

Univerzita Palackého v Olomouci

Přírodovědecká fakulta

Katedra geografie

**Prostorové zkreslení a prostorová neurčitost při agregaci dat**

Diplomová práce

Bc. Klára Novotná

Vedoucí práce: doc. Mgr. Pavel Klapka, Ph.D.

Olomouc 2023

## **BIBLIOGRAFICKÝ ZÁZNAM**

**Autor (osobní číslo):** Bc. Klára Novotná (R200068)

**Studijní obor:** Geografie a regionální rozvoj

**Název práce:** Prostorové zkreslení a prostorová neurčitost při agregaci dat

**Title of thesis:** Spatial bias and spatial uncertainty caused by data aggregation

**Vedoucí práce:** doc. Mgr. Pavel Klapka, Ph.D.

**Rozsah práce:** 61 stran, 1 volná příloha

**Abstrakt:** Diplomová práce se zabývá prostorovým zkreslením vznikajícím při agregaci dat. Jsou zkoumány dva konkrétní prostorové jevy a různé typy členění území. Hodnocení zkreslení spočívá v komparaci výsledků statistických a prostorových analýz získaných v jednotlivých regionálních systémech. Agregací efekty se v různé intenzitě objevily u všech testovaných dat.

**Klíčová slova:** prostorová analýza, prostorové zkreslení, agregace dat, MAUP

**Abstract:** The diploma thesis deals with spatial bias arising when data are aggregated. There are two specific spatial phenomena and various forms of zoning systems being examined. The bias evaluation consists of comparison of the results obtained from statistical and spatial analysis performed in each of regional systems. The aggregational effects have occurred in all of the tested data with various level of intensity.

**Keywords:** spatial analysis, spatial bias, data aggregation, MAUP

Prohlašuji, že jsem diplomovou práci vypracovala samostatně pod vedením doc. Mgr. Pavla Klapky, Ph.D. a že veškeré použité zdroje informací jsem uvedla v seznamu zdrojů na konci práce.

V Olomouci dne 20. 4. 2023

.....

Upřímně děkuji doc. Mgr. Pavlu Klapkovi, Ph.D. za cenné rady a podnětné návrhy, které byly poskytnuty při procesu vzniku této práce.

# UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2020/2021

## ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: Bc. Klára NOVOTNÁ  
Osobní číslo: R200068  
Studijní program: N0532A330021 Geografie a regionální rozvoj  
Studijní obor: Geografie a regionální rozvoj  
Téma práce: Prostorové zkreslení a prostorová neurčitost při agregaci dat  
Zadávající katedra: Katedra geografie

### Zásady pro vypracování

Cílem diplomové práce je analyzovat fenomény prostorového zkreslení a prostorové neurčitosti, které vznikají jako důsledek agregace prostorových dat. Zkoumaným územím bude Česká republika, hodnoceny budou různé typy a hierarchické úrovně prostorových jednotek (regionů), administrativní, funkční apod. Regionální systémy a konkrétní prostorové informace budou zvoleny po dohodě s vedoucím práce.

Rozsah pracovní zprávy: 20 000 24 000 slov  
Rozsah grafických prací: Podle potřeb zadání  
Forma zpracování diplomové práce: tištěná

### Seznam doporučené literatury:

- Alvanides, S., Openshaw, S. (1999): Zone design for planning and policy analysis. In: Stillwell, J., Geertman, S., Openshaw, S. eds.: Geographical Information and Planning. Springer, Berlin-Heidelberg, 299&#x2013;315.
- Anselin, L. (1995): Local indicators of spatial association &#x2013; LISA. *Geographical Analysis* 27 (2), 93&#x2013;115
- Cliff, A. D., Haggett, P. (1970): On the efficiency of alternative aggregations in region building problems. *Environment and Planning A* 2 (3), 285&#x2013;294.
- Fischer, M. M., Getis, A. eds. (2009): Handbook of applied spatial analysis: software tools, methods and applications. Springer Science & Business Media, Berlin.
- Fotheringham, A. S., Rogerson, P. A. eds. (2008): The SAGE handbook of spatial analysis. Sage, London.
- Fotheringham, A. S., Wong, D. W. (1991): The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23 (7), 1025&#x2013;1044.
- Fusco, G., Cagliioni, M., Emsellem, K., Merad, M., Moreno, D., Voiron-Canicio, C. (2017): Questions of uncertainty in geography. *Environment and Planning A* 49 (10), 2261&#x2013;2280.
- Gehlke, C. E., Biehl, K. (1934): Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association* 29 (185), 169&#x2013;170.
- Griffith, D. A. (2018). Uncertainty and context in geography and GIScience: reflections on spatial autocorrelation, spatial sampling, and health data. *Annals of the American Association of Geographers*, 108(6), 1499-1505.
- Haining, R. (2003): Spatial data analysis: theory and practice. Cambridge University Press, Cambridge.
- Halás, M., Klapka, P., Erlebach, M. (2019): Unveiling spatial uncertainty: a method to evaluate the fuzzy nature of functional regions. *Regional Studies* 53 (7), 1029 &#x2013; 1041.

- Holt, D., Steel, D. G., Tranmer, M., Wrigley, N. (1996): Aggregation and ecological effects in geographically based data. *Geographical Analysis* 28 (3), 244-261.
- Klapka, P. (2019): Regiony a regionální taxonomie: koncepty, přístupy, aplikace. Univerzita Palackého v Olomouci, Olomouc.
- Klapka, P., Halás, M., Netrdová, P., Nosek, V. (2016): The efficiency of areal units in spatial analysis: Assessing the performance of functional and administrative regions. *Moravian Geographical Reports* 24 (2), 47-59.
- Kwan, M. P. (2012): The uncertain geographic context problem. *Annals of the Association of American Geographers* 102 (5), 958-968.
- Kwan, M. P. (2018). The limits of the neighborhood effect: Contextual uncertainties in geographic, environmental health, and social science research. *Annals of the American Association of Geographers*, 108(6), 1482-1490.
- Openshaw, S. (1977): A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the Institute of British Geographers, New Series* 2 (4), 459-472.
- Openshaw, S. (1984): *The Modifiable Areal Unit Problem*. CATMOG 38. Norwich, GeoBooks.
- Spurná, P. (2008): Prostorová autokorelace a všudypřítomný jev při analýze prostorových dat? *Sociologický časopis/Czech Sociological Review*, 44(4), 767-787

Vedoucí diplomové práce: doc. Mgr. Pavel Klapka, Ph.D.  
Katedra geografie

Datum zadání diplomové práce: 27. ledna 2021  
Termín odevzdání diplomové práce: 10. dubna 2022

L.S.

---

doc. RNDr. Martin Kubala, Ph.D.  
děkan

---

prof. RNDr. Marián Halás, Ph.D.  
vedoucí katedry

V Olomouci dne 27. ledna 2021

## Obsah

1 Úvod.....	9
2 Cíle práce.....	10
3 Teoretický základ .....	11
3.1 Prostorová analýza .....	11
3.2 Vlastnosti prostorových informací a prostorové vzory .....	12
3.3 Prostorová závislost dat.....	13
3.4 Problémy analýzy prostorových dat .....	16
3.5 Agregace prostorových dat a její význam .....	17
3.6 Problémy spojené s agregací geografických informací.....	18
3.6.1 Problém modifikovatelné územní jednotky.....	19
3.6.2 Problém malých čísel .....	21
3.6.3 Ekologická chyba .....	21
4. Metodika práce .....	23
4.1 Teoretické podklady pro výzkum.....	23
4.2 Výběr prostorových jednotek a práce s nimi .....	23
4.3 Výběr prostorových dat .....	27
4.4 Získání dat a výpočty .....	27
4.4.1 Proces získání ukazatelů na obecní úrovni .....	28
4.4.2 Agregace dat.....	29
4.4.3 Výpočet korelace a variability.....	30
4.5 Práce v prostředí GIS .....	30
4.5.1 Kartogramy.....	30
4.5.2 Analýzy prostorové autokorelace .....	30
5 Analýza vlivu agregace na prostorové informace .....	32
5.1 Variabilita.....	32
5.2 Prostorová distribuce.....	34
5.3 Prostorová závislost.....	40
5.4 Korelační koeficient .....	46

6 Diskuze.....	51
6.1 Limity výzkumu .....	54
7 Závěr.....	55
Summary .....	56
Použité zdroje informací .....	57



# 1 Úvod

Geografický prostor tvoří zásadní rozměr geografického bádání a poznání. Je také ale základní dimenzí vstupující do lidského života. Studium povahy geografického prostředí člověkem pokročilo za prosté poznávání míst či určování polohy objektů. Analýza prostoru odkrývá jeho zvláštní vlastnosti, vztahy v něm a mnohotvárnost problematiky vůbec. Tyto roviny nám při běžném pohybu v geografickém prostoru unikají, nebo pro nás mohou mít spíše abstraktní charakter.

Do prostorové analýzy však vstupuje řada problémů vyplývajících zejména ze specifické povahy geografických informací. Tato práce se zaměřuje na problémy spojené se slučováním prostorových dat. Při seskupování prostorových informací detailní měřítkové úrovni do větších územních celků dochází k projevům takzvaného agregačního zkreslení. Fischer (2006) tvrdí, že způsob uspořádání dat do prostorových jednotek ve smyslu různé velikosti i tvaru může modifikovat výsledky analýz těchto údajů, a to zpravidla v nepředvídatelné míře.

Užívání prostorových informací a manipulace s nimi proniká do rozmanité škály nejen vědních oblastí, včetně rezortů řízení a plánování. Nejčastěji jsou používány a interpretovány údaje právě v agregované formě, tedy sloučené do územních jednotek. Jak uvádí Ira a Matlovič (2020), výzvou současné geografie je export jejich znalostí jiných oblastí. V souvislosti s tímto je jistě užitečné upozorňovat na možná úskalí plynoucí z užívání seskupených prostorových dat. Tato rizika jsou při mnohých výkladech i po téměř století od jejich popsání stále podceňována.

Předkládaná práce má snahu zachytit problematiku prostorové agregační neurčitosti na příkladu prostorových jevů, které jsou v době výzkumu aktuálními tématy ve veřejném prostoru. Hodnocení probíhá z velké části v typech územního členění České republiky běžně využívaných pro statistické účely. Odhalení zkreslení by mohlo přispět k uvědomění si nutnosti citlivého přístupu k interpretaci a zacházení s agregovanými daty.

## 2 Cíle práce

Hlavním cílem předkládané práce je identifikace a analýza potenciálního prostorového zkreslení spojeného s agregací prostorových informací nižší hierarchické úrovně do větších územních celků. Prověřováno bude několik typů územního členění České republiky různé hierarchické úrovně. Většinu z nich tvoří administrativními regiony. Ty představují v České republice nejrozšířenější formu pro uspořádání statistických dat, jejich analýzu a interpretaci.

Přítomnost agregačního zkreslení bude ověřována pomocí porovnání statistických analýz zvolených prostorových jevů provedených v rámci jednotlivých systémů členění. Konkrétně bude sledována statistická závislost zvolených proměnných. Případná odlišnost výsledků korelační analýzy v rámci různých systémů členění prokáže, že zvolené jevy a testované regionální systémy podléhají nežádoucím účinkům agregace.

Dílním cílem je studium distribuce studovaných fenoménů v rámci prostorových jednotek. V souvislosti s hledáním potenciálních agregačních efektů bude důležité zaměřit se na komparaci rozložení jevů mezi systémy členění státu. Explanace příčin pozorovaných prostorových vzorů nebudou předmětem výzkumu. Dalším záměrem je studium přítomnosti prostorové závislosti ve vybraných datech. Oba tyto dílní cíle se vztahují k vlastnostem geografických informací, které mohou být zodpovědné za vznik prostorové neurčitosti při agregaci dat.

### 3 Teoretický základ

Třetí kapitola přináší vhled do teoretických konceptů, které tematicky souvisí s výzkumem. Často i základní analýzy prostorových informací naráží na potíže, které vyplývají z jeho komplikované povahy. Práce se zaměřuje na problémy spjaté s agregací geografických informací a jejichž nevídaným dopadem je prostorové zkreslení. Budou charakterizovány důležité vlastnosti prostoru a také některé aspekty statistického zpracování prostorových informací.

#### 3.1 Prostorová analýza

Obecně lze prostorovou analýzu charakterizovat jako souhrn nástrojů a metod, které slouží k rozboru prostorových vzorů, vztahů mezi nimi a také k modelování těchto vztahů (Fischer, 2006). Zohlednění role prostoru ve výzkumných oblastech také mimo geovědní obory jim může napomoci k dosažení lepších výsledků. Přehlížení významu konceptu místa ostatními vědními oblastmi je kritizováno a autoři věří, že právě praktický charakter prostorové analýzy je klíčem k motivaci vědců k interdisciplinárnímu přístupu (Ward a O'Loughlin, 2002; Fischer a Getis, 2010).

Počátky prostorové analýzy sahají do období kvantitativní revoluce v geografii, tedy na přelom 50. a 60. let minulého století. V 70. a 80. letech následoval relativní útlum zájmu o analýzu prostoru (Fotheringham a Rogerson, 2008). Ten souvisel s odklonem geografie od kvantitativních metod díky nástupu nových směrů – behaviorální či humanistické geografie (Klapka, 2019). Mimořádný nárůst zájmu o prostorovou analýzu zaznamenáváme od konce 80. let. Stalo se tak díky pokroku v oblasti geografických informačních systémů (GIS) a s ním spojenou expanzi zájmu o prostorová data a jejich povahu do dalších oborů jako jsou environmentalistika a ekologie, epidemiologie, studie o veřejném zdraví, sociologie, kriminalistika či ekonomie (Fotheringham a Rogerson, 2008; Spurná 2008). Tyto zůstávají spjaty s prostorovou analýzou i nadále, což dokládají soudobé výzkumy Garretona a kol. (2020) o prostorovém zkreslení segregace měst, Vogela (2016) o prostorových jednotkách pro zkoumání kriminality a znevýhodnění čtvrtí, Haláse a Klapky (2020) o vlivu vlastností prostoru na vymezování funkčních regionů Slovenska či od počátku 20. let mnoho prací zabývajících se prostorovou analýzou pandemie COVID-19 (Ehlert, 2021; Bourdin a kol., 2021).

Prostorová analýza bývá nahlížena z různých perspektiv. Jádrem oboru je však práce s prostorovými daty. Díky nestandardní povaze prostorových dat, která bude rozebrána dále, není v rámci prostorových analýz vždy možná aplikace obvyklých (statistických) metod (Mather a Openshaw, 1974; Haining, 2003; Spurná, 2008). Podle Fischera a Getise (2010) čerpá tato disciplína primárně z matematiky, statistiky a ekonometrie. Haining (2003) vyčleňuje tři hlavní oblasti prostorové analýzy:

- kartografické operace s prostorovými daty
- statistická analýza prostorových dat
- matematické prostorové modelování

Jednotlivá oddělení prostorové analýzy jsou navzájem propojená a společným základem je jim také využívání geoinformatických softwarů. Tyto technologie tvoří v prostorové analýze nástroj pro manipulaci s prostorovými daty, jedná se například o prostorové dotazování, překrývání mapových vrstev (Haining, 2003). Sumarizace, vizualizace dat či identifikace odlehklých hodnot souboru jsou funkce GIS potřebné pro průzkumnou analýzu prostorových dat a prostorovou statistiku (Fischer a Getis, 2010). Tyto softwary jsou také důležitou součástí modelování prostorových vzorů a vztahů (Rey, 2015).

Uvedme nyní některé praktické příklady uplatnění metod prostorové analýzy. Nejběžnější formu představují grafické vizualizace prostorových dat, kupříkladu mapy zobrazující přehled výskytu určitého jevu (nezaměstnanosti) v regionech. Další oblastí je výzkum distribuce prostorových informací poskytující mimo jiné možnost identifikovat lokality, kde dochází koncentraci obyvatel. Síťové analýzy uživatelům umožňují hledat nejkratší či nejrychlejší trasu mezi zvolenými body. Sledování směru a intenzity prostorových interakcí přispívá k definování spádových oblastí. A modelování prostorových interakcí na základě těch zjištěných může být užitečným úhlem pohledu na regionální plánování.

### **3.2 Vlastnosti prostorových informací a prostorové vzory**

Prostorové údaje jsou data, která obsahují atributové informace (vyjádření vlastnosti či hodnoty jevu) a zároveň informace o poloze. Jedná se tedy o veškeré skalární i vektorové charakteristiky, které jsou prostorově lokalizované. Praktickým příkladem mohou být informace ze sčítání lidu, meteorologických měření, epidemiologická data či záznamy z družic. Způsobu uspořádání geografických informací v území říkáme prostorový vzor (Unwin, 1996). Prostorové vzory by neexistovaly bez procesů, které je utvářejí. Mezi těmito procesy a pozorovanou distribucí jevů však není přímá shoda. Proto je porozumění prostorovým vzorům komplikované (Fortin a Dale, 2008). Podle Haininga (2010) jsou některé vlastnosti prostorových dat podmíněny povahou reálného geografického prostoru a jiné jsou získány druhotně během procesu reprezentace či měření.

Prostorové vzory jsou utvářeny také vlastnostmi geografického prostoru a charakteru jevů v něm. Jednou ze základních vlastností prostoru je jeho heterogenita označovaná také nestacionárností či variabilita uspořádání. Anselin (1989) popisuje prostorovou heterogenitu jako existenci rozdílů mezi lokalitami. Jiang (2015) koncept specifikuje a připomíná, že přírodní podmínky jsou na zemském povrchu rozloženy nerovnoměrně a rovněž prostorové vzory humánních charakteristik vykazují velkou variabilitu. Podle Haininga (2003) představuje prostor z hlediska diverzity lokalit jakousi přírodní laboratoř pro prostorovou analýzu. Dále autor tvrdí, že je třeba rozlišovat variabilitu prostorových proměnných ovlivněnou „kompozičními“ charakteristikami jako jsou hmotné zajištění obyvatel či jejich věk. Na druhé straně pak vyčleňuje vliv „kontextu“, kdy roli hraje vystavení charakteristikám dané lokality a jako příklad uvádí rozmístění škodlivých vlivů na lidské zdraví (viz také Maškarinec a kol., 2013).

Jak poznamenává Klapka (2019), geografické jevy se vyznačují hierarchickým uspořádáním ve smyslu nadřazenosti a podřazenosti. Zajímavé je včlenění otázky měřítka do hodnocení variability prostorových fenoménů. Známý koncept Korčákova pravostranně šikmého statistického rozdělení, originálně publikovaný Korčákem (1941), mluví o tom, že malých objektů je na zemském povrchu výrazně více než objektů většího rozměru. Platnost tohoto dokládá například tvrzení Mathera a Openshawa (1974), že normální rozdělení geografických jevů se vyskytuje velmi vzácně.

Zásadní charakteristikou geografického prostoru je také jeho kontinuita neboli spojitost. Jedná se o tendenci lokalit k sobě bližších vykazovat podobné hodnoty atributových proměnných. Existence spojitých prostorových vzorů je vlastně podstatou Toblerova prvního geografického zákona (Tobler, 1970). Ten říká, že „všechno souvisí se vším, ale blízké věci spolu souvisí více než věci vzdálené“, což bylo citováno například Goodchildem (2004) nebo Hainingem (2008). Haining (2010) dále přirovnává prostorovou spojitost ke kontinuitě času, protože události na sebe obvykle plynule navazují a nedochází k náhodným změnám z okamžiku na okamžik. Stejně tak distribuce geografických informací není zcela nahodilá a zemský povrch není tvořen množinou izolovaných individualit (Mather a Openshaw, 1974). Díky kontinuálnímu pokrytí prostoru atributy je možné předpovídat neznámé hodnoty jevů, což je podstatou prostorové interpolace (Haining, 2003). Pokud známe sousední hodnoty, lze předpokládat, že neznámé hodnoty atributu budou podobné těm přilehlým známým (Fotheringham a Rogerson, 2008). Praktickým příkladem, který bývá běžně uváděn pro potvrzení této vlastnosti prostoru, je nadmořská výška. Je pravděpodobné, že nadmořská výška dvou míst, která jsou od sebe méně vzdálená, bude podobnější než u dvou míst, která jsou od sebe vzdálenější. Dokladem prostorové kontinuity v humánní sféře mohou být ceny bytů, které jsou typicky podobné u těch v přilehlých lokalitách (Suprná, 2008; Jiang, 2015). Existence kontinuálních prostorových vzorů je podstatou principu prostorové závislosti neboli autokorelace, které bude následně věnována větší pozornost.

### **3.3 Prostorová závislost dat**

Než bude přistoupeno k rozboru problematiky prostorové závislosti, je vhodné objasnit základy závislosti dat obecně. Závislost neboli korelace vyjadřuje vztah mezi dvěma jevy, přesněji řečeno proměnnými. Jde tedy o to, jak změna hodnoty jedné veličiny ovlivňuje hodnotu druhé. Vzorem z geografické praxe může být hodnocení vzájemného vztahu teploty vzduchu a tlaku vzduchu či souvislosti nezaměstnanosti s úrovní vzdělanosti. Za účelem měření povahy a síly lineárního vztahu mezi proměnnými se nejčastěji používá korelační koeficient, nazývaný také Pearsonův koeficient.

Tento ukazatel nabývá hodnot od -1 do 1, kdy výsledek -1 značí absolutní zápornou korelaci a výsledek 1 naprostou kladnou korelaci. Pokud je hodnota koeficientu rovna 0, pak mezi proměnnými není žádný vztah (Freund a Wilson, 2003). Sílu závislosti pak můžeme odvodit z hodnoty výsledného koeficientu. Primárně rozlišujeme tři úrovně závislosti, které popisuje například Turney (2022):

- slabá závislost: 0 až 0,3 (pozitivní) nebo 0,0 až -0,3 (negativní)

- střední závislost: 0,3 až 0,5 (pozitivní) nebo -0,3 až -0,5 (negativní)

- silná závislost: 0,5 až 1,0 (pozitivní) nebo -0,5 až -1,0 (negativní)

Jak již bylo uvedeno, jevy v prostoru mají tendenci spolu souviset a v důsledku toho jsou hodnoty proměnné v jedné lokalitě ovlivňovány hodnotami atributů sousedních lokalit (Anselin, 1995). Pojem prostorová autokorelace v překladu znamená korelaci dané proměnné se sebou samou v prostoru. Podle Getise (2010) se konceptu prostorové závislosti začala výrazněji věnovat pozornost od 50. až 60. let minulého století, čili v období kvantitativní revoluce. Během tohoto období byly prostorové asociace jedním z ústředních témat geografického výzkumu. Za zásadní bývají považovány práce Berryho (1971) či Cliffa a Orda (1969 a 1973), kteří poprvé použili výraz prostorová „autokorelace“ namísto „závislost“ a rozvinuli charakteristiku problematiky (Spurná, 2008).

Prostorová závislost podmiňuje existenci seskupení podobných hodnot v území, kterým říkáme prostorové shluky. Podobně jako u statistické korelace dat, rozlišujeme pozitivní či negativní korelaci v prostoru. Pokud hovoříme o **pozitivní prostorové autokorelaci**, máme na mysli právě prostorové shlukování. Tedy když se v blízkosti vysokých hodnot nachází vysoké hodnoty proměnné nebo když jsou v blízkosti nízkých hodnot atributu získány rovněž nízké hodnoty. Naproti tomu **negativní prostorová autokorelace** značí, že se v lokalitě vyskytují rozdílné hodnoty, takže v okolí vysokých hodnot pozorujeme nízké hodnoty veličin a obráceně. Jestliže není zjištěna pozitivní ani negativní korelace mezi pozorováními v prostoru a není mezi nimi významný vztah, jedná se o **nulovou prostorovou autokorelaci** (Spurná, 2008; Haining, 2010). Většina geografických jevů je určitým způsobem pozitivně autokorelovaná (Fotheringham a kol., 2003).

Z výše uvedeného je zřejmé, že prostorová autokorelace souvisí nejen s kontinuitou, ale také s heterogenitou prostoru. Meentemeyer (1989) uvádí, že tento obor je nástrojem pro identifikaci prostorové variability, jelikož se zabývá kontrastem mezi nízkými a vysokými hodnotami jevů v prostoru. Se zvyšující se vzdáleností mezi lokalitami podobnost atributů klesá, jak ostatně vyplývá z Toblerova geografického zákona (Haining, 2008).

Pro určení významnosti prostorové závislosti geografických dat bylo vyvinuto několik způsobů. Mnoho z nich je dostupných v geoinformatických softwarech. První skupinou měřicích prostředků jsou globální indexy prostorové autokorelace. Nejznámější statistikou z nich je takzvané Moranovo I kritérium (Anselin, 1995). To je vyjádřeno vzorcem:

$$I = \frac{k \sum_{i=1}^k \sum_{j=1}^k w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^k \sum_{j=1}^k w_{ij} \sum_{i=1}^k (y_i - \bar{y})^2}$$

kde **k** označuje počet objektů, **y<sub>i</sub>** je hodnota ukazatele v jednotce **i** a **y** je aritmetickým průměrem analyzované proměnné (Cliff a Ord, 1973). Výpočetně Moranovo I vychází z Pearsonova korelačního koeficientu. Výsledkem globálních indexů autokorelace a rovněž Moranova I kritéria je číselná hodnota

(rovněž v intervalu od -1,0 do 1,0), která udává míru prostorové autokorelace zkoumaného jevu v celé studované oblasti vstupující do analýzy. U globálních měř existuje riziko zprůměrování podstatných regionálních rozdílů. Můžou tak být skryty lokální prostorové shluky či odlehle hodnoty (Spurná, 2008; Netrdová a Nosek, 2009). Je nepravděpodobné, že je platnost ukazatele možné vztáhnout na všechny části studované oblasti (Maškarinec, 2013).

V 90. letech 20. století si výzkumníci v prostorové analýze začali uvědomovat nutnost změny přístupu ke zkoumání závislostí v prostoru. Je formován koncept druhé skupiny metod měřících prostředků, a sice lokálních indikátorů prostorové autokorelace (Anselin, 1995). Tyto indexy jsou pro potřebu identifikace prostorového shlukování uvnitř sledovaného území vhodnější a jejich předností je možnost zobrazit je na mapě (Klapka a kol., 2016). Spurná (2008) poukazuje na společenskovední oblasti, ve kterých je možné uplatnit tyto metody. Zmiňuje například výzkumy cen nemovitostí, kriminální činnosti či prostorové analýzy voleb.

Významnou prací je výzkum Getise a Orda (1992), ve kterém autoři představili míry  $G_i$  a  $G_i^*$  pro zkoumání lokálních vzorů prostorové závislosti. Nejrozšířenějším typem lokálních charakteristik prostorové autokorelace je sada lokálních indikátorů prostorové autokorelace (LISA) zavedená Anselinem (1995). Díky nim dochází k rozkladu globálních charakteristik a jsou zkoumána jednotlivá pozorování uvnitř celé oblasti. Může tak dojít k odhalení místního vzoru prostorové autokorelace, který by globální míra nezachytila. Součet lokálních indikátorů autokorelace pro jednotlivá místa je pak úměrný globálnímu indexu za celou oblast (Anselin, 1995). Jedná se o lokální verzi globální statistiky Moranovo I. Výsledky analýzy LISA lze nejpřehledněji interpretovat pomocí tzv. Moranova diagramu. V něm v rámci pozitivní prostorové autokorelace (PA) rozlišujeme tzv. **hot spots** – horká místa označující shluky nadprůměrných hodnot a tzv. **cold spots** – chladná místa vztahující se k prostorovým seskupením podprůměrných hodnot proměnné. Pod výrazem **spatial outliers** – prostorové odchylky rozumíme negativní autokorelaci, kdy se v okolí podprůměrné hodnoty nachází hodnota nadprůměrná a opačně (Spurná, 2008). Je třeba dodat, že výsledky poskytnuté metodikou LISA mají pouze popisný charakter. Pomocí těchto statistik nelze objasnit důvody uspořádání hodnot v prostoru ani zjištěných závislostí (Maškarinec, 2013).

vážená hodnota proměnné v blízkých prostorových jednotkách	spatial outliers <b>nízká – vysoká</b> negativní PA	hot spots <b>vysoká – vysoká</b> pozitivní PA
	cold spots <b>nízká – nízká</b> pozitivní PA	spatial outliers <b>vysoká – nízká</b> negativní PA
	hodnota proměnné v prostorové jednotce	

Obrázek 1: Moranův diagram

Zdroj: Spurná (2008); vlastní zpracování

### 3.4 Problémy analýzy prostorových dat

V této části budou rozebrány problematické aspekty, které vstupují do prostorových, potažmo geografických analýz. Pozornost bude věnována konceptu nejistoty/neurčitosti v geografii a nestandardnímu charakteru prostoru a prostorových jevů. Zvláštní zřetel bude brán také na fenomén prostorové autokorelace. Komplikacím souvisejícím s analýzou agregovaných prostorových dat bude věnována pozornost v samostatné kapitole 3.6.

Stejně jako je tomu v mnoha jiných oborech, také v geovědních oblastech komplikuje výzkumy koncept nejistoty a s ní související riziko vzniku omylů. Tato nejistota pramení z různých zdrojů. Jako první příklad uveďme subjektivní vnímání míst, které se mezi jedinci může lišit. Tím pádem různí lidé mohou reagovat na vlivy daného prostředí odlišně, což může vést k pochybením ve studiích vlivů prostředí na individuální úrovni (Kwan, 2018). Dalším aspektem je nejednoznačnost vymezení geografických pojmů. Fusco a kol. (2017) jmenují za složitě definovatelné například výrazy systém, region nebo hranice. Dále uvádí odlišné vymezení města a venkova napříč různými státy, což může komplikovat mezinárodní komparace.

Nejvýznamnějším zdrojem nejistoty a souvisejících pochybení je pravděpodobně komplikovaná povaha prostoru a prostorových informací. Běžně totiž pracujeme pouze se zjednodušenou verzí geografického prostoru, který nás reálně obklopuje. Prostorové údaje jsou pro usnadnění jejich zpracování a vizualizace uloženy v datových maticích. Ty jsou výsledkem přenosu komplexní geografické reality do digitální formy. Diskretizace prostoru vzhledem k jeho komplexitě a heterogenitě s sebou nese riziko přílišného



zjednodušení (Halás a Klapka, 2020). Problematická je jednak volba prvků pro reprezentaci skutečně existujících objektů a procesů a jednak proces sběru a zpracování dat (Haining, 2003, 2010). Fischer (2006) tvrdí, že téměř veškerá prostorová data jsou více či méně chybná. Na vině je podle něj mimo jiné ukládání dat a manipulace s nimi pomocí počítačových procesů. Pochybení dále vznikají, jak bylo zmíněno, při výběru vzorku, měření a výpočtech (Griffith, 2018). Tato nedopatření mohou vyústit v chybné interpretace výsledků analýz či k jejich zkreslení (Fischer, 2006). Jiang (2015) spatřuje problém v tzv. „paradoxu prostorové heterogenity“. Jeho podstatou je, že ačkoli se distribuce geografických informací značně liší od normálního neboli Gaussova rozdělení, používá prostorová statistika metody založené na průměrných hodnotách a normálním rozdělení, včetně metod měření prostorové autokorelace.

Prostorová autokorelace bývá často vymezována nejen jako vlastnost závislosti prostorových dat, ale také jako potenciální problém statistických analýz (Spurná, 2008; Maškarinec, 2013). Neprostorové metody aplikované na prostorová data mohou působit zkreslení výsledků. Tyto postupy totiž předpokládají nezávislost pozorování v prostoru (Fischer, 2006). Problematický je také vztah globálního a lokálního vzoru prostorové závislosti. Je obtížné specifikovat charakter lokální prostorové autokorelace, je-li zároveň prokázána přítomnost globální asociace. Lokální prostorové závislosti mohou být do určité míry vedlejším efektem rozsáhlé globální prostorové autokorelace (Getis, 2010). Griffith (2018) shledává nežádoucím přílišné zaměření studií na pozitivní prostorovou autokorelaci. Ve zkoumání negativní prostorové autokorelace vidí potenciál k identifikaci konkurenčního efektu v území. Obvykle podle něj však zjištěná negativní autokorelace bývá interpretována jako výsledek nevhodně zvolených územních jednotek pro analýzu. Spurná (2008) uvádí, že pro zkoumání lokální prostorové závislosti sociálních jevů je vhodné použít data agregovaná na úrovni obcí či základních sídelních jednotek. Analýza pro větší územní celky by poskytla velmi zjednodušené, a tudíž nepřesné výsledky. Podrobněji bude problematika volby územních jednotek pro prostorové analýzy popsána v následující kapitole. Getis (2010) poukazuje na fakt, že i samotné lokální statistiky prostorové autokorelace mohou podléhat jejím účinkům. Díky testování každé lokality zvláště totiž existuje množství jednotlivých testů, které jsou na sobě navzájem závislé.

### **3.5 Agregace prostorových dat a její význam**

S prostorovými údaji individuální úrovně se setkáváme zřídka. V běžné praxi jsou geografické informace slučovány a v agregované formě jsou dále prezentovány. To znamená, že jev obvykle zkoumáme v určitých územních jednotkách. Uvnitř těchto jednotek ale existují individuální objekty, ve kterých jsou uskutečněna pozorování daných jevů. Příkladem z humánně geografického výzkumu mohou být obce, které obsahují jednotlivé domácnosti a navíc i jednotlivé osoby.

Důvodů, proč jsou používána statistická data a zvláště data prostorového charakteru v agregované formě, je několik. Jako první zmiňme potřebu ochrany soukromých informací, například při sčítání lidu

(Haining, 2008). Seskupení důvěrných údajů poskytnutých jedinci zajistí jejich anonymitu (Holý, 2011). To je příčinou skutečnosti, o které hovoří Klapka a kol. (2016), že prostorová data, která jsou nám běžně dostupná, jsou vykazována skupinově pro určité regiony. Agregace dat je však také užitečnou součástí metodiky zpracování (nejen) geografických informací.

Přehled přínosu slučování geografických informací do územních jednotek v prostorové analýze zdravotnických údajů přináší Wise a kol. (1997). Jejich postřehy však můžeme vztáhnout na prostorové úlohy obecně (viz například Klapka, 2019). Autoři Wise a kol. (1997) hovoří o tom, že díky agregaci údajů do územních celků vyšší úrovně dojde v těchto jednotkách ke zvýšení počtu pozorování. Tato skutečnost snižuje náchylnost dat ke zkreslení způsobenému odlehlými hodnotami. Dalším pozitivem vyplývajícím ze zvětšení statistického vzorku je jeho větší odolnost vůči možným chybám. Ve větším datovém souboru vzniklém seskupením bude mít účinek takovýchto nepřesností menší intenzitu. Jednoduše řečeno, snadněji se „ztratí“. Tato konstatování souvisí s takzvaným efektem malých čísel, který bude rozebrán později. Dále zmíněný kolektiv autorů uvádí, že agregovaná data usnadňují statistické výpočty, jelikož do analýzy vstupuje mnohem méně prvků, než kdybychom počítali s individuálními pozorováními. Se zjednodušením struktury dat souvisí také poslední z přínosů jejich slučování, a sice usnadnění vizualizace. Mapy a grafy, kde jsou prostorové informace agregované do územních celků, jsou přehlednější a je snadnější se v nich zorientovat.

V určitých případech je slučování prostorových informací podmínkou pro to, aby prostorová analýza měla smysl. To zdůvodňuje Openshaw (1984) příkladem ze zkoumání zemědělských výnosů, přičemž vychází z publikace dvojice autorů Yule a Kendall (1950). Tento doklad nezbytnosti agregace spočívá ve skutečnosti, že chceme-li analyzovat výnosy plodin, je nutné do výzkumu zahrnout jak část půdy obsahující jednu plodinu, tak části půdy, na kterých se pěstuje druhá plodina (případně další). Podrobit zkoumání prostorové variability či asociace pouze jednotlivé záhony by postrádalo smysl

### **3.6 Problémy spojené s agregací geografických informací**

Tato problematika je klíčová pro předkládaný výzkum. Je poukázáno na efekty, které vyplývají z prostorové analýzy agregovaných údajů. Tyto jsou spojeny s obtížnou interpretací výsledků a potenciálním zkreslením prostorových vzorů. Řadou studií bylo prokázáno, že zvolený způsob agregace geografických informací do prostorových celků ovlivňuje výsledky analýz těchto dat. Riziko je o to vážnější, že definice prostorových jednotek nemá žádná závazná a jasně daná pravidla.

Agregace prostorových údajů automaticky znamená potřebu územního členění a tím pádem nutnost vedení hranic náležících jednotlivým zónám napříč zájmovou oblastí (Mather a Openshaw, 1974). Tento požadavek je výchozím momentem vzniku potíží. Opět je třeba hledat souvislost s charakterem prostoru. Fusco a kol. (2017) upozorňují na rozpor mezi kontinuální geografickou realitou a ostrým vymezením logických kategorií, k nimž běžně používané hranice prostorových jednotek rozhodně patří. Roli hraje také to, z pohledu které geografické sféry se na problematiku díváme. Alvanides a Openshaw (1999)

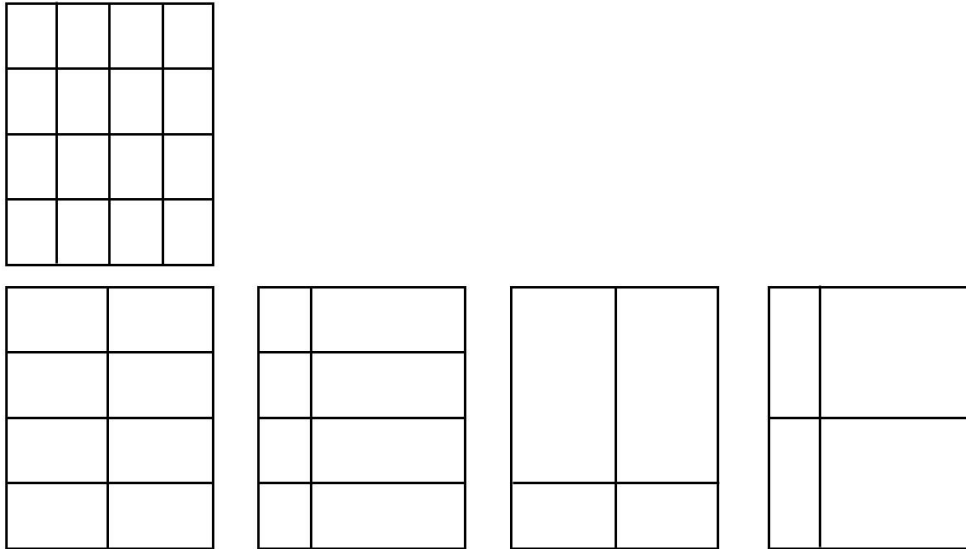
uvádí, že u fyzickogeografických objektů lze v mnohých případech identifikovat hranice (např. ostrov či pole). Jinak je tomu v humánní sféře, kde málo jevů vytváří přirozená rozhraní a jako příklad autoři zmiňují trh práce. Následuje popis tří navzájem propojených problémů prostorové analýzy agregovaných dat.

### 3.6.1 Problém modifikovatelné územní jednotky

Problém modifikovatelné územní jednotky (MAUP z anglického modifiable areal unit problem) je považován za významnou komplikaci prostorové analýzy. Jeho řešení je prakticky nemožné, a proto je potřeba usilovat alespoň o jeho zmírnění. Vyplyvá z výše zmiňované nejednoznačnosti vymezení prostorových jednotek. Jak poznamenává Openshaw (1984), existuje téměř nekonečné množství způsobů územního členění určité oblasti. Alvanides a Openshaw (1999) připomínají, že badatel má díky geoinformatickým softwarům téměř neomezenou moc nad výslednou formou systému jednotek a může ji svévolně modifikovat. Opět zde narážíme na koncept subjektivity v geografii.

Modifikovatelnost územního členění by sama o sobě nebyla tolik problematická, pokud by způsob jeho vymezení neovlivňoval analýzy prostorových informací v jednotkách územního členění. Jak poznamenávají Mather a Openshaw (1974), geografická data jsou citlivá na způsob utřídění agregovaných údajů do plošných jednotek. Podle Klapky a kol. (2016) při nevhodném uspořádání může dojít ke skrytí významných prostorových vzorů a také k chybným interpretacím. Holt a kol. (1996) upozorňují na fakt, že výsledky statistických výpočtů (například korelace dvou proměnných) do značné míry záleží na úrovni agregace dat. Dosáhneme jiných hodnot, pokud danou analýzu provedeme v územních jednotkách s agregovanými daty a odlišných výsledků při stejném výpočtu nad stejnými vstupními, ale nesloučenými daty. Jak autoři dále poznamenávají, nejde pouze o hierarchickou úroveň agregace. Výsledné hodnoty takovéto statistiky budou totiž velmi pravděpodobně odlišné i v případě, že vstupní data sloučíme dvěma různými způsoby do shodného počtu stejně velkých uskupení, takže jiné bude pouze vytyčení hranic jednotek (Holt a kol., 1996; Haining, 2003). Zmíněné dva příklady citlivosti dat nastiňují, že MAUP sestává ze dvou podřazených problémů, které vyvstávají, když hledáme, jakou formou agregovat prostorové informace.

Prvním z nich je **problém měřítka**, který se pojí s problematikou hierarchického uspořádání prostoru. Není totiž jasné, na jaké úrovni měřítka by mělo být členění území provedeno. Řešíme tedy, do kolika prostorových celků by měla být data sloučena, a tudíž jaká by měla být jejich velikost (Openshaw, 1984). Druhou subdivizí MAUP je **problém agregace**, někdy označován také jako problém zonace. Ten se týká již konkrétní stanovené hierarchické úrovně. Jedná se o nejednoznačnost způsobu sloučení menších územních celků do požadovaného počtu větších územních celků (Openshaw, 1977). Jinak řečeno, jde o to, jak konfigurovat hranice výsledných zón (Fotheringham a Wong, 1991). Měřítkovou a zónační variabilitu znázorňuje obrázek 2. Je uveden příklad různého uspořádání výchozího systému šestnácti objektů do osmi a čtyř agregovaných celků.



Obrázek 2: Příklad měřítkové a zónační variability

Zdroj: upraveno podle Openshaw (1984)

Významným projevem agregčního zkreslení je shlazení variability seskupených údajů (Haining, 2003), které je zdrojem problému měřítka. Pokud se v regionech se seskupenými daty vyskytují jednotlivá místa s extrémními hodnotami obklopená průměrnými, budou extrémní pozorování v agregátech méně patrná (Wong, 2008). To proto, že různé individuální hodnoty budou nově reprezentovány souhrnnou hodnotou pro agregované oblasti (Wong, 1996). Agregční účinky ve statistických analýzách byly identifikovány ve 30. letech minulého století, když Gehlke a Biehl (1934) zaznamenali rostoucí závislost proměnných se zvyšující se úrovní agregace, což později prokázali i další autoři, například: Openshaw a Taylor (1979), Openshaw (1984) nebo Fotheringham a Wong (1991). Jak uvádí Andresen (2021), mnohé tyto experimenty ukázaly, že změnami členění území lze dosáhnout různých hodnot vzájemné korelace proměnných. Koeficienty měření závislosti při různých formách agregace ukazovaly na kladnou, jindy na zápornou korelaci. Výpočet korelačního koeficientu souvisí s variabilitou proměnné. Klesá-li variabilita proměnných s agregací, lze očekávat také změnu jejich vzájemné korelace (Fotheringham a Wong, 1991; Wong, 2008). Efekty MAUP ve vícerozměrné analýze však autoři Fotheringham a Wong (1991) označují za nepředvídatelné. Vliv na MAUP může mít také přítomnost prostorové závislosti. Důsledky prostorové asociace jsou patrné především vůči problému zonace, což dokazují Openshaw (1984) a Wong (2008). Regionální systém tak může obsahovat prostorové jednotky, které jsou tvořeny celými shluky podobných hodnot, nebo mohou být tyto oblasti koncentrace rozděleny mezi jednotlivé zóny a jejich význam zanikne (Garreton a kol., 2020).

Přestože bylo vynaloženo mnoho úsilí, nebyl nalezen způsob, jak vyřešit problém modifikovatelné územní jednotky (Andresen, 2021). Přínosy dosavadních výzkumů spočívají ve způsobech, jak zmírnit jeho účinky. Protože MAUP je důsledkem agregace dat, jako řešení se nabízí použití neseskupených

údajů, což však není schůdným řešením (Fotheringham a Wong (1991), jak ostatně vyplývá z již vzpomenutých příkladů nezbytnosti agregace. Zásadní tak zůstává snaha hledat optimální způsob seskupení dat, neboli takový, který minimalizuje zkreslení skutečné distribuce geografických fenoménů. Jak poznamenává Meentemeyer (1989), nejjednodušší cestou k odhalení prostorových vzorů je použití takového způsobu agregace, který maximalizuje variabilitu dané proměnné. To však záleží na povaze zkoumaného procesu (Hipp, 2007). V běžné praxi jsou však jednotky územního členění vytvořené pro jeden konkrétní účel používány k vykazování nejrůznějších statistik (Alvanides a Openshaw, 1999).

### **3.6.2 Problém malých čísel**

Problém či efekt malých čísel (small number problem/effect) souvisí s různou velikostí vzorku v agregovaných jednotkách. Jeho důsledkem jsou potenciální chyby v interpretaci výsledků analýz. Menší statistický vzorek mimo jiné znamená menší odolnost dat vůči nepřesnostem v měření či vůči odlehkým hodnotám souboru (Haining, 2010). Komplikace vznikají ale také tehdy, pokud je počítána míra nějakého jevu v oblastech s podprůměrnou velikostí vzorku, obvykle s nižším počtem obyvatel (Gelman a Price, 1999). Při přepočtu údajů na počet obyvatel v těchto územích bude ve jmenovateli menší číslo a výsledný podíl tak bude vyšší než v ostatních prostorových jednotkách (Haining, 2003).

Problém malých čísel bývá skloňován ve spojitosti s prostorovou analýzou jevů s méně častým výskytem, často rizik – nemocí či trestných činů (Haining, 2003). V důsledku efektu malých čísel může být míra rizika v oblastech s nižší hustotou zalidnění přeceňována. Alvanides a Openshaw (1999) zmiňují různou velikost regionů NUTS, které Evropská unie používá mimo jiné pro alokaci finančních prostředků pro rozvoj slabších území. Upozorňují na kontroverzní fakt, že existuje možnost, že státům s menšími oblastmi NUTS mohou být přidělovány vyšší částky, jelikož vykazují extrémnější hodnoty negativních jevů.

Úvahy o předcházení tomuto problému vedou opět ke snaze o použití optimálních zón, v tomto případě s vyrovnanou velikostí vzorku (populace). Takový výběr je ale značně limitován MAUP a vyřešení jednoho problému by tak podpořilo vznik jiného.

### **3.6.3 Ekologická chyba**

Ekologická chyba je propojena s problémem měřítka v rámci problému modifikovatelné územní jednotky. Jelikož s vyšší úrovní agregace klesá variabilita datového souboru, zvyšuje se míra závislosti mezi proměnnými, což bylo vysvětleno v části 3.6.1. Je nesprávné očekávat platnost charakteristik zjištěných z agregovaných údajů pro individuální úroveň (Wong, 1996). Zpravidla se jedná o falešné korelace proměnných na vyšší měřítkové úrovni, ze kterých jsou vyvozovány závěry pro nižší úroveň (Meentemeyer, 1989).

Pojem ekologická chyba zavedl Robinson ve svém výzkumu publikovaném v roce 1950. Ten měřil vzájemnou korelaci mezi imigranty jiné rasy úrovní gramotnosti obyvatel v USA (Haining, 2010). Zatímco na makroregionální úrovni byla zjištěna velmi silná závislost, na individuální úrovni byl vztah

mezi barvou pleti a IQ nevýznamný (Meentemeyer, 1989). Z výsledků na makroúrovni bychom mohli mylně usoudit, že obyvatelstvo s odlišnou barvou pleti je zodpovědné za vyšší charakteristiky negramotnosti. S větší pravděpodobností však spíše docházelo k usazení přistěhovalců v oblastech s vyšší gramotností například z důvodu jejich lepší prosperity (Pollet a kol., 2014). Jak uvádí Salkeld a Antolin (2020), falešné závislosti v agregovaných datech ukázalo také několik dalších experimentů, například korelaci mezi spotřebou čokolády a četností udělení Nobelovy ceny. Sami pak objevují mimo jiné závislost mezi onemocněním boreliózou a výskytem restaurací se smaženými kuřaty.

Openshaw (1984) podotýká, že pokud by agregované územní jednotky byly vnitřně homogenní, k ekologickému klamu by nedocházelo. To je ale v rozporu s prostorovými vzory geografických jevů, které se vyznačují svojí heterogenitou. Podle Salkelda a Antolina (2020) se ekologické omyly ve vědeckém výzkumu objevují poměrně často a je nezbytné korelace v agregovaných oblastech ověřovat na individuální úrovni. To je ale problematické, pokud nejsou neagregovaná data dostupná (Fischer, 2006).

## 4. Metodika práce

### 4.1 Teoretické podklady pro výzkum

Z důvodu nezbytnosti postihnout teoretická východiska výzkumu, uchopit jejich souvislost s cíli práce a jejich význam, byla výchozím bodem literární rešerše. Základ tvořila doporučená literatura ze zadání práce, která byla postupně doplňována dalšími zdroji. Většinu použitých podkladů tvořily zahraniční vědecké články spojené s tematikou geografického prostoru, jeho analýzy a agregčního zkreslení prostorových analýz. Publikace se zaměřovaly na vlastnosti prostoru, jejich specifika a související příčiny i důsledky prostorové neurčitosti.

Dokumenty se podrobně věnovaly problémům prostorové analýzy, obsahovaly matematické a statistické výpočty a navrhovaly řešení těchto problémů. Tyto napomáhaly proniknutí do metodických konceptů samotného výzkumu. Mnohé z těchto prací přinesly zásadní zjištění, což bylo typické především pro druhou polovinu minulého století. Z tohoto důvodu jsou poměrně hojně citovány zdroje ze 70. až 90. let. Užitečné byly také vybrané kapitoly příruček prostorové analýzy. Tyto příručky přináší souhrnný, ale podrobný popis aspektů oboru, často také interpretují či hodnotí zmíněné důležité výzkumy z 20. století.

### 4.2 Výběr prostorových jednotek a práce s nimi

Oblastí zájmu tohoto výzkumu je území České republiky. Pro následné statistické analýzy prostorových informací bylo nutné vybrat různé typy členění zájmového území. Je použito celkem pět systémů prostorových jednotek. Zastoupeny jsou územní celky různé hierarchické úrovně, ale zároveň také rozdílné typy regionů (administrativní i vymezené na základě funkčních vztahů). Jejich přehled zachycuje tabulka 1.

Tabulka 1: Přehled použitých systémů členění území

typ systému členění	název systému členění	počet územních jednotek v systému
ADMINISTRATIVNÍ	obce	6258
	správní obvody obcí s rozšířenou působností (SO ORP) a hlavní město Praha	206
	okresy a hlavní město Praha	77
	kraje	14
FUNKČNÍ	funkční regiony dle pravidelných denních pohybů mobilních telefonů	113

Zdroj: vlastní zpracování; 2023

## **Administrativní regiony**

Územní celky administrativního typu, které jsou běžnými jednotkami územního členění státu, nebudou blíže charakterizovány. Pro úplné pokrytí území státu bylo hlavní město Praha v rámci analýz uvažováno jako součást regionálních systémů okresů a správních obvodů obcí s rozšířenou působností, i když území Prahy není těmito regiony pokryto (Zákon o územně správním členění státu, 2021). Pro následné analýzy v prostředí GIS byly využity čtyři polygonové vrstvy administrativních regionů z databáze ArcČR verze 4.1 bezplatně poskytované firmou Esri.

Aby bylo možné agregovat prostorové informace na nejdetailnější měřítkové úrovni (obce) do větších územních celků, bylo nutné opatřit pro každou obec kód regionu, do kterého spadá. Kódy byly obcím přiřazeny z datových sad ČSÚ pro Územně analytické podklady.

## **Funkční regiony**

Vybraný funkční regionální systém odráží pravidelné denní toky v rámci České republiky odvozené z lokalizačních dat uživatelů významného mobilního operátora. Sběr dat o poloze mobilních telefonů probíhal po dobu čtyř týdnů, na podzim roku 2019. Výsledkem je 113 funkčních regionů se srovnatelnou rozlohou v rámci systému. Jsou složeny z 1451 tzv. elementárních jednotek (EJ), což jsou seskupení obcí, kdy průměrný počet obcí v takové jednotce je čtyři. Autoři pro zformování systému aplikovali hojně užívaný regionalizační postup CURDS (Halás a kol., 2021). Detailnější charakteristika procesu vzniku popisovaného regionálního systému není pro potřeby této práce důležitá.

Autoři funkčního regionálního systému poskytli pro účely této práce tabulku ve formátu XLSX obsahující kódy obcí a zmiňovaných elementárních jednotek a také polygonovou vrstvu 1451 jednotek. Příslušnost EJ k funkčnímu regionu je vyjádřena kódem v atributové tabulce vrstvy. Kódy náležitosti obcí k EJ se v několika případech neshodovaly s příslušností obcí do EJ v rámci polygonů. Kódy zařazující obce do základních jednotek byly upraveny dle polygonové vrstvy, což bylo umožněno porovnáním vrstvy EJ s vrstvou obcí. Opravy se týkaly nižších desítek obcí. V prostředí GIS, konkrétně v programu ArcMap došlo ke sloučení vrstvy EJ do výsledné vrstvy funkčních regionů. K seskupení došlo pomocí nástroje Dissolve ze skupiny nástrojů Data management. Díky němu lze spojovat prvky na základě zvoleného atributu, kterým byl v tomto případě kód elementární jednotky.

Následující obrázky ukazují vymezení použitých regionálních systémů a jejich názvů s výjimkou obecní úrovně. Plné názvy některých území jsou pro usnadnění přehlednosti zkráceny. Systémy členění jsou prezentovány postupně od nejnižší hierarchické úrovně (SO ORP) po nejvyšší (kraje).





Obrázek 3: Vymezení správních obvodů obcí s rozšířenou působností (206 jednotek)

Zdroj: ArcČR® 4.1, 2022; vlastní zpracování



Obrázek 4: Vymezení funkčních regionů z lokalizačních dat mobilního operátora (113 jednotek)

Zdroj: Halás a kol., 2021; vlastní zpracování



Obrázek 5: Vymezení okresů (77 jednotek)

Zdroj: ArcČR® 4.1, 2022; vlastní zpracování



Obrázek 6: Vymezení krajů (14 jednotek)

Zdroj: ArcČR® 4.1, 2022; vlastní zpracování

### 4.3 Výběr prostorových dat

Vybrány byly dva prostorové fenomény, a sice úhrny nákazy onemocněním COVID-19 od začátku pandemie a výsledky druhého kola českých prezidentských voleb 2023. Jedná se o jevy, které jsou v době vzniku práce ve společnosti aktuálními tématy a mají pro ni značný význam. Záměrem předkládané práce však není objasnit zákonitosti a příčiny jejich distribuce ve zkoumaném území. Zvolené jevy mají sloužit především jako metodický nástroj pro studium účinků agregace prostorových dat na zkruslení jejich vzorů, závislostí a interpretace.

Ke každé ze dvou klíčových statistik byly následně přiřazeny tři pomocné ukazatele pro měření vzájemné korelace v rámci každého ze zájmových regionálních systémů. Bylo nasnadě zvolit takové ukazatele, jejichž vztah s příslušným hlavním jevem musel být takový, aby bylo možné logicky očekávat vzájemnou pozitivní či negativní závislost. Specifikace a bližší popis sledovaných charakteristik bude následovat. Tabulka 2 podává ucelený přehled použitých prostorových informací. Jednotlivým relativním ukazatelům byly přiřazeny zjednodušené názvy, jejich přesné vymezení a výpočet bude rozebrán dále.

Tabulka 2: Přehled zjednodušených názvů ukazatelů pro výpočty korelace

hlavní ukazatele	pomocné ukazatele pro korelaci s hlavním ukazatelem
1) počet případů na 1000 obyv.	2) obyvatelstvo 65+
	3) zaměstnaní ve zdrav./soc. službách
	4) velikost domácnosti
5) podíl hlasů pro zvoleného prezidenta	6) nezaměstnanost
	7) věřící
	8) minimálně maturita

Zdroj: vlastní zpracování; 2023

### 4.4 Získání dat a výpočty

Všechny prostorové informace byly získány z veřejně dostupných zdrojů. Stahovány byly údaje na obecní úrovni, opatřené šestimístními kódy pro jednoznačnou identifikaci konkrétní obce. Operace s daty probíhala většinou v prostředí MS Excel, částečně v programu Statistica. Následuje popis procesu získání jednotlivých charakteristik z tabulky 2 na úrovni obcí. Každá z osmi relativních statistik byla vypočtena z několika výchozích absolutních ukazatelů, které jsou v kapitole 4.4.1 označeny podtržením. Veškeré výpočty byly pro kontrolu provedeny dvakrát. Během druhého pokusu bylo odhaleno zaměnění hodnot tří vstupních proměnných u devíti obcí. Toto pochybení bylo způsobeno chybným přiřazením kódů obcí.

Hodnoty výchozích absolutních ukazatelů, jsou uvedeny ve příloze k práci. Jsou prezentovány v tabulkách ve formě, v jaké z nich byly vypočteny relativní proměnné uvedené v tabulce 2 a ve které

byly agregovány. Jedná se o číselné údaje pro 6258 obcí, včetně čtyř vojenských újezdů, u nichž jsou však hodnoty všech ukazatelů nulové.

#### **4.4.1 Proces získání ukazatelů na obecní úrovni**

##### **1) počet případů na 1000 obyv.**

Tento ukazatel vyjadřuje celkový počet diagnostikovaných případů onemocnění COVID-19 na 1000 obyvatel od začátku pandemie. Byla stažena datová sada Epidemiologická charakteristika obcí z otevřených dat ministerstva zdravotnictví ČR. Jedná se o textovým souborem formátu CSV, který obsahuje epidemiologické údaje pro obce ČR. Z nich byl vybrán ukazatel počty nově diagnostikovaných osob s COVID-19 včetně reinfekcí. Zjevnou nevýhodou těchto dat je zahrnutí reinfekcí. Při vztážení tohoto údaje na počet obyvatel se totiž může stát, že některé osoby jsou v počtu nově diagnostikovaných započteny vícekrát. Hodnota očištěná od reinfekcí však není k dispozici. V popisu datové sady je uvedeno, že jsou evidovány případy nesprávného přiřazení osob ke konkrétním obcím z důvodů nesouladu dat laboratoří a krajských hygienických stanic. Jedná se údajně o menšinu případů, kdy byly některé nově diagnostikované osoby z velkých měst omylem přiřčleněny drobným obcím (Ministerstvo zdravotnictví, 2023).

Epidemiologické charakteristiky jsou v souboru dat strukturované po jednotlivých dnech pro celé období pandemie počínaje dnem 1. 3. 2020. Jelikož takto objemný soubor dat nebylo možné otevřít v programu MS Excel, bylo k jeho načtení využito softwaru Statistica. S využitím filtru došlo k výběru období od 1. 3. 2020 do 31. 12. 2022. Popisné statistiky programu umožnily součet hodnot zvoleného ukazatele pro celé vybrané období tak, abychom dostali za každou obec pouze jeden údaj. Případy nakažených byly evidovány také u dvou ze čtyř vojenských újezdů. Jelikož by nebylo možné vztáhnout zde toto číslo na počet obyvatel, byly hodnoty z těchto území zanedbány. Vynecháno tedy bylo celkem 17 případů nově diagnostikovaných osob a to jak na úrovni obcí tak pro srovnatelnost rovněž na ostatních úrovních členění území.

K přepočtu na 1000 obyvatel byl využit údaj o počtu obyvatel v obci ze sčítání lidu, domů a bytů 2021 (SLDB 2021).

##### **2) obyvatelstvo 65+**

Vyjadřuje podíl obyvatel ve věku 65 a více let. Údaje bylo docíleno podílem absolutního počtu obyvatel ve věku 65 a více let a celkového počtu obyvatel. Oba údaje pochází ze otevřených dat ČSÚ ze SLDB 2021.

##### **3) zaměstnaní ve zdrav./soc. službách**

Ukazatel podílu zaměstnaných osob ve zdravotních či sociálních službách. Jedná se o procentní vyjádření počtu obyvatel zaměstnaných ve zdravotních a sociálních službách z počtu zaměstnaných obyvatel celkem. Tyto ukazatele byly získány ze SLDB 2021.

#### **4) velikost domácnosti**

Vyjádření průměrného počtu členů hospodařících domácností z dat posledního sčítání lidu. Konkrétně jako podíl počtu členů hospodařících domácností celkem a počtu hospodařících domácností.

#### **5) podíl hlasů pro zvoleného prezidenta**

Udává hodnotu podílu platných hlasů pro zvoleného prezidenta Petra Pavla ve druhém kole prezidentských voleb. Zdrojem informací byla veřejná databáze ČSÚ. Požadovaný ukazatel byl vypočten jako podíl počtu platných hlasů pro kandidáta Petra Pavla a celkového počtu platných hlasů.

#### **6) nezaměstnanost**

Ukazatel podíl nezaměstnaných osob ke konci roku 2021 byl vypočten podle Metodiky sledovaných ukazatelů pro územně analytické podklady (2022). Jedná se o podíl počtu dosažitelných uchazečů registrovaných Úřadem práce ČR ke konci roku 2021 a počtu obyvatel ve věku 15 až 64 let ke konci roku 2020. Oba tyto údaje byly získány z dat ČSÚ pro Územně analytické podklady. Pro vyjádření počtu obyvatel v produktivním věku byl záměrně použit ukazatel z roku 2020, jelikož zmíněná metodika používá při výpočtu podílu nezaměstnaných osob ve jmenovateli informaci o rok starší než v čitateli.

#### **7) věřící**

Tato statistika vyjadřuje podíl věřících hlásících i nehlásících se k církvím na celkové populaci podle dat SLDB 2021. Konkrétně jde o podíl součtu věřících hlásících se k církvi, náboženské společnosti nebo směru a věřících nehlásících se k církvi, náboženské společnosti nebo směru na počtu obyvatel obce.

#### **8) minimálně maturita**

Pro výpočet podílu obyvatel minimálně s maturitou bylo nutné získat absolutní počet obyvatel minimálně s maturitou v obci. Z údajů SLDB 2021 o obyvatelstvu dle nejvyššího dosaženého vzdělání bylo tedy nutné sečíst počet obyvatel v kategorii středoškolského vzdělání s maturitou a ve dvou dalších vyšších úrovních vzdělání. Výsledek byl pak vztažen na počet obyvatel ve věku 15 a více let, který byl získán jako součet dvou věkových kategorií zjišťovaných SLDB 2021.

#### **4.4.2 Agregace dat**

K agregaci dat do větších územních celků byly využity kontingenční tabulky programu MS Excel. Data na úrovni obcí byla seskupována podle kódů příslušnosti k jednotlivým administrativním regionům vyšší hierarchické úrovně. V případě funkčního regionálního systému probíhaly součty na dvou úrovních. Nejdříve došlo k agregaci podle elementárních jednotek a poté byla tato data spojena dle kódů funkčních regionů. V kontingenčních tabulkách byly sčítány hodnoty výchozích absolutních údajů (v předchozí podkapitole označeny podtržením). Z nich bylo následně na každé úrovni agregace počítáno osm výsledných statistik týmž způsobem jako u obcí, což je popsáno v předchozí podkapitole.

#### **4.4.3 Výpočet korelace a variability**

Jak již bylo nastíněno, v rámci výzkumu jsou sledovány korelace dvou hlavních jevů s každým z trojice příslušných pomocných ukazatelů. K měření závislosti bylo použito funkce PEARSON v prostředí MS Excel. Jedná se o nástroj, kterým je možno vyjádřit hodnotu Pearsonova korelačního koeficientu pro dvojici souborů statistických dat. Statistická závislost byla posuzována ve všech pěti zájmových typech členění území.

Jak již bylo zmíněno v souvislosti s problémem modifikovatelné územní jednotky, s vyšší úrovní agregace obvykle klesá variabilita souboru. Proto byla ověřena také variabilita osmi relativních ukazatelů ve všech typech agregace. Pro každou proměnnou a pro každý regionální systém byla stanovena hodnota variačního koeficientu. Relativita tohoto ukazatele umožňuje vzájemné porovnání. Výpočet byl proveden rovněž v prostředí MS Excel jakožto procentuální podíl směrodatné odchylky (nástroj SMODCH.P) a aritmetického průměru (funkce PRŮMĚR) dané sady hodnot.

#### **4.5 Práce v prostředí GIS**

Veškeré mapové úlohy byly řešeny v aplikaci ArcMap. Proces získání datových vrstev pro GIS analýzy je uveden v kapitole 4.2 v rámci popisu použitých regionálních systémů. Jedná se o polygonové vrstvy jednotlivých typů členění území.

##### **4.5.1 Kartogramy**

Pomocí nastavení symbolů byly kartograficky znázorněny hodnoty dvou zkoumaných jevů na všech pěti úrovních členění území. Pro tvorbu kartogramů bylo nutné mít příslušné atributy všech pěti použitých polygonových vrstev (pěti systémů územního členění) kompletně vyplněné hodnotami sledovaných jevů. Na úrovni obcí nejsou pro oba sledované fenomény dostupná data pro čtyři polygony, a sice vojenské újezdy, které byly z kartogramu vynechány pomocí příkazu Data Exclusion.

Intervalové stupnice kartogramů bylo zapotřebí stanovit tak, aby bylo možné přehledně porovnat prostorovou distribuci jevů mezi pěti regionálními systémy. Byly hledány takové hraniční hodnoty intervalů, aby byly pro konkrétní mapovaný ukazatel stejné napříč všemi pěti úrovněmi členění území. Požadavek na takovéto rozdělení stupnic je pochopitelně v rozporu s rovnoměrným rozvržením prvků do intervalů. Nejdříve bylo přistoupeno ke stanovení mezních hodnot pomocí kvantilů pro každý regionální systém zvlášť. Následně byly hraniční hodnoty jednotlivých intervalů napříč pěti vrstvami zprůměrovány a zaokrouhleny.

##### **4.5.2 Analýzy prostorové autokorelace**

Nejprve byla hodnocena globální autokorelace dvou hlavních zkoumaných jevů pomocí Moranova I kritéria. Posuzováno bylo všech pět úrovní členění území. Software ArcMap umožňuje výpočet tohoto indikátoru prostřednictvím nástroje „Spatial Autocorrelation (Morans I)“ ze skupiny nástrojů Spatial Statistics Tools, oddělení Analyzing Patterns. Výsledkem této analýzy je číselná hodnota, která udává

míru prostorové autokorelace v rozmezí hodnot od -1,0 do 1,0 ve stejném významu jako Pearsonův korelační koeficient.

V rámci dialogového okna tohoto nástroje je nutné provést několik kroků. Nejprve zadat vstupní vrstvu a vstupní atribut (ukazatel, jehož autokorelaci chceme změřit). Byl proveden výpočet pro dva způsoby konceptualizace prostorových vztahů: INVERSE\_DISTANCE a FIXED\_DISTANCE\_BAND. Dále je v případě druhého způsobu konceptualizace třeba zadat hodnotu parametru „Distance Band or Threshold Distance“. Jedná se o vzdálenost v metrech, na které je shlukování hodnot proměnné maximální. Tato byla získána pomocí výpočtu „Incremental Spatial Autocorrelation“ rovněž ze sady nástrojů Spatial Statistics Tools. Výsledkem této operace je spojnicový graf. Z něj je možné odečíst vzdálenost, na které je shlukování hodnot proměnné vrcholí. Posledním krokem je pak přenastavení standardizace dat na řádkovou (ROW).

K měření lokální prostorové autokorelace v rámci polygonové vrstvy obcí byly využity dvě funkce z nabídky Mapping clusters ze sady nástrojů Spatial Statistics Tools v prostředí ArcMap. Na základě zvoleného atributu dokážou tyto analýzy identifikovat prostorově závislé skupiny polygonů. Nástroj „Cluster and Outlier analysis“ pracuje s lokálním indexem prostorové autokorelace Moranovo I popsaného Anselinem (1995). Nástroj „Hot Spot Analysis“ používá statistiku prostorové asociace  $G_i^*$  popsanou Getisem a Ordem (1992).

Oba nástroje je před spuštěním výpočtu rovněž nutné nastavit prostřednictvím parametrů v dialogovém okně. To se příliš neliší od měření globální prostorové autokorelace. Také u lokálních měř byly testovány dva jmenované způsoby konceptualizace vztahů. Výsledky podle INVERSE\_DISTANCE se však příliš lišily od prostorového rozložení jevů zkoumaných v kartogramech. Proto byly provedeny výpočty pouze s využitím konceptualizace dle FIXED\_DISTANCE\_BAND. Další parametry byly nastaveny shodně jako u výpočtu globální charakteristiky.

Výstupem je kartografické vyjádření lokálních indexů prostorové autokorelace. Výsledné mapy ukazují hot spots (shluky nadprůměrných hodnot), cold spots (shluky podprůměrných hodnot), tedy lokality s pozitivní autokorelací a místa bez významného vztahu mezi hodnotami (not significant). Moranova statistika navíc podává přehled také o negativní prostorové autokorelaci a zobrazuje takzvané spatial outliers (odlehle hodnoty).

## 5 Analýza vlivu agregace na prostorové informace

Výzkum je rozdělen do několika fází. Nejdříve je porovnávána meziregionální variabilita sledovaných jevů v rámci jednotlivých typů členění území. Následně je pozorováno jejich prostorové rozložení v těchto systémech zón. Hodnocena je také spojitost prostorových vzorů, tedy prostorová autokorelace. Nakonec je sledována stabilita vyjádření statistických vztahů mezi dvojicemi proměnných v různých typech agregace

Sledovanými prostorovými jevy jsou celkový počet diagnostikovaných případů onemocnění COVID-19 na 1000 obyvatel pro celé období pandemie a podíl platných hlasů pro zvoleného prezidenta Petra Pavla ve druhém kole prezidentské volby. V rámci výzkumu jsou tyto jevy nazývány zjednodušeně, a sice „covidová“ a „volební“ charakteristika/proměnná/jev.

### 5.1 Variabilita

Zkoumání prostorových vzorů, které jevy vytváří, se logicky pojí se sledováním jejich variability. Tyto prostorové vzory jsou zpravidla různorodé, jak již bylo zmiňováno v souvislosti se základními vlastnostmi geografických informací. Jelikož se míra rozmanitosti fenoménů napříč jednotkami územního členění váže také na další oblasti tohoto výzkumu, je její hodnocení uvedeno přednostně. K tomuto účelu je použit jednoduchý relativní ukazatel variability, konkrétně variační koeficient. Přehled variačního koeficientu dvou hlavních zkoumaných jevů včetně aritmetických průměrů v jednotlivých typech členění státu shrnuje Tabulka 3. Z výpočtů variability na úrovni obcí byly vynechány čtyři vojenské újezdy, ve kterých se nenachází vzorek, ke kterému by bylo možné vztáhnout výpočet dvou sledovaných ukazatelů.

Tabulka jasně ukazuje na pokles variability hodnot ukazatelů se zvyšujícím se stupněm agregace. S agregací se měnil také aritmetický průměr, jenž se až na jednu výjimku zvyšoval a od kterého se výpočet variačního koeficientu odvíjí. Na obecní úrovni dosahuje tento koeficient v případě obou jevů nejvyšších hodnot a dostává se nad 20 %. To lze vysvětlit častějším výskytem extrémních hodnot, tedy nízkých minim a vysokých maxim. Ta jsou pak při výpočtu měř ze sloučených údajů, kde jsou tyto krajní hodnoty seskupeny s jinými, často bližšími průměrným, vyrovnávána. Výsledky v tabulce tedy poukazují na skrývání rozmanitosti prostorových vzorů patrných na nižších hierarchických úrovních.

Připomeňme, že obce, coby nejmenší použité prostorové jednotky, jsou rovněž výsledkem agregace. Při použití ještě drobnější struktury členění území bychom mohli očekávat větší rozmanitost. I přesto by ale musely být použité informace v agregované formě, jak ostatně vyplývá z teoretických poznatků (kapitola 3.5). Aplikace zcela individuálních dat není vhodná pro všechny typy analýz, včetně tohoto výzkumu, neboť je zjišťováno zastoupení daných jevů na určitém vzorku.

U covidové charakteristiky pozorujeme výraznější shlázení rozdílů pod vlivem slučování dat. Na nejnižším stupni agregace je variabilita počtu případů na 1000 obyv. vyšší, než jakou sledujeme u



druhého jevu. Hned v rámci následující hierarchické úrovně ale výrazně klesá, přičemž výsledky prezidentských voleb zůstávají rozmanitější. Ještě ve struktuře čtrnácti krajů je hodnota variačního koeficientu výsledků voleb pouze o 1,5 procentního bodu nižší než je tomu v případě 206 správních obvodů obcí s rozšířenou působností. Nejvýznamněji se ukazatel variability obou jevů změnil při přechodu z členění území na obce k SO ORP. To lze z pohledu problému měřítka přisuzovat skutečnosti, že mezi těmito dvěma stupni dochází k nejzásadnější redukci počtu územních jednotek. Zmíňme také funkční regiony, které v případě volebních výsledků vybočují z trendu změny koeficientu s úrovní agregace, který vidíme u administrativních jednotek. Rozdíly mezi funkčními regiony jsou menší než v rámci okresů, které představují vyšší hierarchickou úroveň členění.

Co se týká covidu a obcí, byly pozorovány výrazné nadprůměrné hodnoty, kdy devět obcí vykazovalo hodnotu nad 1000. Nutno podotknout, že všechna tato území čítají méně než 400 obyvatel. Tyto vysoké počty mohou být pouze důsledkem výskytu reinfekcí nebo také zmíněné metodické chyby zdrojových dat, kdy případy z velkých měst byly nesprávně přiřazeny k populačně malým obcím. V těchto obcích je také možné předvídat působení efektu malých čísel, tedy že počet diagnóz (včetně reinfekcí) vztažený na nepatrný vzorek populace se jeví jako nadprůměrný.

Problém malých čísel se mohl do určité míry týkat také volebních výsledků zvoleného prezidenta na obecní úrovni. Mezi sedmnácti jednotkami, ve kterých byl podíl hlasů vyšší než 85,0 %, se nachází pět s celkovým počtem hlasů nižším než 100. Spíše lze ale za nejvyššími podíly vítězných hlasů vidět jiné příčiny, které nesouvisí s formami agregace dat a jsou vhodnějším předmětem zkoumání pro volební geografii. Statistické rozložení se na obecní úrovni u podílu hlasů pro zvoleného prezidenta v porovnání s covidovou charakteristikou více blíží vzpomínanému Korčákovu rozdělení. Sledujeme tedy hojný výskyt obcí s nízkými hodnotami, čemuž odpovídá také nízký aritmetický průměr tohoto ukazatele pro tento typ členění území.

Tabulka 3: Variační koeficient a aritmetický průměr hl. ukazatelů v různých typech členění území

území	variační koeficient [%]		aritmetický průměr	
	počet případů na 1000 obyv.	podíl hlasů pro zvoleného prezidenta	počet případů na 1000 obyv.	podíl hlasů pro zvoleného prezidenta
OBCE	23,6	21,2	400	54,0
SO ORP	9,1	14,5	425	54,3
FUNKČNÍ REG.	8,0	13,3	426	53,9
OKRESY	7,6	13,8	426	55,2
KRAJE	6,4	13,0	430	56,9

Zdroj: ČSÚ, Ministerstvo zdravotnictví ČR; 2023

## 5.2 Prostorová distribuce

Tento oddíl se zaměřuje na jeden z dílčích cílů práce, a sice analýzu prostorové distribuce dvou základních proměnných. Ta je porovnávána napříč pěti zájmovými formami agregace dat. Prostorové rozložení hodnot jmenovaných jevů je sledováno v kartogramech, jež jsou vytvořeny zvláště pro každý typ členění území.

V teoretických základech práce bylo upozorněno na skutečnost, že způsob členění území, v němž je variabilita ukazatele maximální, je pro sledování prostorové distribuce nejvhodnější. Jak vyplývá z předchozí podkapitoly, nejvyšší rozmanitost zájmových proměnných se projevila ve struktuře obcí, tedy na nejnižší hierarchické úrovni. Z tohoto důvodu je zvláštní pozornost při popisu a komparaci jednotlivých typů členění věnována obecní úrovni. Již při hodnocení variability se projevilo působení problému modifikovatelné územní jednotky. Nahlédnutí na prostorové rozložení geografických informací při různých formách jejich seskupení umožní posouzení účinků agregace z jiné perspektivy.

Mapy na obrázku 8 vyjadřují počet diagnostikovaných případů onemocnění COVID-19 (včetně reinfekcí) na 1000 obyvatel dané územní jednotky. Následující obrázek 9 je kartografickým vyjádřením druhého z hlavních sledovaných jevů. Z uvedené intervalové stupnice lze vyčíst, v jakém rozmezí se pohybuje podíl z platných hlasů odevzdaných v dané územní jednotce ve druhém kole prezidentských voleb připadající zvolenému prezidentovi.

Určitá variabilita prostorového rozložení obou jevů je patrná na první pohled ve všech pěti typech členění území. Zároveň je evidentní, že prostorová distribuce ukazatelů není zcela náhodná. V rámci jednotlivých systémů členění státu je více či méně patrná podobnost hodnot mezi sousedícími územními celky. Jak již bylo předesíláno, projevy variability či spojitosti prostorových vzorů jsou limitovány formou agregace. V rámci tohoto výzkumu v tomto ohledu pozorujeme především vliv hierarchické úrovně, neboli efekt měřítka.

Zaměřme se také na samotný rozbor prostorové distribuce konkrétních sledovaných fenoménů ve sledovaném území, jež je nejvíce patrná z obecní struktury a do jisté míry také z dalších úrovní agregace. V nejpodrobnější formě zobrazení dat je u covidové charakteristiky možné najít výraznou převahu jednotek zbarvených dle nejnižšího a nejvyššího intervalu (viz obrázek 8). To souvisí s četnějším výskytem extrémních pozorování, která byla popsána při zkoumání variability souborů. Mezi oblastí s vyššími hodnotami covidové proměnné můžeme obecně zařadit severovýchod Čech, dále pás sahající od zázemí Plzně po zázemí Prahy, oblast střední a východní Moravy a také severovýchod Moravskoslezského kraje. Poněkud rozdrobeně se vyšší míry nachází také v Jihočeském kraji a v severních Čechách. Naopak nízké úhrny případů na 1000 obyvatel se hojně vyskytují v západních Čechách, na jihu Jihočeského kraje, jihozápadě Moravy a významně také v severních částech Olomouckého a Moravskoslezského kraje.

Co se týká volebních výsledků, můžeme z obrázku 9 také rozpoznat hojně zastoupení obcí zbarvených barvami nejnižšího a nejvyššího intervalu na rozdíl od ostatních regionálních systémů. Ve srovnání s prvním studovaným jevem pozorujeme poněkud výraznější prostorové koncentrace podobných hodnot na obecní úrovni. Na základě této skutečnosti můžeme předpokládat vyšší míru prostorové autokorelace výsledků prezidentských voleb.

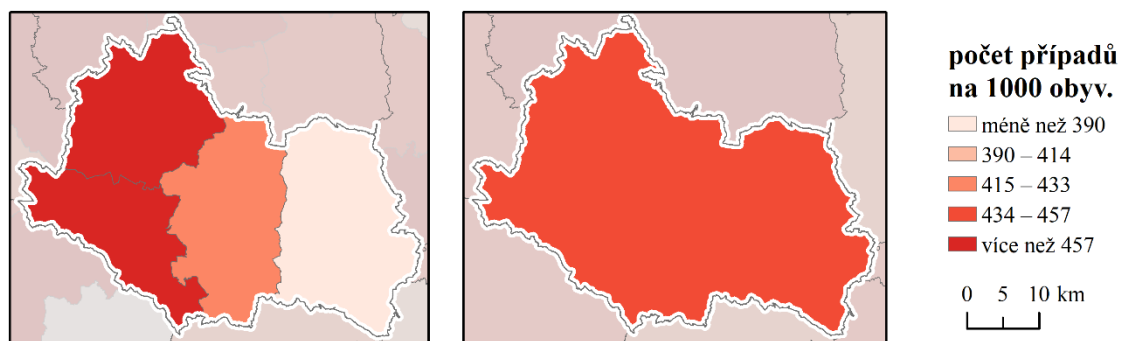
Vysoké hodnoty volebního jevu identifikujeme opět na severovýchodě Čech a v širším zázemí Prahy, respektive ve Středočeském kraji. Dobře rozpoznatelné jsou také koncentrace nadprůměrných podílů v okolí Plzně, Českých Budějovic a Brna, vždy včetně těchto měst. O něco méně soustředěné jsou vysoké podíly hlasů vítězného kandidáta v oblasti kraje Vysočina a na východě Moravy. Zajímavé jsou také výsledky v příhraničních horských obcích na jihozápadě a severozápadě Čech, které vykazují vyšší hodnoty než přilehlé jednotky. Nejvýraznější oblast s nízkými hodnotami se táhne při severozápadním až západním okraji Čech. Další se přibližně shoduje s územím okresu Znojmo. Nutno zmínit také území podprůměrných podílů hlasů pro zvoleného prezidenta v oblasti severní Moravy s přesahem směrem na jihozápad s výjimkou Olomoucka.

V regionálních systémech vyšší hierarchické úrovně než obecní je zastoupení jednotek patřících do intervalů blízkých se průměrné hodnotě podstatně vyšší. Platí to pro obě zkoumané proměnné a je evidentní, že příčinou je již několikrát zmiňovaná nižší variabilita těchto regionálních systémů. Je možné se domnívat, že zachování výraznějších meziregionálních rozdílů volebních výsledků souvisí s vyšší prostorovou koncentrací podobných hodnot tohoto jevu. Například oblasti s podprůměrnými obecními hodnotami na severozápadě Čech poměrně dobře kopírují vymezení Ústeckého a Karlovarského kraje. Oblast Středočeského kraje je naopak poměrně homogenně pokryta obcemi s vyššími výsledky podílů. Z uvedených skutečností tak vyplývá, že čím vyšší je vnitřní homogenita agregovaných jednotek, tím je systém odolnější vůči potlačení zjevnosti prostorových vzorů. Územní koncentraci podobných hodnot ale nelze jednoznačně považovat za výhodu. Při odlišném vymezení hranic prostorových zón existuje riziko rozčlenění shluků podobných hodnot mezi více jednotek a tím pádem skrytí jejich existence.

Nejvýmluvněji lze ilustrovat zastření prostorových vzorů pomocí komparace nejvyšší a nejnižší hierarchické úrovně členění. Předpokládejme, že uživatel je konfrontován pouze s mapami pro soustavu čtrnácti krajů, které vidíme na obrázcích 8 a 9. Při jejich interpretaci je odkázán na předpoklad platnosti hodnoty proměnné v daném kraji pro všechna místa uvnitř něj. I zcela nezasvěcený uživatel by však při následném porovnání s nejdetailejším způsobem členění připustil scestnost tohoto předpokladu. Výskyt obcí či shluků obcí s podprůměrnými (či nadprůměrnými) hodnotami uvnitř krajů s nadprůměrnými (či podprůměrnými) hodnotami jsou na obou obrázcích zjevné díky všudypřítomné prostorové variabilitě a koncentraci.

Analýza se nyní zaměří na několik lokalit, kde lze poměrně dobře identifikovat prostorovou neurčitost, respektive výskyt agregčních nepřesností v prostorových vzorech. Nejprve se budeme věnovat

obrázku 8, tedy covidovému ukazateli. Jmenované propojení Plzně a Prahy územím nadprůměrných hodnot odpovídá vymezení funkčních regionů. Ve zbylých třech regionálních systémech je koncentrace vysokých hodnot mezi městy méně patrná. Také zmiňované rozdrobené nadprůměrné covidové míry v Jihočeském kraji jsou mezi nejvyšší počítány pouze na území SO ORP Písek. Správní obvod Vyškov patří do nejvyššího intervalu navzdory skutečnosti, že téměř 30 % jeho rozlohy tvoří vojenský újezd Březina s nulovou hodnotou sledované míry. Právě zahrnutí vojenského újezdu však snižuje vzorek populace pro výpočet dané míry a napomáhá tak projevu efektu malých čísel. Zařazení do nejvyššího intervalu sledujeme ve funkčním regionu města Náchod, kdy je tomuto území po sloučení s oblastí Broumova ve struktuře okresů přidělen nižší rozsah hodnot. K redukci nadprůměrných hodnot dochází též uvnitř okresu Písek, ve kterém „zaniká“ správní obvod i funkční region města Hranice. Okres Svitavy je ukázkovým příkladem zastření variability pozorovatelné na nižších hierarchických úrovních (viz výřez na obrázku 7). Naproti tomu okres Bruntál dobře zachycuje oblast s podprůměrnými počty případů na 1000 obyv. Ta je v rámci ostatních forem členění nadhodnocována, zejména v případě SO ORP i funkčního regionu Krnov. Za příklad agregačního zkreslení na krajské úrovni uveďme Olomoucký kraj, ve kterém jsou podprůměrné hodnoty na severu a průměrné až nadprůměrné hodnoty na jihu reprezentovány rozpětím hodnot prostředního intervalu.



Obrázek 7: Distribuce covidové charakteristiky v rámci okresu Svitavy a jeho SO ORP

Zdroj: ČSÚ, Ministerstvo zdravotnictví ČR, 2023; ArcČR® 4.1, 2022; vlastní zpracování

Další hodnocení se bude věnovat fenoménu prezidentských voleb a bude se vztahovat k obrázku 9. Nej přesněji variabilitu sledovanou na obecní úrovni zachycuje systém 206 správních obvodů ORP, ovšem i v něm jsou patrné účinky agregačních efektů. Toto tvrzení platí i pro první komentovaný jev na obrázku 8. Věrné zachycení nízkých podílů hlasů pro zvoleného prezidenta podává správní obvod Frýdlant, přičemž od úrovně funkčních regionů je tato oblast výrazně nadhodnocována. Podobně je tomu také v případě SO ORP Nový Bydžov, který jediný řadí do nejnižšího intervalu oblast podprůměrných hodnot severovýchodně od Prahy. Dříve uvedená oblast koncentrace vysokých hodnot v Plzni a okolí je v okresní struktuře poněkud podhodnocena díky existenci okresů Plzeň-sever a Plzeň-jih. Naproti tomu shluk nadprůměrných podílů v aglomeraci Českých Budějovic se v rámci okresního a funkčního

regionálního systému promítá dále na jihovýchod až ke státním hranicím. U seskupení obcí s nadprůměrnými hodnotami jevu zasahujícího do okresů Šumperk a Bruntál pozorujeme redukcí vysokých hodnot v případě všech regionálních systémů. Poněkud nespolehlivá by mohla být interpretace výsledků Jihomoravského kraje, jehož výrazná rozmanitost je utvářena především koncentrací vysokých hodnot do Brna a jeho širšího zázemí, což kontrastuje se zmiňovanými nízkými podíly na Znojemsku.

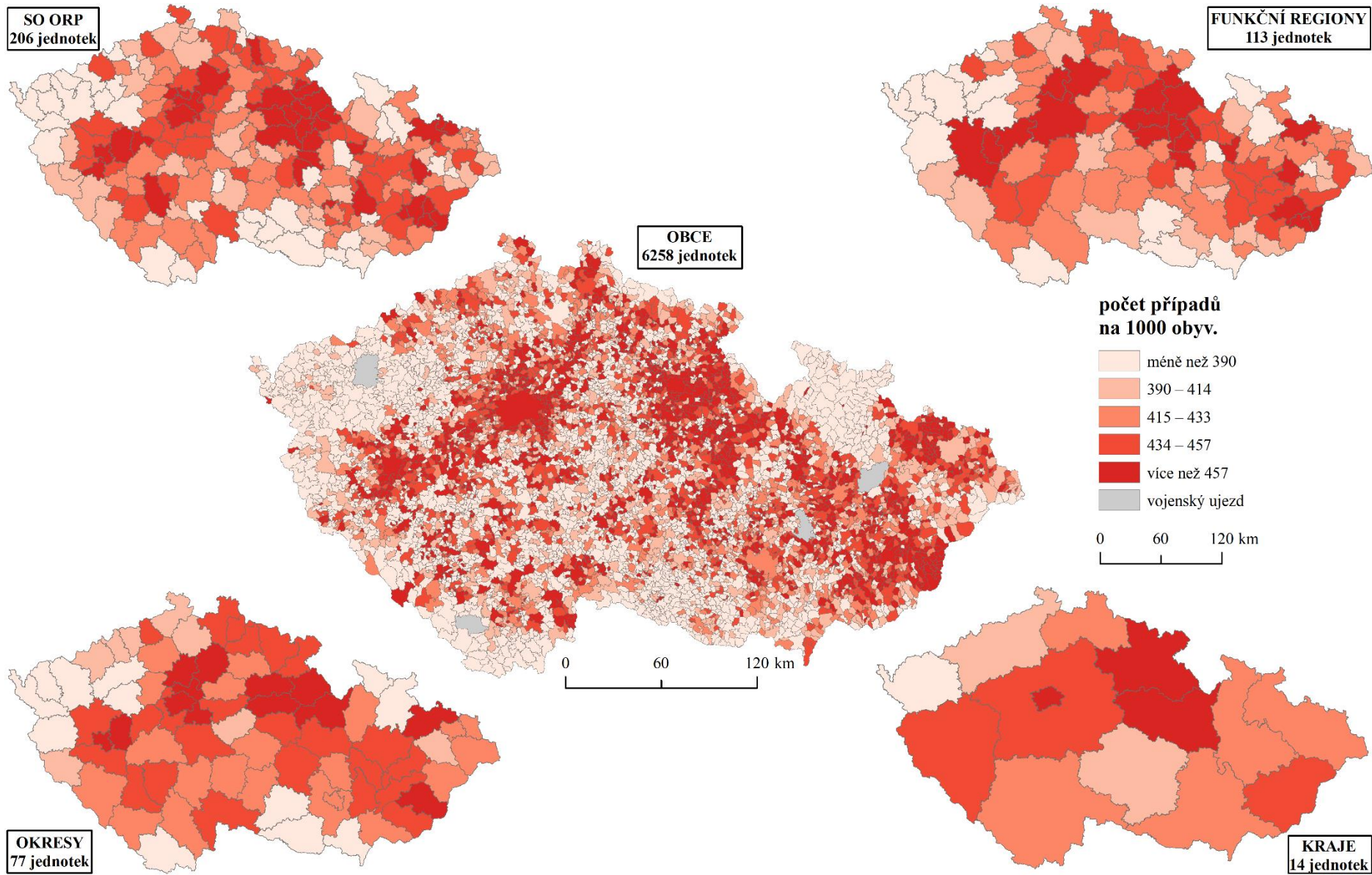
Následují obrázky 8 a 9, jejichž označení je z důvodů formátování uvedeno zde:

Obrázek 8: Prostorová distribuce covidové charakteristiky v pěti typech členění území (červená)

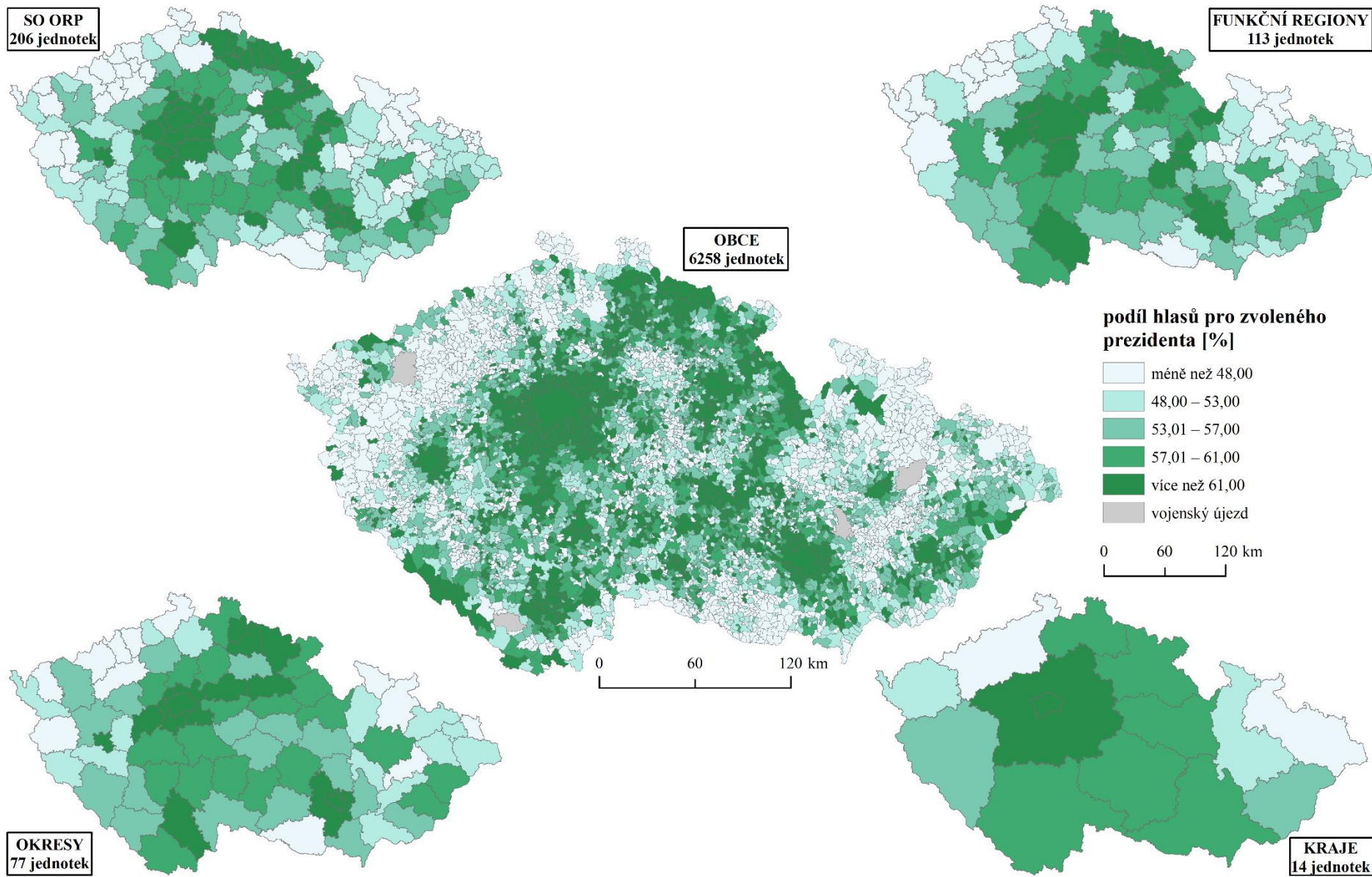
Zdroj: ČSÚ, Ministerstvo zdravotnictví ČR, 2023; ArcČR® 4.1, 2022; vlastní zpracování

Obrázek 9: Prostorová distribuce volební charakteristiky v pěti typech členění území (zelená)

Zdroj: ČSÚ, 2023; ArcČR® 4.1, 2022; vlastní zpracování







### 5.3 Prostorová závislost

Tento úsek výzkumu se zaměřuje na druhý z dílčích záměrů předkládané práce. Jedná se o měření prostorové závislosti studovaných jevů. Přítomnost závislosti blízkých pozorování proměnné může být zdrojem nebo alespoň zesilovacím efektem agregacího zkreslení, jehož analýza je ústředním cílem práce. Posuzována je globální míra prostorové autokorelace dvou základních ukazatelů v jednotlivých typech agregace dat. Využito je globální Moranovo I kritérium. Pro rozkrytí lokálních vzorů prostorové asociace je v rámci obecního členění zjišťována také lokální míra prostorové autokorelace. Hodnocena je pomocí dvou nástrojů k měření této charakteristiky. Metodika měření je uvedena v kapitole 4.5.

Existence prostorové závislosti v geografických informacích pramení z kontinuity prostoru, ale je spojena také s variabilitou prostoru. Dříve zkoumaný variační koeficient měřil statistickou rozmanitost regionálního systému. Jeho výsledky však nevypovídají nic o koncentraci či disperzi jevu v území. Očekává se, že globální indikátor prostorové autokorelace pomůže detailněji vystihnout prostorovou variabilitu sledovaných proměnných. Oproti pozorování prostorových distribucí na mapách v předchozí části podává Moranovo I kritérium jiný pohled na prostorové vzory a dokáže kvantifikovat intenzitu shlukování podobných hodnot. Číselné vyjádření umožní srozumitelnější porovnání míry shlukování podobných hodnot daného jevu mezi jednotlivými typy agregace.

V tabulce 4 pozorujeme míru globální prostorové autokorelace dvou hlavních jevů v pěti typech územního členění. Ve všech zkoumaných systémech členění území oba jevy vykazují pozitivní prostorovou autokorelaci (kladné hodnoty Moranova indexu). Dochází tedy k územnímu shlukování podobných (podprůměrných či nadprůměrných) hodnot, které je pro většinu prostorových fenoménů typické. Výše koncentrace podobných výsledků proměnných se napříč jednotlivými regionálními systémy poměrně výrazně liší.

Je patrné, že charakter změny intenzity prostorového shlukování s rostoucí úrovní agregace je odlišný od trendu změny meziregionální variability. Zatímco stupeň regionální diferenciacie s rostoucí hierarchickou úrovní postupně klesal, míra prostorového shlukování obou fenoménů je nejnižší ve struktuře krajů a obcí. Významněji jsou jevy koncentrovány v případě SO ORP, funkčních regionů a okresů s nejvyšší intenzitou v posledním jmenovaném regionálním systému. Zaměříme-li se na tyto tři prostřední hierarchické úrovně územního členění, pozorujeme v případě obou proměnných nižší výsledky Moranova I kritéria v rámci funkčních regionů.

Nelze jednoznačně posoudit, který z pěti systémů jednotek je náchylnější k agregacím omylům působením prostorové autokorelace. Ta má totiž vliv zejména na variace vymezení regionů na stejné měřítkové úrovni (problém zonace v rámci MAUP). Předkládaná práce však hodnotí územní členění různé hierarchické úrovně a věnuje se tak spíše problému měřítka. Vzhledem k nejbližšímu počtu regionů by bylo možné zvážit porovnání funkčních regionů (113 jednotek) a okresů (77 jednotek) jakožto podobně podrobných forem agregace dat. Díky nižší intenzitě prostorové autokorelace



studovaných jevů na úrovni funkčních regionů je zde možné očekávat menší riziko zkreslení prostorových vzorů sledovaných charakteristik než v rámci okresů.

Na všech úrovních agregace byla potvrzena vyšší míra prostorové asociace u volební charakteristiky očekávaná na základě empirických poznatků z předchozího oddílu. Ve třech prostředních regionálních systémech z hlediska hierarchické úrovně vykazuje volební ukazatel středně silnou prostorovou závislost, jež se v rámci okresů blíží k silné pozitivní autokorelaci. Zatímco prostorová závislost covidového ukazatele je v rámci obcí a krajů téměř nulová. V ostatních formách územního členění pozorujeme mírně či středně významnou kladnou autokorelaci.

Tabulka 4: Globální prostorová autokorelace dvou hlavních jevů na obecní úrovni

území	globální Moranovo I	
	počet případů na 1000 obyv.	podíl hlasů pro zvoleného prezidenta [%]
OBCE	0,099	0,264
SO ORP	0,343	0,417
FUNKČNÍ REG.	0,282	0,366
OKRESY	0,394	0,458
KRAJE	0,072	0,129

Zdroj: ČSÚ, Ministerstvo zdravotnictví ČR; 2023; vlastní zpracování

Lokální indikátory prostorové autokorelace (LISA) jsou nástrojem pro detailnější analýzu prostorové složky variability ukazatelů. Díky nim je možné identifikovat konkrétní oblasti prostorového shlukování podobných hodnot jevů a zobrazit je na mapě. Výpočty LISA jsou na základě teoretických poznatků aplikovány pouze v rámci obcí, kde se očekávají jejich nej přesnější výsledky. Dosavadní zjištění výzkumu ukazují na vliv sousedství v případě obou sledovaných charakteristik. Z map distribuce jevů byly patrné územní shluky podobných hodnot a globální Moranův index prokázal autokorelaci obou jevů v případě všech typů členění. Na základě toho lze také v rámci lokální analýzy očekávat výskyt tzv. hot spots a cold spots.

Obrázky 10 a 11 zachycují výsledky dvou typů lokálních měř prostorové autokorelace hodnocených prostorových informací. Jedná se o dva nástroje představené v teoretických základech výzkumu. Konkrétně o častěji používanou lokální modifikaci Moranova I kritéria (nahore) a lokální  $G_i^*$  míru prostorové závislosti (dole). Lokální Moranovo I zobrazuje oblasti s pozitivní autokorelací (hot spots a cold spots) a zároveň místa s negativní autokorelací (spatial outliers). V rámci prostorových odchylek je rozlišován také příslušný kvadrant Moranova diagramu, tedy zdali se jedná o nadprůměrné hodnoty obklopené podprůměrnými či obráceně. Vytyčována jsou také místa s nulovou autokorelací (not significant), ve kterých není pozorován vztah mezi blízkými pozorováními proměnných. Druhý z lokálních indikátorů autokorelace postihuje pouze pozitivní prostorovou autokorelaci (hot spots a cold

spots) a nulovou autokorelací (not significant), ale navíc je uvedena také míra spolehlivosti zjištěných prostorových vztahů.

V souladu s globálním hodnocením potvrzují také LISA prostorovou koncentraci podobných hodnot v případě prověřovaných ukazatelů. Co se týká dvojice použitých metod, ty ve výsledcích projevují výraznou shodu, alespoň pokud jde o vytyčení prostorově závislých oblastí. Lokální Moranova statistika v rámci shluků identifikuje ještě prostorově odlehle hodnoty. Díky tomu je prostorová distribuce vystižena podrobněji. Naproti tomu lokální míra  $G_i^*$  má tendenci více prostorovou distribuci zobecnit. Zjištěné lokální vzory autokorelace více či méně korespondují s rozložením jevů na obecní úrovni pozorovaným v kartogramech v části 5.2. Následuje komparace výsledků LISA s prostorovými vzory ve zmíněných kartogramech.

Pokud porovnáme prostorovou závislost dvojice jevů, výraznější shlukování vykazuje volební charakteristika (obrázek 11), což opět souhlasí s globálními výsledky. V případě covidového ukazatele (obrázek 10) zase registrujeme více oblastí s negativní autokorelací. To ukazuje na vyšší prostorovou variabilitu jevu, přičemž tato proměnná ukázala také větší statistickou meziobecní variabilitu. Pozoruhodné je, že identifikujeme několik zón, v nichž dochází k výskytu hot spots či cold spots shodně pro oba jevy. Nadprůměrné hodnoty počtu případů na 1000 obyvatel i podílu hlasů pro zvoleného prezidenta se souhlasně koncentrují v Praze a přilehlých oblastech, na severovýchodě Čech a v menší míře na východě Moravy. Uskupení nízkých hodnot proměnných se překrývají na severu Moravy (Jesenicko, Buntálsko) a na jihozápadě Moravy, v západních Čechách a částečně na východě Středočeského kraje. V rámci obou jevů se částečně shodují také místa, kde indikátory LISA identifikovaly vyšší náhodnost, tedy nulovou autokorelací. Jedná se o dvě zóny vinoucí se přibližně severojižním směrem, první mezi Klatovskem a Litoměřickem a druhá od Břeclavska po Šumpersko.

Pokud hovoříme o covidové proměnné (obrázek 10), mapované oblasti hot spots a cold spots odpovídají výraznějším výskytům podprůměrných a nadprůměrných hodnot zdůrazňovaných při popisu distribuce jevu v kartogramu. Výsledky lokální analýzy autokorelace však rozmístění krajních hodnot znázorňují zřetelněji. Ve spojitosti s tímto zmiňme také v kartogramech komentované rozdrobené vyšší míry covidové charakteristiky. Ty v oblasti Jihočeského kraje se v rámci Moranovy lokální analýzy promítly jako prostorové odchylky typu vysoká – nízká. Tedy nadprůměrné hodnoty v blízkosti podprůměrných. Ty na severu Čech jsou zachyceny jako izolované hot spots obklopené prostorovými odchylkami nízká – vysoká, nebo jsou tato seskupení považována za nevýznamná. Spojitost horké skvrny mezi zázemím Plzně a Prahy je přerušena, což bylo zjevné v rámci okresů a regionálního systému SO ORP.

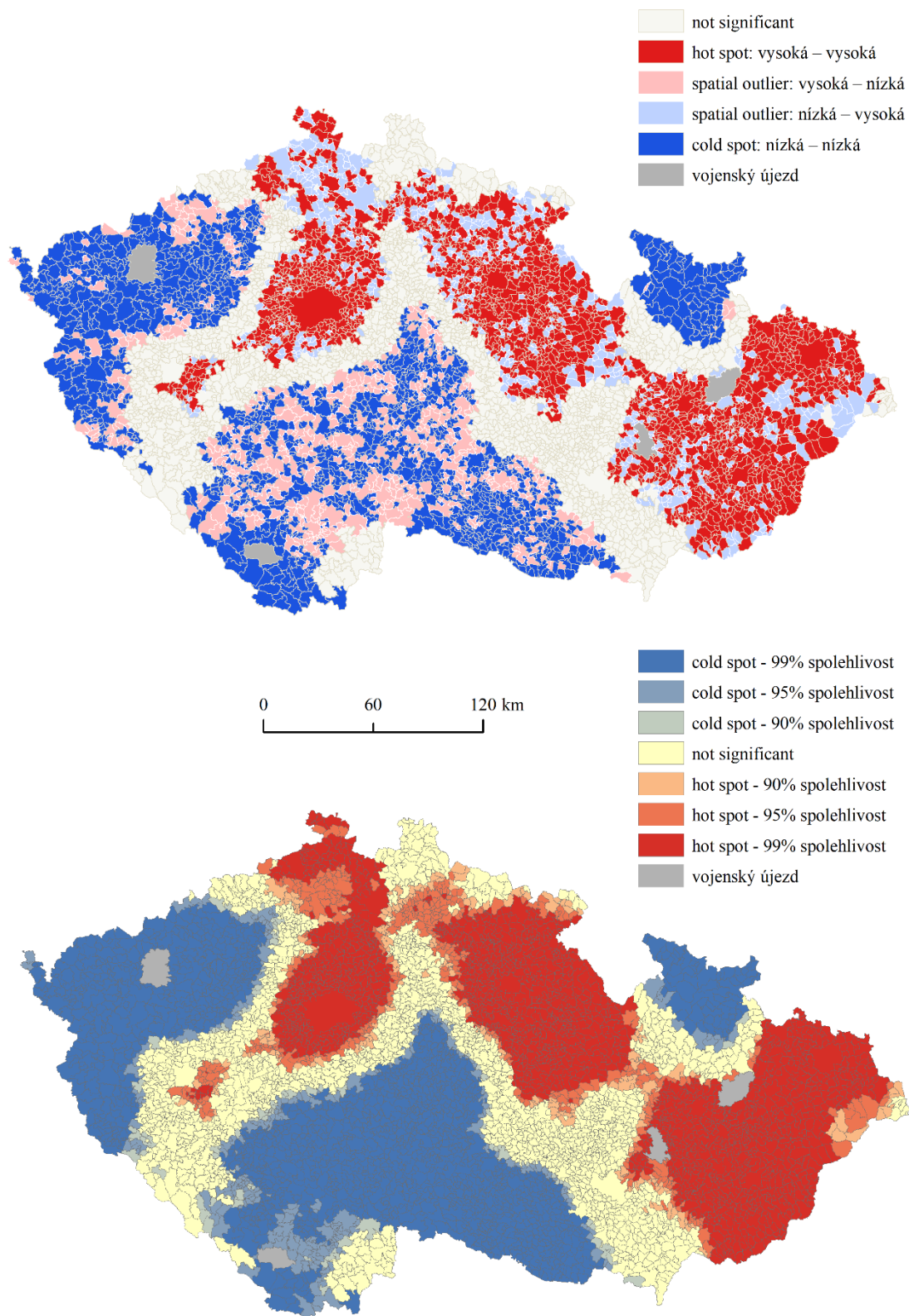
Volební podpora zvoleného prezidenta vykazuje souvislejší územní koncentraci a totéž platí pro obce s jeho nízkou podporou. Klasifikace oblastí hot spots a cold spots opět dochází ke shodě s většinou koncentrací dříve pozorovaných nízkých a vysokých podílů (obrázek 9). V rámci lokální autokorelační analýzy jsou však mnohé tyto shluky prezentovány jako velmi výrazně spojitě. Za hot spots jsou

pokládány také lokality s dobře patrnými shromážděními hodnot nejvyššího intervalu v dříve jmenovaných větších městech a jejich okolí. Výjimkou je Plzeň a zázemí, které je v rámci  $G_i^*$  statistiky pohlceny do západočeské zóny podprůměrných hodnot a Moranovým ukazatelem zachyceny alespoň jako prostorová odchylka vysokých hodnot v obklopení nízkých. Stejným způsobem indikátory LISA naložily s dříve jmenovanou izolovanou oblastí vyšší podpory zvoleného prezidenta na území okresů Šumperk a Bruntál, jejíž shlazení jsme pozorovali z agregovaných údajů.

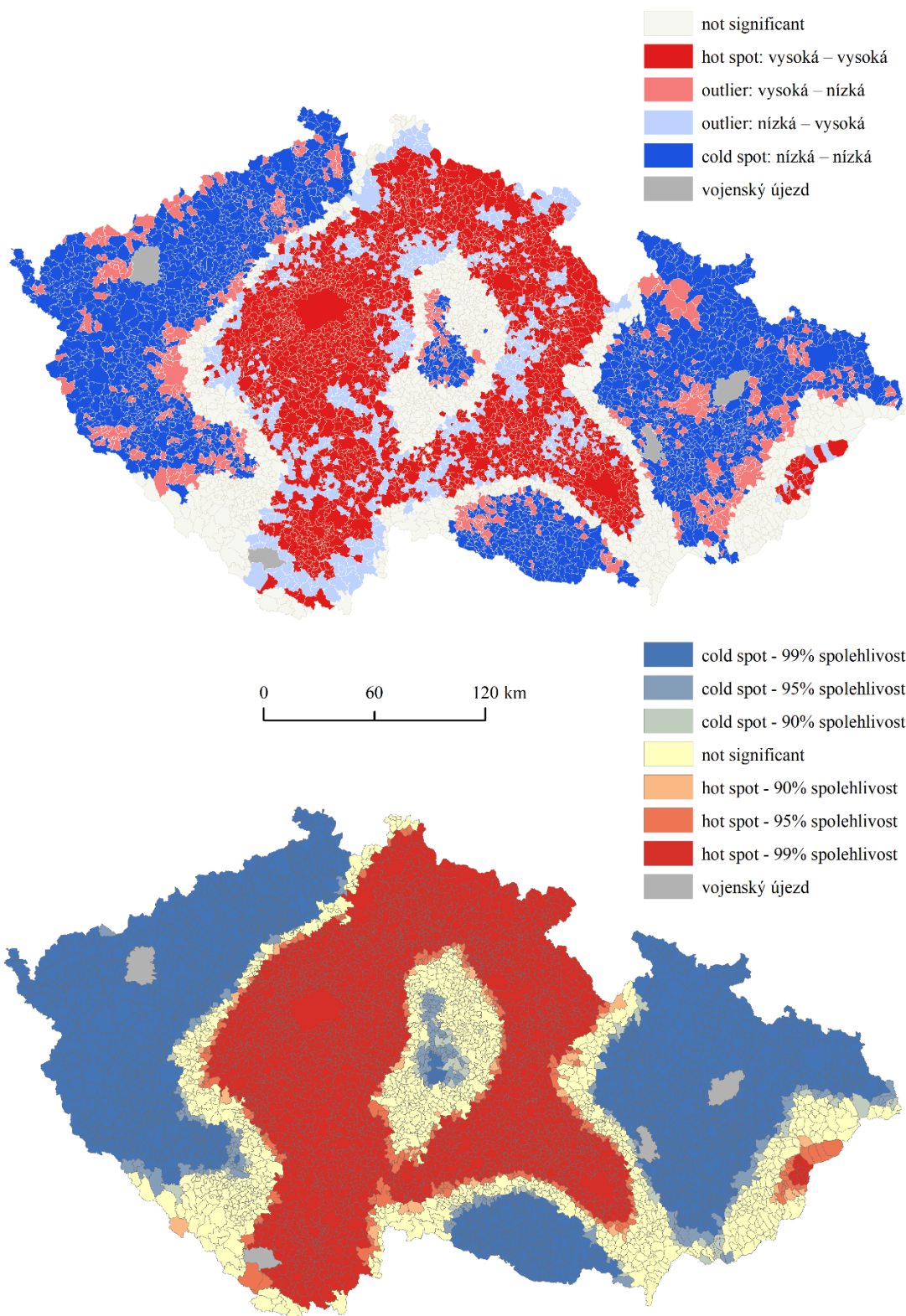
Mezi shluky podprůměrných hodnot včleňují lokální indikátory také východní rozhraní Středočeského kraje. Tato chladná skvrna je v rámci některých regionů v kartogramech mírně patrná. Lépe je v územním členění rozpoznatelná nedaleká dříve zmiňovaná izolovaná zóna podprůměrných hodnot na severovýchod od Prahy projevující se např. v SO ORP Nový Bydžov. LISA ji ale hodnotí zčásti jako prostorovou odchylku a zčásti jako neautokorelovanou oblast. Nakonec zmiňme ještě českobudějovický funkční region a okres, u kterých bylo upozorněno na nadhodnocení jeho jihovýchodní části, kterou lokální Moranova analýza označuje za prostorovou odchylku nízkých hodnot v blízkosti horké skvrny.

Výzkum prostorové autokorelace prokázal její přítomnost (jak negativní tak pozitivní) při hodnocení obou ukazatelů. Konkretizoval také jisté aspekty územního rozložení jevů na obecní úrovni, zejména prostorovou variabilitu. Výsledky globálního indexu dokládají, že meziregionální variabilita (měřená variačním koeficientem) vypovídá poměrně málo o prostorové proměnlivosti geografických informací. Pokud jde o hodnocené proměnné, volební podpora zvoleného prezidenta je prostorově závislejší. Přestože rozšíření onemocnění se ukázalo být do jisté míry také územně koncentrovaným, u volební preference jakožto společensky podmíněnějšího jevu má prostorová blízkost větší účinek.

Z objevených různorodých lokálních vzorů autokorelace logicky vyplývá problematičnost jejich zachycení při variacích vymezení reprezentačních prostorových jednotek. Hodnoceným regionálním systémům se daří územní shluky zachycovat různě a na některá problematická místa bylo upozorněno. Důležité je zejména územní vymezení zón stejné měřítkové úrovně ve spojitosti s různou prostorovou konfigurací shluků, což v rámci tohoto výzkumu nelze příliš posoudit. Určitou roli však může hrát také hierarchická úroveň (velikost jednotek) vzhledem k velikosti shluků, čehož se dotýká případ okresu a funkčního regionu České Budějovice a volebního ukazatele.



Obrázek 10: Lokální prostorová autokorelace covidové charakteristiky na obecní úrovni  
 Zdroj: ČSÚ, Ministerstvo zdravotnictví ČR, 2023; ArcČR® 4.1, 2022; vlastní zpracování



Obrázek 11: Lokální prostorová autokorelace volební charakteristiky na obecní úrovni  
 ČSÚ, 2023; ArcČR® 4.1, 2022; vlastní zpracování

## 5.4 Korelační koeficient

Závěrečná část výzkumu se zaměřuje na hlavní cíl výzkumu, a sice odhalení a popis agregáčnických účinků na prostorové vzory. Zkoumání agregáčnických účinků bude probíhat na sledování korelačního koeficientu. Tento koeficient je snadno interpretovatelný a výhodou je jeho relativnost. Pozorována je závislost zkoumaných jevů s trojicí pomocných ukazatelů a to v rámci každého z pěti forem členění území. Důležitá je komparace výsledků korelační analýzy mezi jednotlivými typy agregace dat. Budeme sledovat, zdali se závislost proměnných měřená v různých regionálních systémech bude lišit od výsledků na nejpodrobnější úrovni členění. V takovém případě je jisté, že zkoumané formy vymezení prostorových jednotek zkreslují výsledky prostorových analýz. Pro zjištění okolností výběru a získání pomocných ukazatelů a výpočtů závislosti dat viz metodiku práce (4. kapitola).

Předcházející části výzkumu poukázaly na skutečnosti, na základě kterých je očekáváno ovlivnění měřených korelačních koeficientů agregáčnickými efekty. V souvislosti se změnou typu agregace byl prokázán pokles meziregionální variability dvojice základních proměnných, shlazení územních rozdílů a nerovnoměrná distribuce byly pozorovány také v kartogramech a autokorelační míry doložily prostorovou koncentraci jevů.

Jelikož se k výpočtu korelačního koeficientu vztahuje intenzita meziregionální variability, jsou uvedeny také hodnoty variačního koeficientu všech proměnných pro pět typů systémů členění. Je posuzován vztah výsledků korelační analýzy k variabilitě proměnných v daných regionálních systémech. Pokud jde o počet případů onemocnění na 1000 obyv., je zkoumána míra jeho závislosti s podílem osob ve věku 65 a více let, dále s podílem zaměstnaných ve zdravotních a sociálních službách a s průměrným počtem členů hospodařící domácnosti. U druhého ze sledovaných jevů bude měřena závislost s podílem nezaměstnaných osob, s procentním zastoupením věřících obyvatel a s podílem obyvatel minimálně s maturitou.

Podíváme-li se na výsledky korelačních koeficientů v grafech (obrázky 12 a 14), zjišťujeme, že se napříč typy agregace různými způsoby mění. Ve většině případů dochází k zachování povahy závislosti, avšak u covidového jevu pozorujeme jeden případ změny z kladného vztahu na záporný. Dále se budeme věnovat detailnímu hodnocení zjištěných korelací se zaměřením na komparaci závislostí napříč systémy členění území a vztahem s variačním koeficientem. Důležitá je odlišnost hodnoty koeficientu korelace v regionálních systémech od jeho výsledku v rámci obecního členění. V případě obcí jevy vykazují nejvyšší variabilitu a závislosti zjištěné na této úrovni nejpřesněji vystihují skutečný vztah proměnných. Prve bude pozornost orientována na covidový jev (obrázky 12 a 13).

Mezi počtem nakažených na 1000 obyv. a podílem obyvatel ve věku 65 a více let byla v rámci obcí zjištěna slabá negativní závislost. Ta se na všech vyšších úrovních agregace jeví jako ještě slabší, v případě funkčních regionů je nejnižší. Hodnoty variačního koeficientu této dvojice proměnných jsou si velmi blízko v rámci všech pěti typů členění území. Nejpodobnější míru závislosti v porovnání

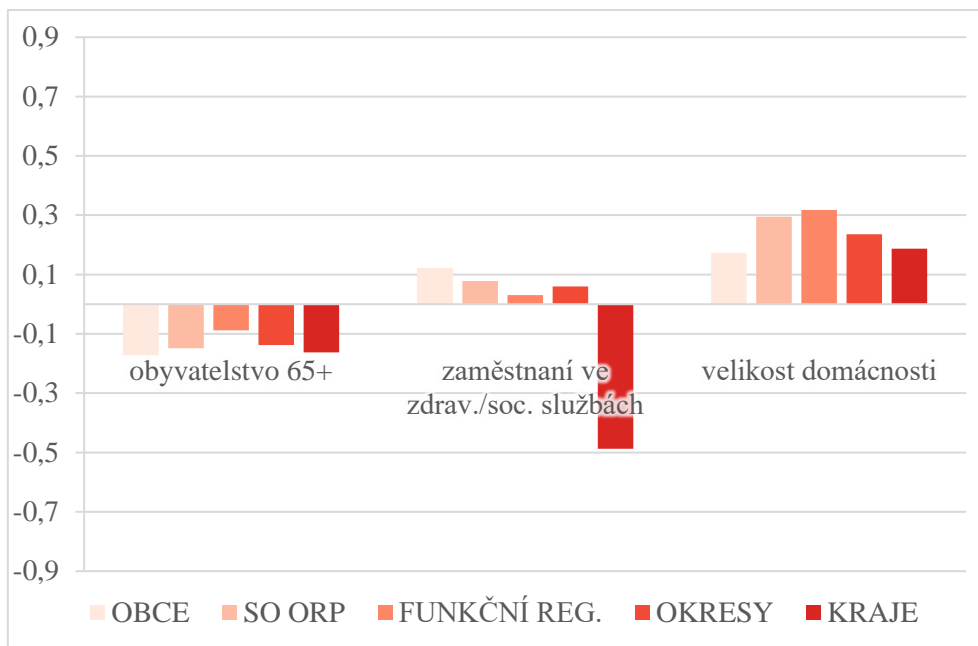


s obecní nacházíme v případě krajů, kde ale můžeme zesílení negativního vztahu připisovat spíše nízké variabilitě obou proměnných. Kromě krajské struktury s obecní mírou korelace nejvíce koresponduje úroveň správních obvodů.

Výsledky závislosti covidového ukazatele a zaměstnanosti ve zdravotních a sociálních službách jsou natolik pozoruhodné, že byly motivací ke zopakování celého procesu výpočtů, které však přineslo totožný výsledek a nedostatky (jak bylo popsáno v rámci metod výzkumu) byly objeveny pouze v zaměnění vstupních hodnot necelé desítky obcí, kromě toho se jednalo o jiné proměnné. V rámci obcí a tří následujících hierarchických úrovní členění ukazuje korelační koeficient na slabě pozitivní vztah, přičemž ve funkčním regionálním systému je nejméně významný. V krajském členění je však korelace mezi těmito proměnnými hodnocena jako středně negativní, téměř silně negativní. Závislost počtu případů na zaměstnanosti ve jmenovaných službách v rámci krajů má tedy velmi odlišný charakter než vztahy zjištěné na nižších hierarchických úrovních. Ukazatel zaměstnanosti ve zdrav. a soc. službách je z hlediska variability s postupným zvyšováním agregace nejpodstatněji zredukován.

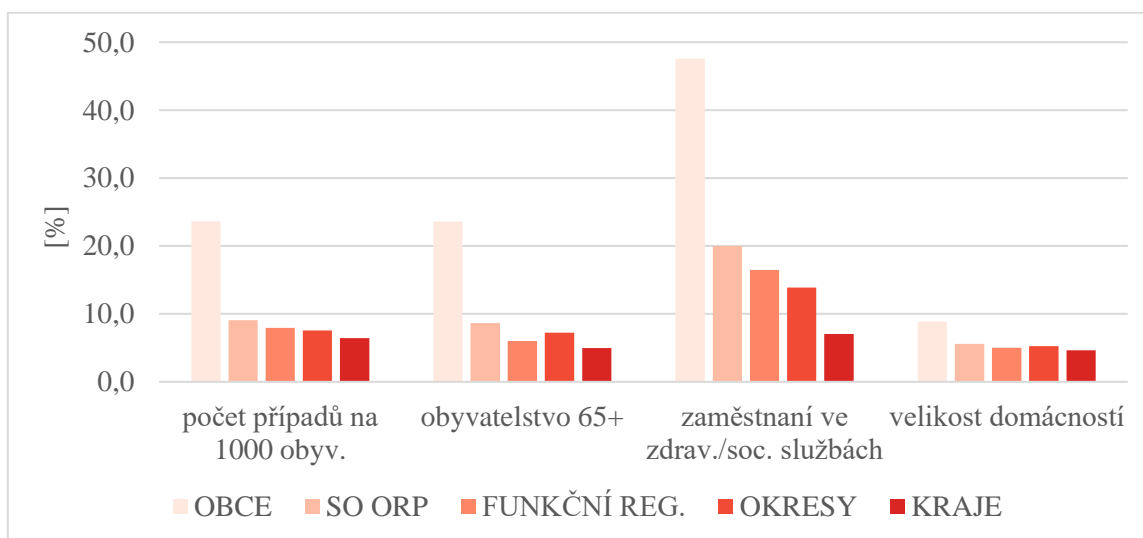
Kladné hodnoty Pearsonova korelačního koeficientu pro covidový jev a průměrný počet členů domácnosti byly získány v každé z pěti forem seskupení dat. V rámci obcí se právě na této ze tří pomocných proměnných ukázal být počet případů na 1000 obyv. nejvíce závislý, přestože zjištěný vztah opět klasifikujeme jako slabý. Na vyšších úrovních pozorujeme silnější vztah a nejvíce se od obecní struktury liší funkční regionální systém, který jako jediný ukazuje střední sílu závislosti. Od okresního členění velikost závislosti opět klesá a nejvíce se hodnotě korelačního koeficientu v obcích blíží opět kraje, přestože je zde variabilita pomocné proměnné nejnižší. Nutno ale poznamenat, že ukazatel velikosti domácnosti vykazuje napříč typy členění nejnižší meziregionální variabilitu a je také s úrovní agregace nejméně pozměněn.

Mezi covidovým jevem a zvolenými nezávisle proměnnými byl na obecní úrovni zjištěn velmi slabý vztah, přičemž nejsilněji s počtem případů nakažení koreloval ukazatel početnosti domácnosti. V případě věkové skupiny starší 64 let se jednalo o negativní závislost a zaměstnanost ve zdravotních a sociálních službách či velikost domácnosti hodnoty covidové proměnné ovlivňovaly kladně. S vyšší úrovní agregace se korelace mezi ukazateli zesilovala i zeslabovala, byla zaznamenána také změna znaménka korelačního koeficientu. Došlo také ke změně intenzity vztahu ze slabé na střední.



Obrázek 12: Hodnota korelačního koeficientu covidové charakteristiky s příslušnými pomocnými proměnnými v pěti typech členění území

Zdroj: ČSÚ, Ministerstvo zdravotnictví ČR, 2023; vlastní zpracování



Obrázek 13: Hodnota variačního koeficientu covidové charakteristiky a třech pomocných proměnných v pěti typech členění území

Zdroj: ČSÚ, Ministerstvo zdravotnictví ČR, 2023; vlastní zpracování

Hodnoty pomocných nezávisle proměnných pro volební jev jsou ve srovnání s covidovými značně variabilnější (viz obrázek 15). První z nich je podíl nezaměstnaných osob, jehož variační koeficient v obecní struktuře sahal nad 65 % a s rostoucí agregací byl zredukován na 25 % v krajském seskupení. Zjištěná povaha korelace s touto pomocnou proměnnou je negativní a pokud jde o obecní členění, pouze v něm je síla vztahu střední (viz obrázek 14). S rostoucí hierarchickou úrovní administrativního členění

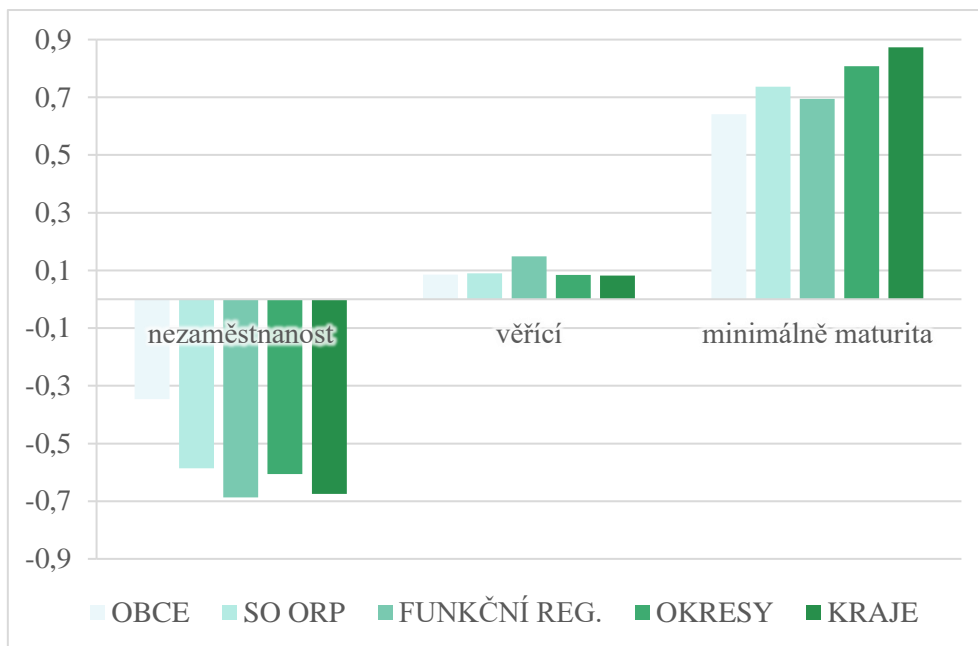


se jedná o stupňující se silně negativní vztah, což odpovídá poklesu variability pomocné proměnné. Nejnižší sahá hodnota korelačního koeficientu v rámci funkčního regionálního systému a zde se jedná o nejvýraznější zkreslení, co se týká volebního jevu. Ovlivnění výsledku variabilitou souboru je možné připustit i zde, neboť rozmanitost hodnot podílu hlasů pro zvoleného prezidenta mezi funkčními regiony byla nižší než mezi okresy.

Nezávisle proměnná podíl věřících je s rostoucí agregací z hlediska variability modifikována méně než ukazatel nezaměstnanosti. Ve srovnání s volební charakteristikou u ní pozorujeme výrazně vyšší meziobecní rozmanitost. Pozorovaný korelační koeficient ukazující její vztah s volebním výsledkem v rámci obcí zůstává napříč administrativními typy členění téměř nepozměněn a je kladného charakteru. Jedná se o nejméně významný zjištěný vztah tohoto výzkumu. Nad 0,1 se hodnocení závislosti dostává pouze v případě funkčních regionů.

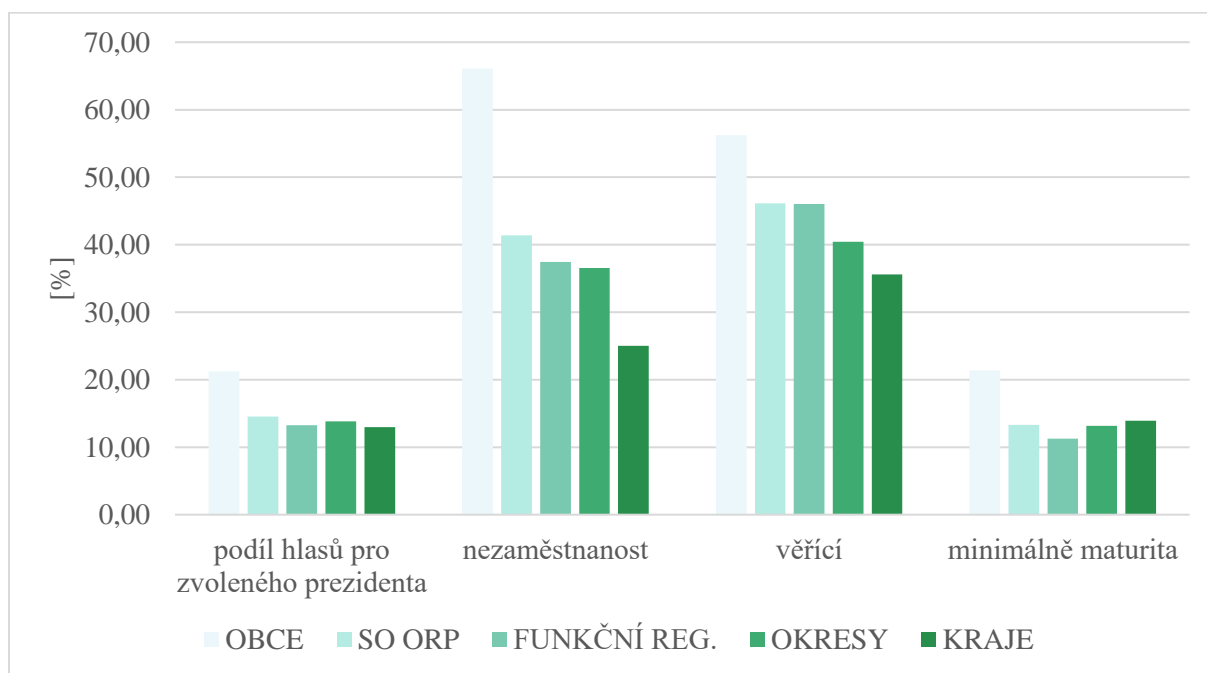
Poslední hodnocený vztah je naopak nejsilnějším v rámci výzkumu a jedná se o souvislost volebního jevu s ukazatelem vzdělanosti, konkrétně s podílem obyvatel minimálně s maturitním vzděláním. Co se týče variability pomocné proměnné v administrativních regionálních systémech (mimo obecní členění), liší se mezi nimi pouze o desetiny procenta a na krajské úrovni je dokonce vyšší než v případě správních obvodů ORP. Nejnižší je ve struktuře funkčních regionů. Pozorovaná hodnota Pearsonova koeficientu je kladná a na nejdetailnější měřítkové úrovni je vyšší než 0,64. Nárůst hierarchické úrovně v případě administrativních celků se projevuje postupným zesilováním vztahu, kdy krajské uspořádání modifikuje hodnotu ukazatele závislosti na vyšší než 0,87. Funkční regiony se v případě tohoto vztahu ukázaly být nejodolnější formou seskupení dat vůči zkreslení, jelikož se zjištěná korelace nejméně liší od nejpodrobnější struktury agregace.

Negativní vztah volební charakteristiky s nezaměstnaností i její pozitivní vztah s vzdělanostní úrovní obyvatelstva dokázaly zkoumané formy slučování vstupních dat zesílit. V případě podílu nezaměstnaných osob intenzivněji. Oproti covidovému jevu je trend změny závislosti s rostoucí agregací jednostranný, výjimkou jsou funkční regiony. Proměnná věřící podle výsledků ovlivnila podíl hlasů pro zvoleného prezidenta velmi nepatrně a rovněž účinek agregace na zjištěný vztah byl v porovnání s ostatními zkoumanými dvojicemi ukazatelů zanedbatelný. Také u tohoto výroku platí výjimka pro funkční systém členění, v němž se závislost volebního jevu a víry projevuje významněji.



Obrázek 14: Hodnota korelačního koeficientu volební charakteristiky s příslušnými pomocnými proměnnými v pěti typech členění území

Zdroj: ČSÚ, 2022, 2023; vlastní zpracování



Obrázek 15: Hodnota variačního koeficientu volební charakteristiky a třech pomocných proměnných v pěti typech členění území

Zdroj: ČSÚ, 2022, 2023; vlastní zpracování

## 6 Diskuze

Všechny problémy prostorových informací spojené s agregací dat, na které upozorňují teoretické základy práce byly v rámci výzkumu prokázány. Oba zkoumané prostorové jevy ve všech formách členění podléhají účinkům agregčního zkreslení. Ovlivnění těchto efektů charakteristikami prostorových dat jako jsou variabilita či kontinuita jsou patrné. Z výsledků však nelze vyvodit jednoznačné zákonitosti způsobu ani míry zjištěného prostorového zkreslení.

Zvyšující se hierarchická úroveň územního členění znamenala pokles meziregionální variability dvou hlavních jevů. Výjimkou bylo funkční regionální členění a covidový jev. Shlazení rozdílů v rámci statistického souboru údajů při slučování individuálních prostorových dat do větších územních celků proklamují například Wong (1996) či Haining (2003). Sledována byla také variabilita pomocných proměnných pro korelační analýzu, také zde se nejvyšší rozmanitost hodnot projevila mezi obcemi. U některých z těchto ukazatelů se od jednostranného poklesu variability s rostoucí agregací odlišovalo funkční regionální členění, kde byla redukce meziregionálních rozdílů vyšší než v rámci okresů. Proměnná minimálně maturita se v kontrastu s tvrzením zmíněných autorů ukázala být variabilnější ve struktuře čtrnácti krajů než v rámci 206 správních obvodů ORP. Příčinu lze hledat v prostorovém rozložení jevu, přesněji ve více regionalizovaných shlucích podprůměrných a nadprůměrných hodnot, které jsou lépe zachyceny krajskou strukturou. Jak totiž zdůrazňuje Meentemeyer (1989) nebo Hipp (2007), různé prostorové procesy a jevy fungují na různých měřítkových úrovních.

Prostorové vzory existující v nejpodrobnějším členění území byly při použití územních systémů s většími jednotkami méně zřetelné. Mapy distribuce jevů v různých typech členění poukázaly zejména na problém agregace související s problémem modifikovatelné územní jednotky zkoumaný například Openshawem (1984), Fotheringhamem a Wongem (1991) a dalšími. Testované regionální systémy prostorové vzory obecní úrovně díky zprůměrování extrémních hodnot zobecňovaly a docházelo také k nadhodnocení či podhodnocení lokalit. Výsledky tedy potvrzují tvrzení Mathera a Openshawa (1974), že na způsobu roztržení údajů do plošných jednotek záleží. Prostorové zkreslení se většinou zvyšovalo s rostoucí hierarchickou úrovní. Rozmanitost výsledků byla mnohem více patrná ze struktury správních obvodů ORP než z krajského členění. V rámci map distribuce sledovaných proměnných však bylo upozorněno na konkrétní místa, kde okresní členění podhodnocovalo či nadhodnocovalo lokální vzor jevu méně než kupříkladu funkční regiony či správní obvody ORP. Důležité je tedy i to, jak jsou v prostoru konfigurovány hranice regionů (Fotheringham a Wong, 1991).

Byl prokázán také problém malých čísel, který souvisel například s pozorováním maximálních hodnot v populačně malých obcích. Problematická je nedostatečná velikost vzorku pro výpočet výsledné míry, v důsledku čehož může dojít k přecenení výsledků, jak upozorňují Gelman a Price (1999) nebo Haining (2003). Tento efekt mohl souviset také se zahrnutím vojenských újezdů s nulovým vzorkem pro výpočet

měř zejména do správních obvodů ORP a kde tak mohlo dojít k nadhodnocení jevu, jak bylo upozorněno v případě covidové proměnné a SO ORP Vyškov.

Že geografické informace jsou téměř vždy pozitivně autokorelované tvrdí Mather a Openshaw (1974), Fotheringham a kol. (2003) nebo Spurná (2008). Měření prostorové závislosti ve zkoumaných datech potvrdilo pozitivní autokorelaci obou jevů ve všech typech členění území. Nejvýraznější podobnost hodnot prostorově blízkých jednotek byla pro oba jevy zjištěna v případě agregace dat do okresů.

Jak poznamenává Fischer (2006), zjištěná míra autokorelace v regionálním systému může být ovlivněna souladem mezi vymezením administrativních jednotek a prostorovým vzorem studovaného fenoménu. Problém modifikovatelné územní jednotky tedy mohl souviset také s měřením prostorové závislosti. Pokud se podíváme na mikroregionální úroveň, za kterou lze považovat územní systémy SO ORP, funkčních regionů a okresů, byla zde zaznamenána výraznější prostorová závislost sousedních jednotek. V rámci obou jevů se funkční systém členění vyznačoval nejnižší úrovní shlukování na mikroregionální úrovni. Na rozdíl od administrativních jsou funkční regiony definovány na základě prostorových interakcí a vybočují ze skladebnosti do vyšších hierarchických úrovní, která platí pro ostatní použité regionální systémy.

Rozdílné úrovně shlukování vykazovaly čtyři různé jevy zkoumané pomocí metod lokální prostorové autokorelace na území České republiky ve výzkumu Klapky a kol. (2016). Lokální analýza prostorové autokorelace v tomto výzkumu potvrdila pozitivní povahu shlukování dvou hodnocených jevů. Identifikovala klastry podobných hodnot, takzvané cold spots a hot spots, ale také oblasti, kde spolu jevy prostorově nesouvisí. V případě obou zkoumaných proměnných identifikovaly výsledky LISA shluky většinově na makroregionální úrovni, bylo poukázáno na rozsáhlé oblasti, kde oba zkoumané jevy shodovaly v umístění hot a cold spots. Makroregionální úroveň je v rámci výzkumu představována systémem čtrnácti krajů. V některých případech dochází k podobnosti shluků s územním vymezením krajů, zejména cold spots a Karlovarský a Ústecký kraj nebo hot spots a Středočeský, Královéhradecký a Pardubický kraj.

Přítomnost prostorové závislosti v datech zvyšuje náchylnost ke zkreslení prostorových analýz v souvislosti s MAUP. Na tom se shodují autoři Gehlke a Biehl (1934), Openshaw a Taylor (1979) nebo Garretton a kol. (2020). Také lokální analýza autokorelace poukázala na některé případy prostorové neurčitosti. Bylo upozorněno například na částečné zaniknutí klastru podprůměrných hodnot východně od Prahy patrného z obecního členění, pokud jsou data agregována do větších prostorových celků.

Nejlépe uchopitelný a matematicky doložitelný způsob potvrzení agregačních efektů přinesla závěrečná část výzkumu, která zkoumala výsledky korelačního koeficientu mezi sledovanými jevy a pomocnými proměnnými v pěti typech členění území. Na stejném principu proběhlo první odhalení agregačního zkreslení ve 30. letech (Gehlke a Biehl, 1934). Z výsledků je jisté, že sledované geografické informace

v rámci hodnocených forem jejich seskupení podléhají agregačním efektům, neboť docházelo ke změnám korelačního koeficientu u každé z testovaných proměnných.

Nejčastěji je autory potvrzován nárůst korelace mezi proměnnými se zvyšováním úrovně agregace (Robinson, 1950; Openshaw, 1977). Ten je vysvětlován nepřímou úměrou s variabilitou, která klesá s nárůstem hierarchické úrovně (Blalock, 1964; Fotheringham a Wong, 1991). Objevovaly se však také práce, kde se změnou agregace nedocházelo k monotónnímu vývoji hodnot korelačních koeficientů (například Blalock, 1964; Clark a Avery, 1976). Tento výzkum nepotvrdil čistě rostoucí trend změny závislosti s vyšší hierarchickou úrovní ani v jednom z testovaných vztahů. Průkaznější bylo zesílení závislosti v rámci volebního jevu a o postupný nárůst šlo pouze v případě administrativních regionů. Intenzivnější zesílení korelací u volební proměnné lze přisuzovat skutečnosti, že se jedná o výrazněji autokorelovaný jev, ale také je možné hledat příčinu ve výběru pomocných ukazatelů, které s danou proměnnou souvisely více.

U celkem tří ze šesti sledovaných vztahů bylo pozorováno také zeslabení korelace. Nutno dodat, že se jednalo o nejslaběji korelované dvojice proměnných. U covidové proměnné byla zaznamenána také změna povahy vztahu ze slabě pozitivního na téměř silně negativní při přechodu na nejvyšší úroveň agregace. O změně znaménka korelačního koeficientu v důsledku agregace dat hovoří Fotheringham a Wong (1991) nebo Andresen (2021).

Z výzkumu vyplývá, že interpretace korelací zjištěných z agregovaných dat by mohla být problematická. Jsou totiž více či méně zkreslené a existuje tak riziko ekologického omylu. Chybnost předpokladu platnosti korelací zjištěných z agregovaných dat na individuální úrovni potvrzují Robinson (1950), Openshaw (1984) či Salkeld a Antolin (2020). Nejvýraznější změnou závislosti u covidového jevu je korelace se zaměstnaností ve zdravotních a sociálních službách. Ta je na krajské úrovni téměř silně negativní, zatímco v rámci obcí je slabě pozitivní. Pokud jde o volební charakteristiku, mohli bychom se dopustit ekologické chyby při tvrzení, že výsledky zvoleného prezidenta silně negativně ovlivnil podíl nezaměstnaných osob, přičemž na obecní úrovni je síla vztahu střední.

Co se týká hodnocených typů členění, v rámci každého z nich podléhala zkoumaná data určité míře agregačního zkreslení. Vhodný typ členění území České republiky pro svůj výzkum z oblasti sociálních procesů posuzovali také Musil a Müller (2008) a naráželi na problém zániku jistých procesů v důsledku zprůměrování dat pro agregované jednotky. Rovněž upozornili na kolísání výsledků v rámci obecního členění způsobené velkými rozdíly ve velikosti vzorku. S touto nekonzistencí se setkáváme také ve výsledcích předkládané práce (viz kapitola 5.1).

Ve všech fázích výzkumu bylo patrné odlišné chování použitého funkčního regionálního systému. Při zkoumání změn korelačních koeficientů došlo k největšímu zkreslení vztahů zjištěných ze struktury obcí funkčními regiony, a to u čtyř ze šesti hodnocených dvojic proměnných. Administrativní regiony (okresy a správní obvody ORP) porovnávali s jinými funkčními členěními z hlediska vhodnosti pro

prostorové analýzy Klapka a kol., přičemž nenalezli významné rozdíly ve schopnosti zachytit prostorové vzory (2020). Autoři se však zaměřovali primárně na srovnání systémů podobné hierarchické úrovně.

## 6.1 Limity výzkumu

Výběr prostorových informací a územních systémů byl limitován dostupností dat. V případě použitých členění šlo o dostupnost datových vrstev prostorových systémů obsahující kódy obcí a kódy jejich příslušnosti k dané územní jednotce. Pro možnost zaměření na problém agregace (zonace) by bylo vhodné použít regiony stejné hierarchické úrovně, které však nebyly k dispozici. Zajímavé by bylo také hodnocení historických územně správních typů členění (staré okresy/kraje) či hodnocení variant návrhů současných administrativních jednotek.

Volbu prostorových informací ke zkoumání omezovala potřeba dostupnosti dat na podrobné měřítkové úrovni (obce). Ne zcela vhodný výběr lze připustit především v případě pomocných proměnných pro zkoumání korelací. Bylo by možné očekávat zajímavější výsledky pro ukazatele, které by vykazovaly silnější závislost se zkoumanými jevy, než tomu bylo u slabě korelovaných proměnných (zaměstnanost ve zdrav./soc. službách či podíl věřících). Nevhodně byla zvolena proměnná podílu obyvatel ve věku 65 a více let v případě covidové proměnné. Tento ukazatel by se patřilo porovnávat spíše s počtem úmrtí v důsledku tohoto onemocnění, zatímco zkoumaná charakteristika vyjadřovala počet nově diagnostikovaných případů. I přesto zde byl zjištěn slabě pozitivní vztah.

Je třeba připomenout také již zmiňovanou nevýhodu datového souboru covidové proměnné spočívající v zahrnutí reinfekcí, které mohly způsobit nadhodnocení výsledků. Počty diagnostikovaných případů očištěné od opakovaných výskytů onemocnění u stejných osob nejsou k dispozici. Jejich použití by však mohlo přinést přesnější výsledky, co se týče covidového jevu.

Soubor analýz dat prováděných v rámci výzkumu obsahoval četné výpočty a manipulaci s obsáhlými datovými soubory. V důsledku tak velkého počtu operací je možné předpokládat vznik pochybení a nepřesností, které nebyly odhaleny při procesu zopakování výpočtů osmi ukazatelů.

Hodnocení variability studovaných statistických souborů bylo omezeno pouze na výpočet variačního koeficientu, jakožto relativního ukazatele vycházejícího ze směrodatné odchylky a aritmetického průměru. Výzkum regionální variability sledovaných proměnných by bylo možné rozvinout pomocí použití Theilova indexu. Ten umožňuje rozklad celkové variability na tu mezi regiony a tu uvnitř jednotlivých regionů. Dá se tak díky němu zjistit, do jaké míry je celková variabilita tvořena meziregionálními rozdíly.

## 7 Závěr

Záměrem diplomové práce bylo zkoumání zákonitostí prostorové neurčitosti a prostorového zkreslení, které jsou potenciálním rizikem při slučování geografických informací. Výzkum byl realizován na konkrétních vstupních datech, a sice dvou pro českou společnost aktuálních fenoménech, kterými jsou úhrn nákazy onemocněním COVID-19 a výsledek prezidentských voleb 2023. Hodnocenými formami agregace těchto dat byly administrativní a funkční systémy členění území České republiky. Byly zkoumány také změny vlastností sledovaných informací, které se určitým způsobem pojí s rizikem agregačních efektů.

U obou hodnocených jevů se projevila náchylnost k agregačnímu zkreslení. Jednotlivé fáze zkoumání poukázaly na skrývání prostorových vzorů a také na kolísání výsledků statistických analýz proměnných v rámci každého z testovaných forem seskupení. Bylo prokázáno také působení specifických vlastností zkoumaných prostorových dat (heterogenita, kontinuita) na projevy agregačního zkreslení. Nebyl však rozpoznán princip, na základě kterého by bylo možné jednoznačně objasnit charakter a především výslednou míru pozorovaných nepřesností. Částečně se však potvrdil nárůst zkreslení se zvyšující se úrovní agregace.

Výzkum dokazuje, že informace prezentované ve struktuře agregovaných prostorových jednotek mohou vykazovat určitou míru nepřesnosti a mohou se lišit od skutečných prostorových vzorů existujících na individuální úrovni. Dva hodnocené fenomény a především jejich prostorová distribuce jsou jedněmi z obvyklých předmětů prostorových analýz. Velmi čtené rozbory, interpretace a zejména kartografická vyjádření prostorového rozmístění onemocnění COVID-19 i volebních výsledků v agregované formě bylo možné zaznamenat také v mediálním prostoru. Potvrzení skutečnosti, že odlišné vymezení územních jednotek pro vyjádření jevů by velmi pravděpodobně poskytlo rozdílné výsledky, je přínosem výzkumu s přesahem do dalších, zejména společenskovedních oblastí.

Na závěr se nabízí úvaha o potenciálních směrech rozvinutí studie. Jak již bylo zmíněno, existuje řada vědeckých prací navrhujiících metodiku potlačení, přesněji řečeno zmírnění agregačních efektů. Mnohé z těchto metod redukce agregačního zkreslení spočívají v principu tzv. prostorového filtrování. Jedná se o odfiltrování prostorové závislosti dat pomocí jejího rozkladu na prostorovou a neprostorovou složku. Aplikace těchto postupů na zkoumaná prostorová data by mohla přinést zajímavé výsledky.

## Summary

The thesis deals with a complex problem called the spatial bias or spatial uncertainty that often occurs when aggregated data are used. Aggregational bias is associated with almost each of spatial data analysis as geographical information is usually categorised in some kind of classes (spatial units). Many researches have shown that the problem is connected to the nature of spatial data which is very unique. Geographical space is heterogenous but the variability is some kind of continuous and spatially dependent. That means there are spatial concentrations of similar values being observed.

The main aim of this research is to identify potential aggregational effects in chosen spatial information and to investigate the connection with its variability and spatial dependence. There are two geographical phenomena being investigated. The first variable expresses the number COVID-19 diagnoses per 1000 inhabitants and the second is the percentage of votes for the president elect in Czech presidential election 2023. The investigation is focused on the Czech Republic and there are five forms of zoning systems of interest. The research consists of several parts which are focused on examining variability, continuity, spatial distribution and the change of correlation coefficient with increasing level of aggregation.

The variability test has shown that the variability of both variables was decreasing as the level of aggregation was higher. Also the maps of spatial distribution displayed that aggregated forms can cover diverse spatial patterns that are noticeable at the individual level. The tests for spatial association have confirmed that the examined data are spatially dependent. The electoral variable has shown to be a bit more spatially concentrated. Also the tests for the stability of correlation coefficient were affected by spatial bias as the results were fluctuated across the zoning systems.

The research explored that the aggregational bias is present in all of the examined information and zoning systems. Also the affection by specific nature of the data was noticed. The role of variability or spatial dependence for intensity of the observed spatial bias was not clear in all cases. However, this kind of uncertainty is typical for spatial bias phenomena.



## Použité zdroje informací

- ALVANIDES, S. a S. OPENSHAW, 1999. Zone design for planning and policy analysis. *Geographical information and planning: European perspectives*. 299–315.
- ANDRESEN, M. A., 2021. Modifiable areal unit problem. *The Encyclopedia of Research Methods in Criminology and Criminal Justice*. **2**, 854–855.
- ANSELIN, L., 1989. *What is special about spatial data? Alternative perspectives on spatial data analysis*. UC Santa Barbara: National Center for Geographic Information and Analysis.
- ANSELIN, L., 1995. Local Indicators of Spatial Association – LISA. *Geographical Analysis*. **27**(2), 93–115.
- ArcČR® 4.1: Digitální vektorová databáze České republiky, 2022. ČÚZK, ČSÚ, ARCDATA PRAHA.
- BLALOCK, J. H. M., 1964. *Causal inferences in non-experimental research*. Chapel Hill, NC: University of North Carolina Press.
- BOURDIN, S., a kol. Does lockdown work? A spatial analysis of the spread and concentration of Covid-19 in Italy. *Regional Studies*, 2021, 55.7: 1182–1193.
- CLARK, W. A. V. a K. L. AVERY, 1976. The effects of data aggregation in statistical analysis. *Geographical analysis*. (8), 428–438.
- CLIFF, A. D. a J. K. ORD, 1973. *Spatial Autocorrelation*. Pion.
- COVID-19: Epidemiologická charakteristika obcí, 2023. *Katalog otevřených dat* [online]. Ministerstvo zdravotnictví České republiky [cit. 2023-02-25]. Dostupné z: [https://opendata.mzcr.cz/cs\\_CZ/dataset/covid-19-epidemiologicka-charakteristika-obci/resource/b1abc429-8f36-4867-8641-ea7980449011](https://opendata.mzcr.cz/cs_CZ/dataset/covid-19-epidemiologicka-charakteristika-obci/resource/b1abc429-8f36-4867-8641-ea7980449011)
- ČSÚ a územně analytické podklady: Aktuální údaje za všechny obce ČR (data mimo SLDB), 2022. *Český statistický úřad* [online]. [cit. 2023-03-17]. Dostupné z: [https://www.czso.cz/csu/czso/csu\\_a\\_uzemne\\_analyticke\\_podklady](https://www.czso.cz/csu/czso/csu_a_uzemne_analyticke_podklady)
- EHLERT, A., 2021. The socio-economic determinants of COVID-19: A spatial analysis of German county level data. *Socio-economic planning sciences*. (78).
- FISCHER, M. M., 2006. Spatial analysis in geography. *Analysis and GeoComputation: Selected Essays*. 17–28.
- FISCHER, M. M. a A. GETIS, 2010. Introduction. In: FISCHER, Manfred M. a Arthur GETIS, ed. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin: Springer, s. 1–26. ISBN 978-3-642-03646-0.

- FORTIN, M.-J. a M. R. T. DALE, 2008. Spatial autocorrelation. In: *The SAGE Handbook of Spatial Analysis*. SAGE Publications, s. 89–104. ISBN 978-1-4129-1082-8.
- FOTHERINGHAM, A. S. a D. W. WONG, 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and planning A*. **23**(7), 1025–144.
- FOTHERINGHAM, A. S., C. BRUNDSON a M. CHARLTON, 2003. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley.
- FOTHERINGHAM, A. S. a Peter A. ROGERSON, 2008. Introduction. In: FOTHERINGHAM, A. Stewart a Peter A. ROGERSON, ed. *The SAGE Handbook of Spatial Analysis*. SAGE Publications, s. 1–5. ISBN 978-1-4129-1082-8.
- FREUND, R. J. a W. J. WILSON, 2003. *Statistical methods*. New York: Academic Press.
- FUSCO, G., a kol., 2017. Questions of uncertainty in geography. *Environment and Planning A*. **49**(10), 2261–2280.
- GARRETON, M., A. BASAURI a L. VALENZUELA, 2020. Exploring the correlation between city size and residential segregation: comparing Chilean cities with spatially unbiased indexes. *Environment and urbanization*. **32**(2), 569–588.
- GEHLKE, C. E. a K. BIEHL, 1934. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*. **29**(185A), 169–170.
- GELMAN, A. a P. N. PRICE, 1999. All maps of parameter estimates are misleading. *Statistics in medicine*. **18**(23), 3221–3234.
- GETIS, A. a J. K. ORD, 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical analysis*. **24**(3), 189–206.
- GETIS, A., 2010. Spatial autocorrelation. In: FISCHER, M. M. a A. GETIS, ed. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin: Springer, s. 255–275. ISBN 978-3-642-03646-0.
- GOODCHILD, M. F., 2004. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers*. **94**(2), 300–303.
- GRIFFITH, D. A., 2018. Uncertainty and context in geography and giscience: reflections on spatial autocorrelation, spatial sampling, and health data. *Annals of the American Association of Geographers*. **108**(6), 1499–1505.
- HAINING, R. P., 2003. *Spatial data analysis: theory and practice*. Cambridge university press.

- HAINING, R. P., 2008. The Special Nature of Spatial Data. In: FOTHERINGHAM, A. S. a P. A. ROGERSON, ed. *The SAGE Handbook of Spatial Analysis*. SAGE Publications, s. 5–24. ISBN 978-1-4129-1082-8.
- HAINING, R. P., 2010. The Nature of Georeferenced Data. In: FISCHER, M. M. a A. GETIS, ed. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*. Berlin: Springer, s. 197–214. ISBN 978-3-642-03646-0.
- HALÁS a kol., 2021. Population movements based on mobile phone location data: the Czech Republic. *Journal od Maps*. **17**(1), 116–122, DOI: 10.1080/17445647.2021.1937730
- HALÁS, M. a P. KLAPKA, 2020. Heterogenita a kontinuita geografického prostoru: příklad funkčních regionů Slovenska. *Geografie*. **125**(3), 319–342.
- HIPP, J. R., 2007. Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American sociological review*. **72**(5), 659–680.
- HOLT, D. a kol., 1996. Aggregation and ecological effects in geographically based data. *Geographical analysis*. **28**(3), 244–261.
- HOLÝ, V., 2011. Jak a proč chráníme důvěrné statistické údaje. *Statistika a my*. ČSÚ, 26–27.
- IRA, V. a R. MATLOVIČ, 2020. Challenges and opportunities for human geography: a few remarks. *Geographia Polonica*. **93**(4), 525–537.
- KLAPKA, P., a kol., 2016. The efficiency of areal units in spatial analysis: Assessing the performance of functional and administrative regions. *Moravian Geographical Reports*. **24**(2), 47–59.
- KLAPKA, P., 2019. *Regiony a regionální taxonomie: koncepty, přístupy, aplikace*. Palacký University Olomouc. ISBN 978-80-244-5448-1.
- KORČÁK, Jaromír, 1941. Přírodní dualita statistického rozložení. *Statistický obzor*. **22**, 171–222.
- KWAN, M. P., 2018. The limits of the neighborhood effect: Contextual uncertainties in geographic, environmental health, and social science research. *Annals of the American Association of Geographers*. **108**(6), 1482–1490.
- MAŠKARINEC, P. a kol., 2013. Prostorová analýza prezidentských voleb v České republice v roce 2013. *Sociológia-Slovak Sociological Review*. **45**(5), 435–469.
- MATHER, P. a S. OPENSHAW, 1974. Multivariate methods and geographical data. *Journal of the Royal Statistical Society: Series D (The Statistician)*. **23**(3-4), 283–308.
- MEENTEMEYER, V., 1989. Geographical perspectives of space, time, and scale. *Landscape Ecology*. **3**, 163–173.

- Metodika sledovaných ukazatelů: Informace k vybraným ukazatelům pro potřeby územně analytických podkladů [online], 2022. Oddělení výstupních databází Českého statistického úřadu [cit. 2023-03-02]. Dostupné z: [https://www.czso.cz/csu/czso/csu\\_a\\_uzemne\\_analyticke\\_podklady](https://www.czso.cz/csu/czso/csu_a_uzemne_analyticke_podklady)
- MUSIL, J. a J. MÜLLER, 2008. Vnitřní periferie v České republice jako mechanismus sociální exkluze. *Sociologický časopis/Czech Sociological Review*. **44**(2), 321–348.
- NETRDOVÁ, P. a V. NOSEK, 2009. Přístupy k měření významu geografického rozměru společenských nerovnoměrností. *Geografie-Sborník české geografické společnosti*. **114**(1), 52–65.
- OPENSHAW, S., 1977. A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the institute of british geographers*. **2**(4), 459–472.
- OPENSHAW, S., 1984. The modifiable areal unit problem. CATMOG 38. GeoBooks, Norwich, England.
- OPENSHAW, S. a P. J. TAYLOR, 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: WRIGLEY, N., ed. *Statistical Applications in the Spatial Sciences*. London: Pion, s. 127–144.
- POLLET, T. V. a kol., 2014. What can cross-cultural correlations teach us about human nature. *Human nature*. **25**, 410–429.
- REY, S., 2015. Mathematical models in geography. In: *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*. Elsevier, s. 785–790.
- ROBINSON, A. H., 1950, Ecological correlation and the behaviour of individuals. *American Sociological Review* **15** 351-357.
- SALKELD, D. J. a M. F. ANTOLIN, 2020. Ecological fallacy and aggregated data: a case study of fried chicken restaurants, obesity and Lyme disease. *EcoHealth*. **17**, 4–12.
- SPURNÁ, P., 2008. Prostorová autokorelace–všudypřítomný jev při analýze prostorových dat?. *Sociologický časopis/Czech Sociological Review*. **44**(4), 767–787.
- TOBLER, W. R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic geography*. **46**(2), 234–240.
- TURNERY, S. (2022). *Pearson Correlation Coefficient (r): Guide & Examples*. Scribbr. Dostupné z: <https://www.scribbr.com/statistics/pearson-correlation-coefficient/>
- UNWIN, D. J., 1996. GIS, spatial analysis and spatial statistics. *Progress in Human Geography*. **20**(4), 540–551.

Veřejná databáze: Volba prezidenta 2023, 2023. : Český statistický úřad [online]. Praha [cit. 2023-03-28]. Dostupné z: <https://vdb.czso.cz/vdbvo2/faces/cs/index.jsf?page=statistiky#katalog=34015>

VOGEL, M., 2016. The modifiable areal unit problem in person–context research. *Journal of Research in Crime and Delinquency*. **53**(1), 112–135.

*Výsledky sčítání 2021 - otevřená data* [online], 2023. Český statistický úřad [cit. 2023-02-16]. Dostupné z: <https://www.czso.cz/csu/czso/vysledky-scitani-2021-otevrena-data>

WARD, M. D. aj. O'LOUGHLIN, 2002. Spatial processes and political methodology: Introduction to the special issue. *Political Analysis*. **10**(3), 211–216.

WISE, S. a kol., 1997. Regionalisation tools for the exploratory spatial analysis of health data. *Recent developments in spatial analysis: spatial statistics, behavioural modelling, and computational intelligence*. 83–100.

WONG, D., 1996. Aggregation effects in geo-referenced data. In: *Practical handbook of spatial statistics*. CRC Press, s. 83–106.

WONG, D., 2008. The modifiable areal unit problem (MAUP). In: FOTHERINGHAM, A. Stewart a Peter A. ROGERSON, ed. *The SAGE Handbook of Spatial Analysis*. SAGE Publications, s. 105–124. ISBN 978-1-4129-1082-8.

YULE, G. U. a M. G. KENDALL, 1950. *An introduction to the theory of statistics*. 14th ed. London: Griffin.

Zákon č. 51/2020 Sb., o územně správním členění státu a o změně souvisejících zákonů (zákon o územně správním členění státu)