 4. 2016

Administrativní členění: ČÚZK

Tematický podklad: ČSÚ, RÚIAN

© Univerzita Palackého v Olomouci





Souřadnicový systém: WGS 1984 UTM Zone 33N

Příloha 1

PRŮMĚRNÝ VĚK

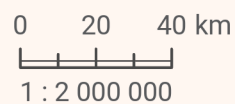
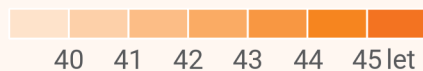
v SO ORP ČR v roce 2019

ADMINISTRATIVNÍ ČLENĚNÍ

-  státní hranice
-  krajská hranice
-  hlavní město
-  krajské město

PRŮMĚRNÝ VĚK

v SO ORP ČR v roce 2019



PRŮMĚRNÝ VĚK

Autor: Bc. Jan ZAPLETAL

Vznik dat: agregace z polygonové vrstvy

Zdrojová data: Počet a věkové složení obyvatel (ČSÚ)

Cílové zóny: SO ORP ČR

Platnost dat ke dni: 31. 12. 2019

Administrativní členění: ČÚZK

Tematický podklad: ČSÚ, RÚIAN

© Univerzita Palackého v Olomouci





Souřadnicový systém: WGS 1984 UTM Zone 33N

Příloha 2

NEZAMĚŠTNANOST v České republice v roce 2019

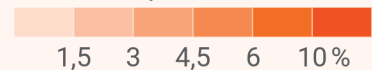
Podíl nezaměstnaných osob vyjadřuje podíl dosažitelných uchazečů o zaměstnání ve věku 15–64 let ze všech obyvatel ve stejném věku (MPSV, 2021). Výsledná hodnota je vynásobením 100 převedena na procenta.


ADMINISTRATIVNÍ ČLENĚNÍ

-  státní hranice
-  krajská hranice
-  hlavní město
-  krajské město

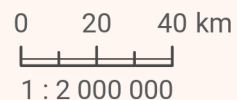
PODÍL NEZAMĚŠTNANÝCH OSOB

v České republice v roce 2019



 bez výskytu jevu

 5 km²



PODÍL NEZAMĚŠTNANÝCH OSOB

Autor: Bc. Jan ZAPLETAL

Vznik dat: vektorová disagregace z obcí ČR

Typ proměnné: intenzivní

Rozloha buňky: 5 km²

Platnost dat ke dni: 31. 12. 2019

Administrativní členění: ČÚZK

Tematický podklad: ČSÚ, RÚIAN

© Univerzita Palackého v Olomouci

Souřadnicový systém: WGS 1984 UTM Zone 33N

Příloha 3

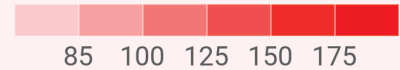
INDEX STÁŘÍ

v Olomouckém kraji v roce 2019

Index stáří vyjadřuje, kolik obyvatel ve věku 65 a více let připadá na 100 dětí do 15 let věku (ČSÚ, 2021). Stanovuje se jako podíl počtu obyvatel v postproduktivním věku ku počtu obyvatel v předproduktivním věku. Tato hodnota je poté vynásobena 100.

INDEX STÁŘÍ

v Olomouckém kraji v roce 2019



bez výskytu jevu

○ 0,5 km²

ADMINISTRATIVNÍ ČLENĚNÍ

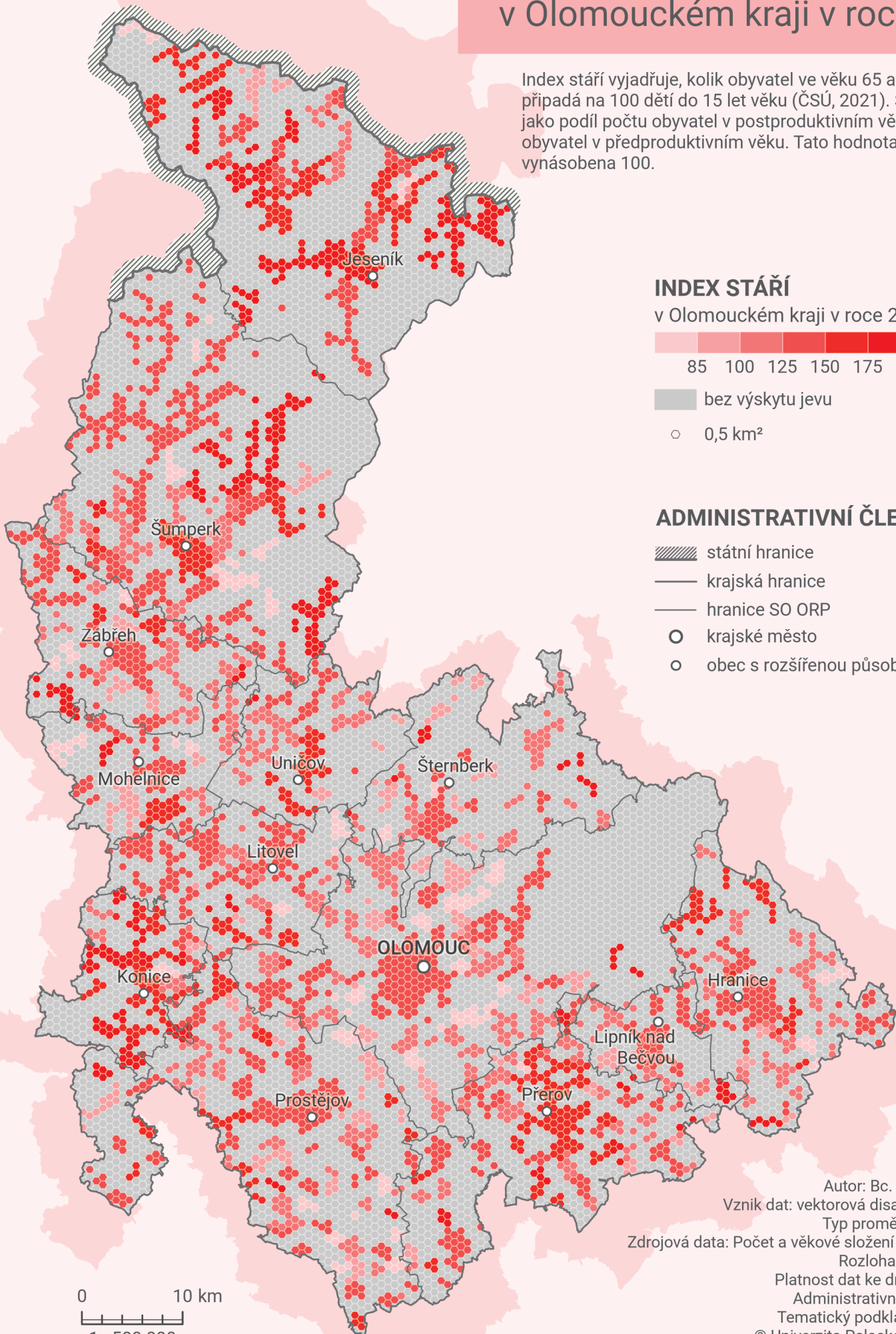
▨ státní hranice

— krajská hranice

— hranice SO ORP

○ krajské město

○ obec s rozšířenou působností



0 10 km
1 : 500 000

INDEX STÁŘÍ

Autor: Bc. Jan ZAPLETAL

Vznik dat: vektorová disagregace z obcí

Typ proměnné: extenzivní

Zdrojová data: Počet a věkové složení obyvatel (ČSÚ)

Rozloha buňky: 0,5 km²

Platnost dat ke dni: 31. 12. 2019

Administrativní členění: ČÚŽK

Tematický podklad: ČSÚ, RÚIAN

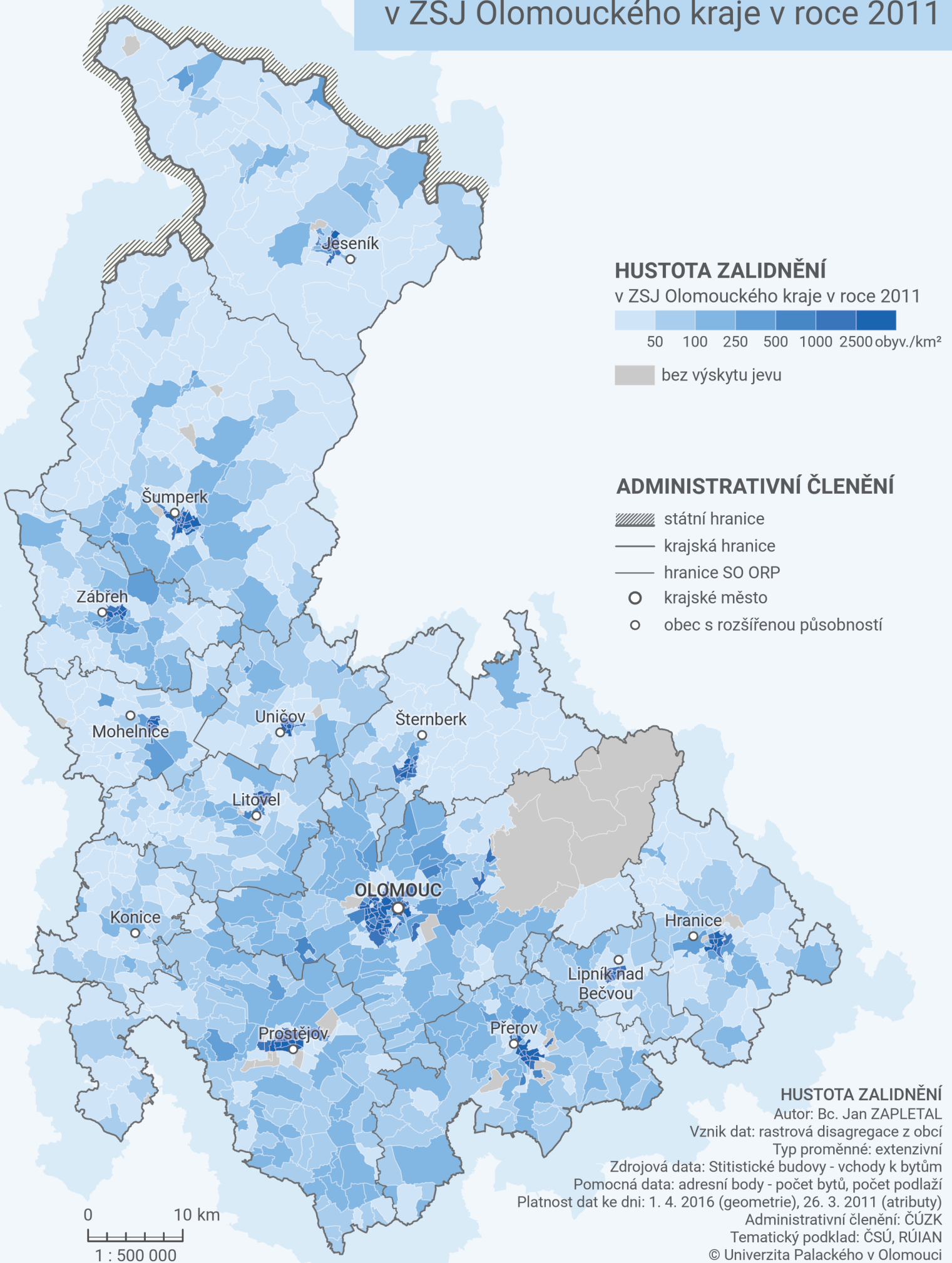
© Univerzita Palackého v Olomouci

Souřadnicový systém: WGS 1984 UTM Zone 33N

Příloha 4

HUSTOTA ZALIDNĚNÍ

v ZSJ Olomouckého kraje v roce 2011



HUSTOTA ZALIDNĚNÍ v Olomouckém kraji v roce 2011

