



Ekonomická
fakulta
Faculty
of Economics

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

Jihočeská univerzita v Českých Budějovicích
Ekonomická fakulta
Katedra aplikované matematiky a informatiky

Diplomová práce

Geograficky vážená regrese a její aplikace v oblasti regionálního rozvoje

Vypracoval: Bc. Milan Šindler

Vedoucí práce: RNDr. Renata Klufová, Ph. D.

České Budějovice 2015

ZADÁNÍ DIPLOMOVÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Bc. Milan ŠINDLER**
Osobní číslo: **E11473**
Studijní program: **N6208 Ekonomika a management**
Studijní obor: **Strukturální politika EU a rozvoj venkova**
Název tématu: **Geograficky vážená regrese a její aplikace v oblasti regionálního rozvoje**
Zadávající katedra: **Katedra aplikované matematiky a informatiky**

Z á s a d y p r o v y p r a c o v á n í :

Cílem práce je aplikace geograficky vážené regrese při modelování regionálních rozvojových procesů. Součástí práce bude klasifikace prostorových regresních modelů, včetně diskuse jejich využití v různých oblastech. Podrobně budou popsána specifika geograficky vážené regrese a následně bude aplikována na konkrétním příkladu/příkladech z oblasti regionálního rozvoje.

Metodický postup:

1. Studium odborné literatury - literární přehled - klasické regresní modely versus prostorové regresní modely a jejich využití v odborné literatuře.
2. Geograficky vážená regrese - popis metody, její specifika a omezení, geograficky vážená regrese v odborné literatuře.
3. Metodická část - geograficky vážená regrese v programu ArcGIS, příp. v jiných SW aplikacích (GeoDA, R) - srovnání, její specifika a omezení, konkrétní model z oblasti regionálního rozvoje.
4. Akvizice dostupných dat, jejich příprava pro regresní analýzu a analýzu v prostředí GIS - návrh konceptuálního modelu.
5. Vlastní analytická část (počítačový model, dílčí analýzy, predikce).

Rozsah grafických prací: 20 map či výkresů
Rozsah pracovní zprávy: 50 stran
Forma zpracování diplomové práce: tištěná
Seznam odborné literatury:


1. Fotheringham, A. S., CH. Brundson a M. Charlton. *Quantitative Geography: Perspectives on Spatial Data Analysis*. SAGE Publications, 2002, 270 s.
2. Fotheringham, A. S., Ch. Brunsdon a M. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, 2002, 269 s.
3. Robinson, G. M. *Methods and Techniques in Human Geography*. Wiley, 1998, 556 s.
4. Ward, M. D. a K. S. Gleditsch. *Spatial Regression Models (Quantitative Applications in the Social Sciences)*. SAGE Publications, 2008, 102 s.
5. ESRI. ArcGIS Desktop 9.3 Help. *Webhelp.esri.com [online]*. [cit. 2012-06-08]. Dostupné z: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Regression_analysis_basics
6. Geographically Weighted Regression. *Ncg.nuim.ie [online]*. [cit. 2012-06-08]. Dostupné z: <http://ncg.nuim.ie/ncg/GWR/>

Vedoucí diplomové práce: RNDr. Renata Klufová, Ph.D.
Katedra aplikované matematiky a informatiky

Datum zadání diplomové práce: 4. června 2012
Termín odevzdání diplomové práce: 12. dubna 2013


doc. Ing. Ladislav Rolínek, Ph.D.
děkan

L.S.


prof. RNDr. Pavel Tlustý, CSc.
vedoucí katedry

V Českých Budějovicích dne 6. června 2012

Prohlášení

Prohlašuji, že svoji diplomovou práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47 zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své diplomové práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích dne 16. 4. 2015

Milan Šindler

Poděkování

Na tomto místě bych rád poděkoval RNDr. Renatě Klufové, Ph.D. za odborné rady, cenné připomínky, ochotu a čas, který mi byla ochotna věnovat při vytváření této diplomové práce.

Obsah

1	Úvod.....	11
1.1	Cíl práce	11
1.2	Hypotézy	11
2	Literární rešerše	12
2.1	Prostorová data.....	12
2.2	Geografické informační systémy	12
2.2.1	ArcGIS	14
2.3	Regresní analýza	14
2.3.1	Metoda nejmenších čtverců	15
2.3.2	Geograficky vážená regrese.....	16
2.4	Prostorové váhy.....	19
2.5	Prostorové shlukování.....	23
2.6	Prostorová autokorelace	24
2.6.1	Globální míry	25
2.6.2	Lokální míry	26
2.7	MAUP – Modifiable Area Unit Problem	26
2.8	Příklad použití geograficky vážené regrese	27
2.9	Výstupy geograficky vážené regrese	29
3	Metodika	30
3.1	Tvorba databáze	30
3.2	Vstupní proměnné	30
3.2.1	Hrubá míra migračního salda.....	30
3.2.2	Nezaměstnanost	30
3.2.3	Počet bytů	31
3.2.4	Hustota zalidnění	31

3.2.5	Hustota silniční sítě.....	31
3.2.6	Koeficient ekologické stability	31
3.3	Řešená území	32
3.3.1	Obce	32
3.3.2	Obce s rozšířenou působností	32
3.3.3	Okresy	32
3.3.4	Kraje.....	32
4	Aplikace vybraných metod	33
4.1	Regresní analýza	33
4.1.1	Obce – globální model.....	33
4.1.2	Obce – lokální model.....	38
4.1.3	Obce s rozšířenou působností – globální model.....	41
4.1.4	Obce s rozšířenou působností – lokální model.....	45
4.1.5	Okresy – globální model.....	48
4.1.6	Okresy – lokální model.....	52
4.1.7	Kraje – globální model	55
4.1.8	Kraje – lokální model	59
5	Závěr	63
I.	Seznam použitých zdrojů.....	65
II.	Summary.....	67
III.	Seznam obrázků a tabulek	68

1 Úvod

Prostorová data jsou taková data, která obsahují geografickou informaci, a často je s nimi manipulováno s pomocí geografických informačních systémů. Rovněž mají specifické vlastnosti, k jejichž zpracování se hodí použití prostorových statistických metod. Často se však používají i metody, které nemusí být pro zpracování prostorových dat vhodné. Proto v této práci v aplikační části bude demonstrováno použití vícenásobné lineární regrese a geograficky vážené regrese s následnou komparací vhodnosti metod. Jako vstupní socioekonomické proměnné poslouží hrubá míra migračního salda, nezaměstnanost, počet bytů, hustota zalidnění, hustota silniční sítě a koeficient ekologické stability.

Pro potřeby práce bude nutné sestavit rozsáhlou databázi na vybraném administrativním dělení ČR, jímž budou obce, ORP, okresy a kraje. Na následně sestavený regresní model v prostředí ArcGIS s výše jmenovanými proměnnými bude aplikována jak metoda nejmenších čtverců, tak i geograficky vážená regrese. V samotném závěru práce dojde ke komparaci výkonnosti metod za pomoci vybraných ukazatelů.

1.1 Cíl práce

Cílem práce je aplikace geograficky vážené regrese při modelování regionálních rozvojových procesů. Podrobně budou popsána specifika geograficky vážené regrese a následně bude aplikována na konkrétním příkladu z regionálního rozvoje.

1.2 Hypotézy

- a) Na prostorových datech použití geograficky vážené regrese může vést k lepším výsledkům v případě, že jsou data prostorově autokorelována.
- b) Problém modifikovatelné územní jednotky (MAUP) – se zmenšováním počtu jednotek do vyšších celků dochází ke zvyšování kvality modelu.

2 Literární rešerše

2.1 Prostorová data

V případě, kdy mají kterákoliv data vztah ke konkrétním místům v prostoru, která jsou uložena jako souřadnice a topologie, tak hovoříme o takzvaných prostorových datech. Klufová (2012) o prostorových datech píše, že *„téměř všechno, co se stane, se stane někde. Lidé jsou do značné míry vázáni svými aktivitami na povrch či spíše nejbližší část zemského povrchu. Cestují po něm, vytvářejí výkopy, do nichž pokládají potrubí a kabely, těží nerostné suroviny, pěstují různé druhy plodin apod. Informace o tom, kde se jednotlivé lidské aktivity odehrávají, má velký význam. Právě proto, že má lokalizace pro uskutečňování většiny lidských činností takový význam, je potřeba ji brát při jejich řešení a plánování v úvahu.“*

Prostorová data obsahují informace o jejich umístění a atributovou informaci. Data, které mají zaznamenané atributové informace s rozdílnou lokací, které jsou přímo součástí dat. Pro účely porozumění procesům, které generují pozorované atributové hodnoty, slouží prostorová analýza dat, která využívá informace o umístění. Prostorová analýza je důležitým nástrojem z toho důvodu, že většinu dat lze označit za prostorová. Nezbytným nástrojem pro spravování a zpracování prostorových dat jsou geografické informační systémy (GIS) (Fotheringham S. , Brunson, Charlton, & Rogerson, 2009).

2.2 Geografické informační systémy

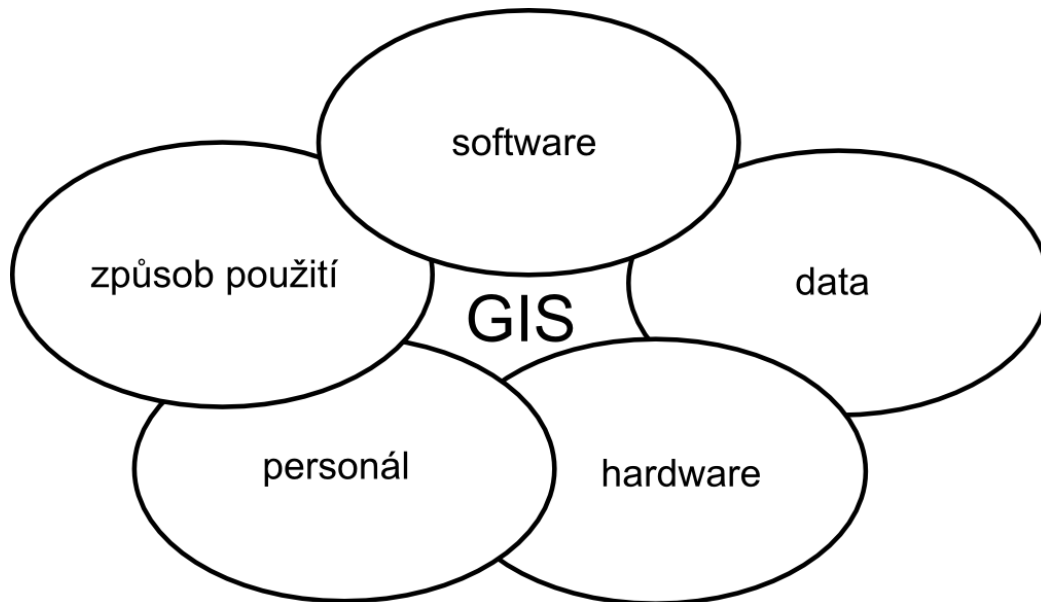
Jednoznačně definovat geografické informační systémy (GIS - Geographic Information Systems) není jednoduché. Pojem GIS reprezentuje integraci mnoha oborů. Proto výkladů, které definují geografické informační systémy je vícero.

Výrobce programu ArcGIS společnost ESRI (2015), definuje GIS jako *„počítačově založený nástroj pro mapování a analyzování záležitostí a událostí, které se staly na Zemi. Technologie GIS integruje běžné databázové operace, jako je dotaz a statistická analýza s unikátní vizualizací a geografickou analýzou z nabízených map“*

Nezávislá vládní agentura United States Geological Survey (2007) má opět rozdílný pohled na GIS: *„GIS je počítačový program, který sestavuje, uchovává, manipuluje a zobrazuje geograficky umístěná data. Údaje, které jsou identifikovány svým umístěním. Provozovatelé také považují celkově GIS jako systém obsahující zúčastněné osoby a data, které jdou do systému.“*

Všechny výše psané definice mají pět základních společných složek: software, data, hardware, personál a způsob použití dat.

Obrázek 1: Základní složky GIS

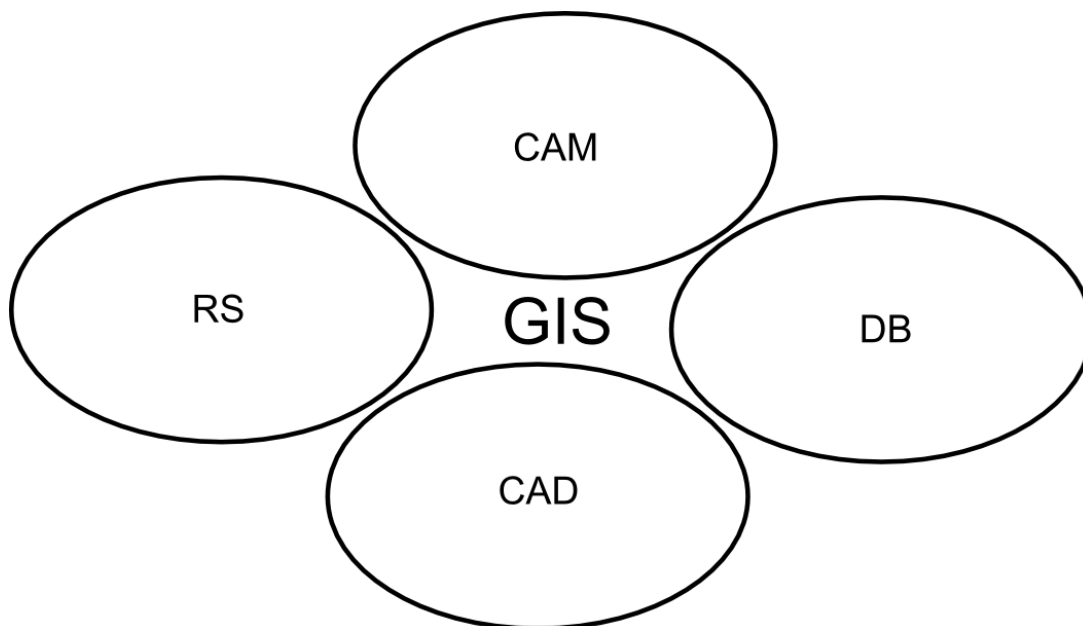


Zdroj: Klufová, Rost, & Klicnarová (2012)

GIS má mnoho způsobů využití. Příkladem využití může být katastrální a mapové plánování ve státní správě, využití v oblasti ochrany životního prostředí, tematické kartografii, zemědělství, urbanistickém plánování, vojenství, matematických studiích, obchodu, zdravotnictví a dálkovém průzkumu Země.

Rozhodně lze říci, že GIS není počítačovým systémem pro tvorbu map, ani CAD programem. GIS je vztažený k dalším složkám informačních technologií, kterými jsou CAD - počítačově podporované kreslení, CAM - počítačově podporované modelování, DB - databázové systémy, RS - dálkový průzkum.

Obrázek 2: Vztah GIS k ostatním informačním technologiím



Zdroj: Klufová, Rost, & Klicnarová (2012)

2.2.1 ArcGIS

Pro účely vyhotovení této práce poslouží mimo jiné program ArcGIS od společnosti ESRI. ArcGIS je sada integrovaných aplikací, které se skládají z:

- ArcMap – hlavní mapová aplikace, která umožňuje tvorbu map, atributových dotazů, analýz prostorových vztahů a tvorbu finálních výstupů.
- ArcCatalog – organizuje prostorová data a umožňuje prohlížení, vyhledávání a přidávání dat
- ArcToolbox – obsahuje nástroje pro zpracování geodat, datovou konverzi, mapové projekce a souřadnicové systémy.

2.3 Regresní analýza

Regrese je označení pro širokou škálu metod určených pro modelování vztahu mezi závislými proměnnými a sadou jedné nebo více nezávislých proměnných. Proměnné y označují závisle proměnné. Nezávislé proměnné jsou někdy označovány jako x -proměnné, nebo také jako prediktory. Regresní model se vyjadřuje rovnicí (Fortheringham & Charlton, 2009).

Regresní analýza rozlišuje mezi dvěma typy regresních modelů. Prvním typem regresního modelu je model lineární, druhým typem regrese je model nelineární. Za nelineární regresní modely jsou považovány takové modely, jejichž výstup první derivace regresní funkce dá opět funkci jednoho, nebo více regresních parametrů (Klufová, Rost, & Klicnarová, 2012).

Nejjednodušší forma lineární regresního modelu je:

$$y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \quad \text{pro } i=1 \dots n, \quad (\text{I.})$$

kde:

- nezávislou proměnnou v této rovnici představuje y_i , zde měřené v některém místě i , další nezávislou proměnnou je x_i ,
- ε_i představuje výraz pro chybu,
- β_0 a β_1 představují parametry.

Parametry β_0 a β_1 musí být odhadnuty tak, že hodnoty $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ jsou minimalizovány přes n pozorování v datovém souboru. Hodnota \hat{y}_i je předpokládané, nebo přesně stanovená hodnota pro i -té pozorování vztažené k i -té hodnotě x (Fortheringham & Charlton, 2009).

2.3.1 Metoda nejmenších čtverců

Rost (2012) o metodě nejmenších čtverců tvrdí, že „*k odhadu neznámých parametrů, které zpravidla označujeme za regresní koeficienty, používáme nejčastěji tzv. metodu nejmenších čtverců. Podstata této metody spočívá na volbě odhadu jednotlivých složek vektoru θ tak, aby byl součet čtverců residuí minimální.*“

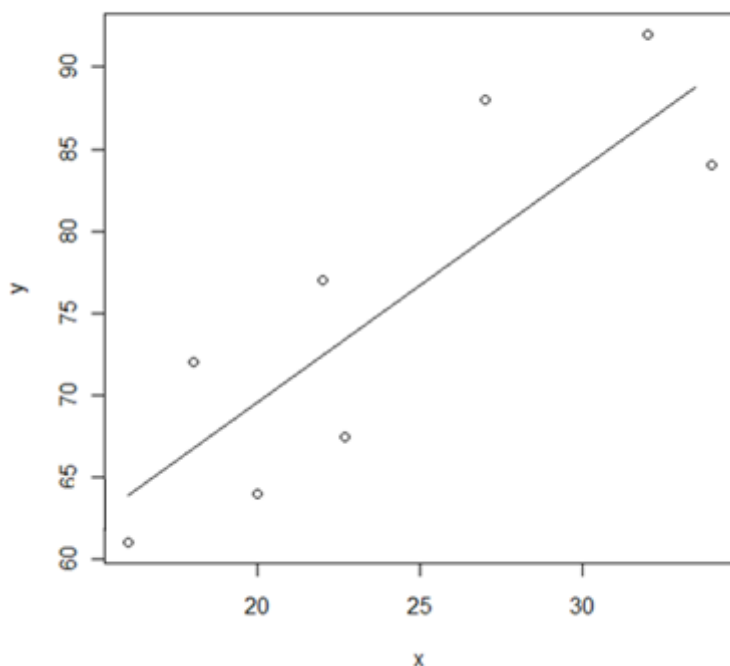
V případě, že chápeme součet čtverců jako funkci neznámých parametrů, lze následující funkci formulovat jako:

$$Q(\theta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon \rightarrow \min. \quad (\text{II.})$$

Úloha musí splňovat podmínku:

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

Obrázek 3: Znárodnění metody nejmenších čtverců



Zdroj: Bogard (2011)

2.3.2 Geograficky vážená regrese

Geograficky vážená regrese (GWR - Geographically Weighted Regression) bere při vytváření modelu v potaz pozorování pro vytváření modelu s ohledem na jejich umístění. Vícenásobná lineární regrese zahrnuje analýzu v globálním měřítku. Geograficky vážená regrese představuje analýzu na lokální úrovni. Regresní analýza může vytvářet rozdílné výsledky v závislosti na prostorové referenční oblasti aplikace. U globálních parametrů jsou veškeré místní zvláštnosti přehlédnuty (Fotheringham, Brunson, & Charlton, 2002). Metoda geograficky vážené regrese rozšiřuje metodu nejmenších čtverců lineárního regresního modelu pro počítání prostorové struktury a odhaduje samostatný model a místní parametry pro každou geografickou polohu dat založené na místní podmnožině dat používajících rozdílnou váhu schémat. Model geograficky vážené regrese může být zapsán ve tvaru:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^k \beta_j(u_i, v_i) x_{ij} + \varepsilon_i, \quad (\text{III.})$$

kde:

- hodnota y_i je výsledná proměnná v souřadnicové pozici,
- i představuje konkrétní pozici o souřadnicích (u_i, v_i)
- hodnoty β_0 a β_j jsou označením pro místní odhadovaný průsečík a vliv proměnné i na konkrétní pozici.

Odhady v modelu geograficky vážené regrese lze získat na kterémkoliv místě, což umožňuje vytvoření mapy znázorňující souvislé plochy hodnot parametrů a posouzení prostorové variability těchto parametrů (Matthews & Yang, 2012).

Kalibrace modelu předpokládá, že pozorovaná data blízko pozorovaného místa mají větší vliv na hodnocení u_i a v_i , než data, která jsou umístěna dále od i . Rovnice měří vztahy plynoucí z modelu kolem každé pozice i . Metoda nejmenších čtverců dává náhled pro to, jak geograficky vážená regrese funguje. V modelu geograficky vážené regrese je hodnota ve vztahu s jeho blízkostí k umístění. Matice vah pozorování již nejsou konstantní, ale liší se podle i . Vzdálenější data mají menší váhu než z data z bližšího pozorování (Fotheringham, Brunson, & Charlton, 2002).

Z toho vyplývá rovnice:

$$\hat{\beta}(u_i, v_i) = (X^T W_{(u_i, v_i)} X)^{-1} X^T W(u_i, v_i) y, \quad (\text{IV.})$$

kde:

- $\hat{\beta}$ představuje odhad β ,
- $W(u_i, v_i)$ jsou matice vah, která mají specifika pro body i .

$W(u_i, v_i)$ matice se zapisuje ve tvaru:

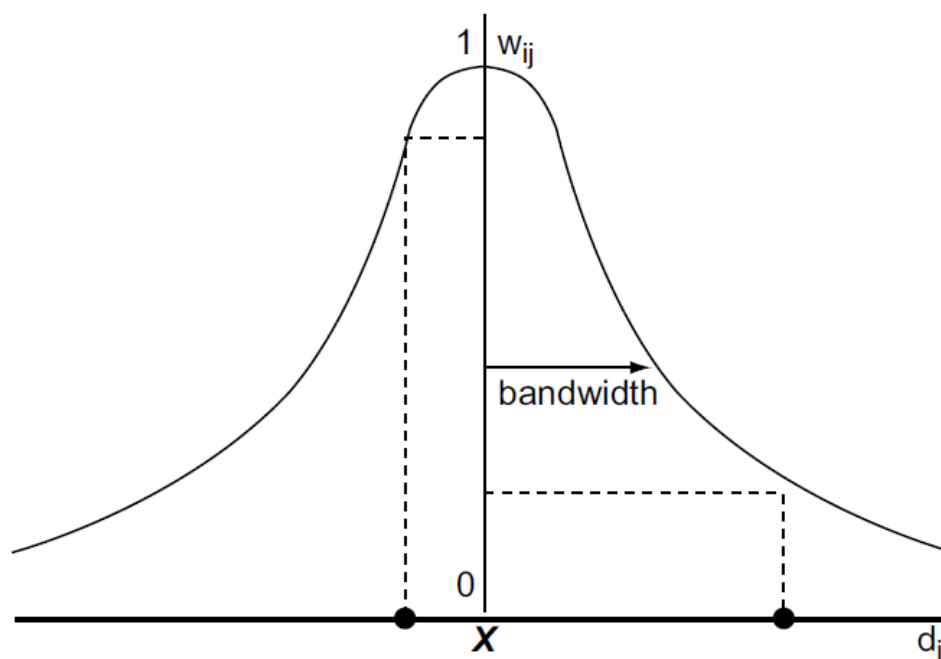
$$W(u_i, v_i) = \begin{pmatrix} w_{i1} & 0 & 0 & \dots & 0 \\ 0 & w_{i2} & 0 & \dots & 0 \\ 0 & 0 & w_{i3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & w_{in} \end{pmatrix}, \quad (\text{V.})$$

kde:

- w_{in} představuje váhu bodu n pro odhadnutí lokálních parametrů v místě i ,
- váhy w_{ij} se odlišují podle i .

Právě toto odlišuje geograficky váženou regresi od klasické regrese, v klasické regresi je váha matic konstantní. K váze hodnoty pozorované v lokalitě j ve vzdálenosti od bodu i se vztahuje celá řada existujících možných funkcí vah. Funkce vah mají tendenci blížit se ke Gaussově funkci odrážející charakter mnoha prostorových procesů. Geograficky vážená regrese zohledňuje lokalizované verze všech parametrů regresní diagnostiky a to i s koeficientem determinace. U globálního modelu klasické regrese mají všechna pozorování váhu 1, což znamená, že odhadované parametry nemají žádné prostorové změny (Klufová, Rost, & Klicnarová, 2012).

Obrázek 4: Schéma fungování funkce prostorových vah



Zdroj: Fotheringham, Brunson, & Charlton (2002)

Obrázek 4 popisuje výše zmíněné fungování prostorových vah, kde

- x je bod regresní analýzy,
- w_{ij} představuje váhu datové bodu j v regresním bodě i ,
- j je datovým bodem,
- d_{ij} je vzdálenost bodu i od regresního bodu j .

2.4 Prostorové váhy

V globálním modelu má každé pozorování jednotkovou váhu. Prvním krokem ke stanovení prostorových vah při kalibraci modelu je vyloučit taková pozorování, která se nacházejí dále než stanovená hodnota d od regresního bodu. Nejjednodušší prostorovou funkci vah lze zapsat ve tvaru:

$$w_{ij} = 1 \text{ pokud } d_{ij} < d \quad (\text{VI.})$$

$$w_{ij} = 0 \text{ jinak}$$

Výše uvedené váhové schéma by značně zjednodušilo postup kalibrace, protože v každém regresním bodě by mohla být podmnožina výběrových bodů ke kalibraci modelu. Tato funkce má však problém nespojitosti. Při změnách i může v okolí regresního bodu

docházet k drastickým změnám odhadovaných koeficientů. Přestože se náhle změny prostorových parametrů skutečně objevují, tak v tomto případě se změny odhadů mohou objevit spíše v důsledku uspořádání výběrových bodů než podpůrných procesů (Klufová, Rost, & Klicnarová, 2012).

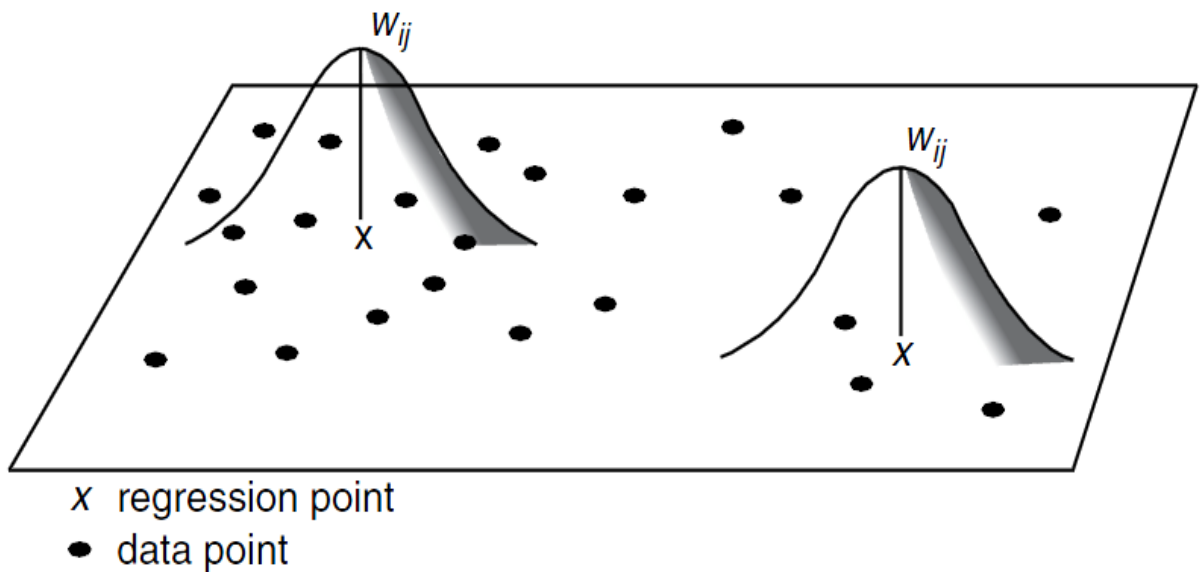
Podle Fotheringhama (2009) je jedním ze způsobů jak vyřešit problém nespojitosti definování vah w_{ij} jako spojité funkce d_{ij} . Funkce je pak zapsána ve tvaru:

$$w_{ij} = \exp[-1/2 (d_{ij}/b)^2], \quad (\text{VII.})$$

kde b označuje šířku pásma. Pakliže se i a j shodují, váha bodu je rovná 1. U dalších bodů se váhy budou snižovat podle Gaussovy křivky s rostoucí vzdálenosti mezi i a j .

Přestože přesná funkce prostorových vah může být vyjádřena v mnoha tvarech, tak se používá dvou rozšířených typů váhových funkcí. Těmi jsou fixní (viz Obrázek 5) a adaptivní (viz Obrázek 6). Prostorově vážená funkce s fixní šířkou jádra je konstantní v celém rozsahu zkoumané oblasti. Tato funkce má nežádoucí vlastnost, že v oblastech řídkého rozmístění datových bodů budou mít výsledné lokální odhady parametrů vysokou míru chyb, která je příčinou nejistoty odhadů způsobených nedostatkem dat u některých regresních bodů (Fotheringham S. , Brunson, Charlton, & Rogerson, 2009).

Obrázek 5: Prostorově vážená funkce s fixní šířkou jádra



Zdroj: Fotheringham, Brunson, & Charlton (2002)

Další možností je použití prostorově vážené funkce s adaptivní šířkou jádra. Tato převážně preferovaná funkce prostorových vah má relativně malou šířku pásma kolem regresního bodu, když jsou datové body kolem něj hustě rozšířené. Tam, kde jsou datové body řidčeji rozmístěné, je pásmo širší, protože jádro se rozšiřuje, aby zahrnulo více datových bodů.

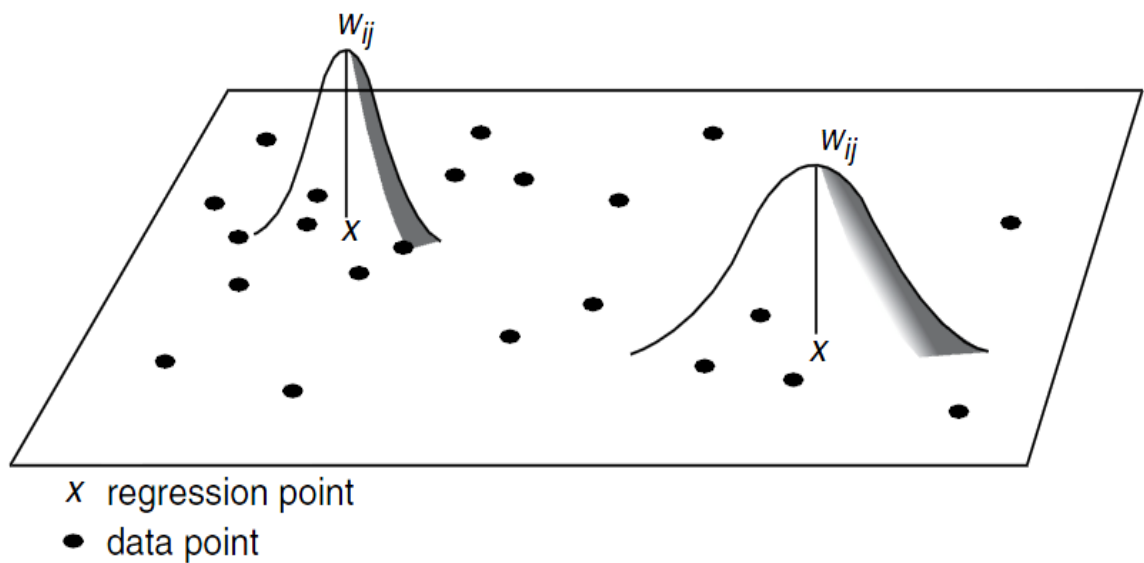
Existuje vícero funkcí, které lze použít k sestavení prostorově vážené funkce s adaptivní šířkou jádra a funkci lze zapsat také následovně:

$$w_{ij} = [1 - (d_{ij}^2 / h^2)]^2 \text{ pokud } j \text{ je jedno z } N - \text{ tého nejblížešího } i, \quad (\text{VIII.})$$
$$w_{ij} = 0 \text{ jinak,}$$

kde:

- h představuje šířku pásma,
- N je parametr, který je odhadován.

Obrázek 6: Prostorově vážená funkce s adaptivní šířkou jádra



Zdroj: Fotheringham, Brunson, & Charlton (2002)

Fotheringham (2009) o GWR říká, že „výsledky GWR se mohou zdát relativně necitlivé na volbě funkce prostorové váhy, pokud je funkce vzdálenostně spojitá. Při použití kterékoliv váhové funkce bude nějaký výsledek. Nicméně výsledek je citlivý na stupni vzdálenosti - rozkladu. Proto musí být získána optimální hodnota h nebo N . To lze nalézt minimalizací křížově-validačního skóre (CV) nebo Aikakova informačního kritéria $AICc$ “.

$$CV = \sum_i [y_i - \hat{y}_i^*(h)]^2, \quad (\text{IX.})$$

kde $\hat{y}_i^*(h)$ je předpovídaná hodnota y_i s daty z bodu i vynechaným z kalibrace.

$$AICc = \text{odchylka} + 2k \left(\frac{n}{n-k-1} \right), \quad (\text{X.})$$

kde:

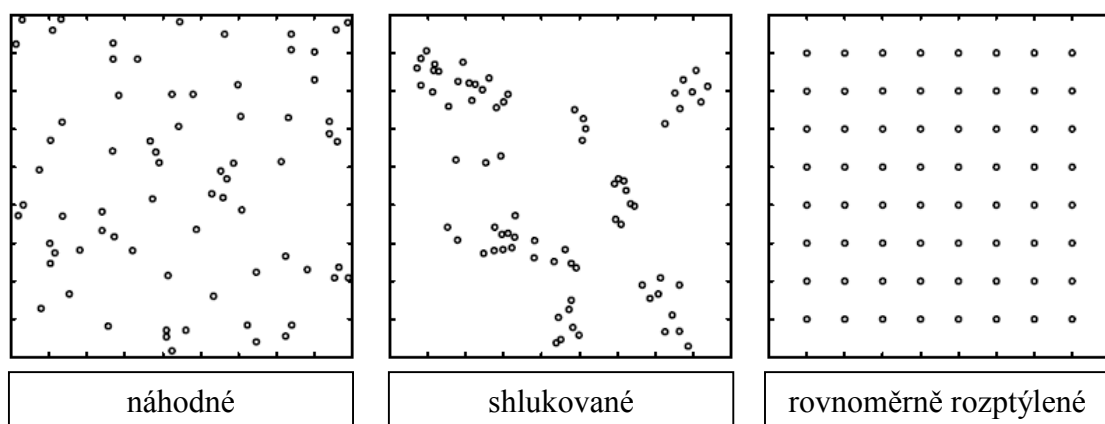
- n je počet datových bodů,
- k je počet parametrů v modelu.

Nižší hodnoty obou statistik ukazují vyšší přesnost modelu. Ideální šířka pásma je kompromisem mezi zkreslením a rozptylem. Příliš malá šířka pásma vede k vyššímu rozptylu v lokálních odhadech z důvodu nižšího počtu datových bodů použitých při místní kalibraci. Naopak příliš velká šířka pásma může znamenat vyšší zkreslení v lokálních odhadech z toho důvodu, že data jsou čerpána ze vzdálenějších míst od regresního bodu. Když se šířka pásma blíží k nekonečnu, tak lokální model bude mít tendenci stát se globálním modelem s počtem parametrů k . V případě, kdy se šířka pásma blíží nule, se lokální model stáhne kolem dat s počtem n parametrů. Proto se počet parametrů v lokálních modelech pohybuje mezi k a n a závisí na šířce pásma. Počet parametrů nemusí být celé číslo a považuje se za efektivní počet parametrů v modelu (Netrdová & Nosek, 2009).

2.5 Prostorové shlukování

Seskupování souborů prostorových dat do shluků, kde si objekty jsou ve shluku podobné, a zároveň jsou rozdílné od objektů jiných shluků, nazýváme prostorovým shlukováním. Prostorové shlukování může být použito na seskupení podobných prostorových objektů. Předpokladem je, že prostorové struktury mají tendenci se spíše seskupovat nežli náhodně rozmisťovat. Měla by být změřena statistická významnost prostorových shluků za pomoci zkoušky předpokladů dat. Zkouška statistické významnosti dat je nezbytností před přistoupením ke shlukové analýze (Fotheringham S. , Brunson, Charlton, & Rogerson, 2009).

Obrázek 7: Základní typy prostorového shlukování dat



Zdroj: Fotheringham S. , Brunson, Charlton, & Rogerson (2009)

Fotheringham (2009) o náhodných strukturách píše, že „*struktura CSR (complete spatial randomness) je synonymem pro homogenní Poissonův proces. Struktury procesu jsou nezávisle a rovnoměrně rozloženy v prostoru, mohou se objevit kdekoliv a vzájemně na sebe nepůsobí. Nicméně struktury generované náhodně mohou být shlukované, nebo rovnoměrně rozptýlené.*“

Prostorové shlukování je statisticky významnější, když data vykazují shlukovanou strukturu, než náhodnou nebo rovnoměrně rozptýlenou. Pro kvantifikaci odchylek struktur náhodného shlukování může být použito několika metod. První typ deskriptivní statistiky je založen na kvadrátech, což jsou dobře definované oblasti, které často mají obdélníkový tvar. Dalším typem statistiky je metoda založená na vzdálenostech mezi vzorky. Takovou statistikou je třeba Ripleyho K-funkce.

Problémem shlukování označených procesů prostorových bodů je zobecnění problému prostorových shluků. Místo jednoho prostorového umístění pro každou kategorii existuje umístění vícero. Klasické shlukové postupy zvládají homogenní prostorové procesy, a proto nemůžou shlukovat označené procesy prostorových bodů. Existuje technika pro získávání dat ze shlukově označených bodů, která je založena na poznání, že podobnost uvnitř shluku musí být signifikantně vyšší, než mezishluková podobnost. Besagova L-funkce je upravenou Ripleyho K-funkcí a je používána k měření korelace mezi sledovanými a očekávanými páry bodů v určité vzdálenosti od sebe (Fotheringham S. , Brunson, Charlton, & Rogerson, 2009).

2.6 Prostorová autokorelace

Netrdová a Nosek (2009) tvrdí, že „*prostorová autokorelace, neboli korelace jednoho jevu se sebou samým v prostoru, je jedním ze základních aspektů prostorových dat, který se projevuje statisticky významným uspořádáním hodnot jevu v prostoru. Na prostorovou autokorelaci lze přitom nahlížet jednak jako na problém analýzy prostorových dat, neboť porušuje základní předpoklad mnoha standardních parametrických statistických testů, že jednotlivá pozorování jsou navzájem nezávislá, jednak jako na nástroj analýzy. Pomocí prostorové autokorelace lze popsat určitou vzájemnou velikostně-prostorovou všesměrnou závislost, což lze využít zejména při určování míry shlukování sledovaného jevu.*“

2.6.1 Globální míry

Jedním z nejpoužívanějších ukazatelů je Moranův index I prostorové autokorelace, který je popsán vzorcem:

$$I = \frac{N \sum_i \sum_j W_{ij} (X_i - \bar{X})(X_j - \bar{X})}{(\sum_i \sum_j W_{ij}) \sum_i (X_i - \bar{X})^2}, \quad (\text{XI.})$$

kde:

- N je počet analyzovaných jednotek,
- X_i je hodnota proměnné v konkrétním umístění,
- X_j je hodnota proměnné v dalším umístění,
- X je aritmetický průměr sledované proměnné,
- W_{ij} je prostorovou vážící funkcí.

Hodnota Moranova indexu I nabývá hodnoty od -1 do 1. Negativní autokorelaci představuje záporná hodnota, pozitivní autokorelace odpovídá kladné hodnotě. Kladná prostorová autokorelace znamená, že ve sledované oblasti dochází k prostorovému shlukování hodnot, naopak záporná hodnota znamená šachovnicový vzorek hodnot (Klufova, Rost, & Klicnarová, 2012).

Dalším používaným ukazatelem je Gearyho index C , který je podobný Moranovu I . Rovnice má tvar:

$$C = \frac{[(N-1)] [\sum_i \sum_j W_{ij} (X_i - X_j)^2]}{2 (\sum_i \sum_j W_{ij} (X_i - \bar{X})^2)}, \quad (\text{XII.})$$

kde:

- X_i je hodnota proměnné v umístění i ,
- X_j je hodnota proměnné v dalším umístění,
- W_{ij} je prostorovou vážící funkcí.

Rozdíl mezi Moranovo indexem I a Gearyho indexem C je porovnávání atributů. Moranovo index porovnává hodnoty atributů pozorovaných umístění jako odchylky od průměru. Gearyho index porovnává atributy mezi sebou. Hodnoty Gearyho indexu se obvykle pohybují mezi 0 a 2. Hodnota menší než 1 znamená pozitivní prostorovou

autokorelaci, hodnota vyšší než 1 znamená negativní prostorovou korelaci. Pokud je hodnota rovná 1, je prostorová korelace nulová.

2.6.2 Lokální míry

Při výpočtu lokálního kritéria se vychází z předpokladu homogenity. Pokud nedojde ke splnění tohoto předpokladu, je použití jednoho globálního kritéria bezvýznamné, protože se může v prostoru měnit. Lokální statistiky vztahující se ke konkrétnímu místu je vhodné použít i v případě neprokázání globální prostorové autokorelace. Zkratka LISA je souhrnné označení pro lokální indikátory prostorové autokorelace - Local Indicators of Spatial Association (Klufova, Rost, & Klicnarová, 2012).

Vzorec lokálního Moranova indexu I má tvar:

$$I_i = z_i \sum_j w_{ij} z_j \quad (\text{XIII.})$$

Pozorování z_i a z_j jsou odchylky od průměru. Vysoké hodnoty indexu znamenají shlukování podobných hodnot, naopak nízké hodnoty znamenají shlukování rozdílných hodnot.

Pro výpočet lokálního Gearyho indexu se používá vzorec:

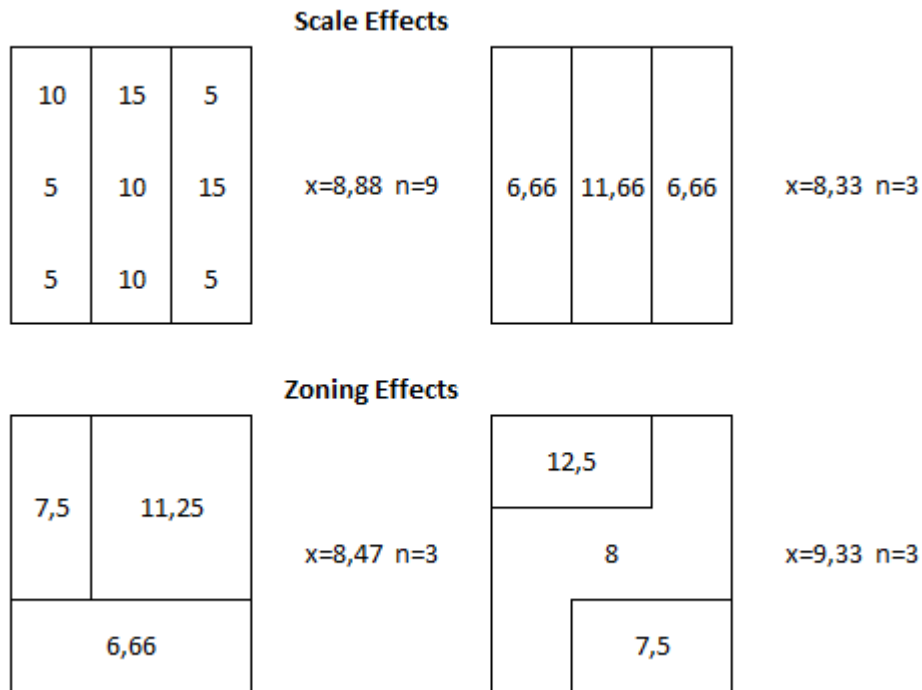
$$c_i = \sum_j w_{ij} (z_i - z_j)^2 \quad (\text{XIV.})$$

Vysoké hodnoty indexu představují shlukování rozdílných hodnot, nízké hodnoty indexu znamená shlukování blízkých hodnot.

2.7 MAUP – Modifiable Area Unit Problem

Problém modifikovatelné územní jednotky – MAUP se skládá ze dvou samostatných problémů, které jsou však úzce provázané. Openshaw (1983) popisuje jako první z nich „problém rozsahu (scale effect), který způsobuje rozdíly ve výsledcích, které mohou být často získány, když data pro jednu sadu územních jednotek jsou postupně agregovány do několika menšího počtu větších územních jednotek. Příkladem je, když jsou sčítací čtvrti agregovány do obvodů a okresů, pak se výsledky mění se vzrůstajícím rozsahem území.“

Obrázek 8: MAUP – scale effects a zoning effect



Zdroj: Openshaw (1983)

O druhém problému MAUP se Fotheringham (2009) rozepisuje „*když počet místních jednotek je fixní nebo relativně stálý, ale hranice jsou překresleny, aby se přizpůsobily změnám, tak se v podstatě jedná o územní proces. Data získaná na základě různých územních systémů stejné oblasti nám dávají různá vyobrazení regionu a různé analytické výsledky. Nekonzistence výsledků na základě údajů z různých územních systémů je známý jako územní problém.*“

Problematika MAUP je zachycena na Obrázek 8. Horní polovina zachycuje efekt rozsahu, kdy jsou jednotky shlukovány to jednotek vyšších, dolní polovina pak územní efekt, kdy dochází ke změně hranic. Proměnná x zobrazuje průměr a n počet pozorování.

2.8 Příklad použití geograficky vážené regrese

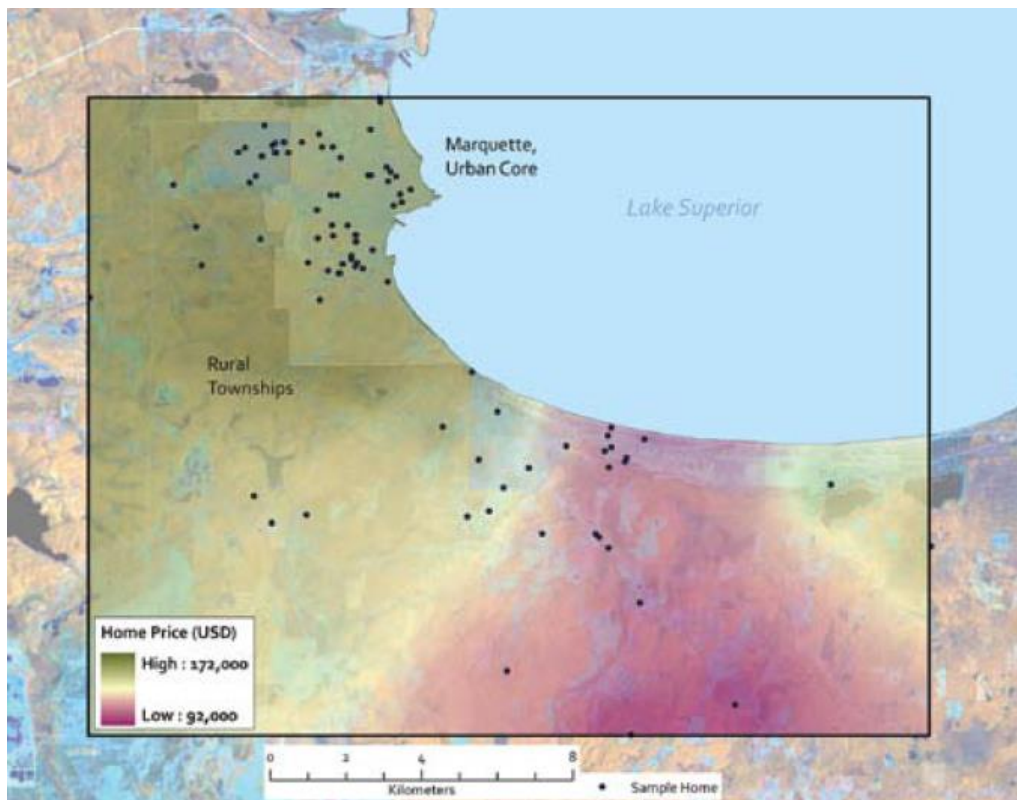
Robert Legg a Tia Bowe z Northern Michigan University (2009) použili metodu GWR pro analyzování prodejních cen rodinných domů v závislosti na jejich umístění a dalších souvisejících proměnných. Před aplikováním GWR byla použita metoda lineární regrese. Při aplikaci autoři zjistili omezení vyplývající z lineárního regresního modelu, který často

nadhodnocoval či podhodnocoval ceny v městských čtvrtích. Proto autoři použili GWR, aby zvýšili přesnost modelu a vyřešili některé z residuálních chyb.

Pro tvorbu modelu byl použit vzorek 93 domácností, jejichž poloha byla zaznamenána v souřadnicovém systému UTM. Ve výčtu parametrů ovlivňujících cenu byl zaznamenán počet obytných místností, rozloha domu a pozemku. Pro tvorbu globálního modelu byla použita lineární regrese. Globální model byl použit k předvídání cen domů v Marquette ve státě Michigan. Následně bylo na stejném datovém souboru použito metody GWR, která dala samostatné regresní koeficienty pro každý z 93 sledovaných domů.

Lineární regresní model byl vyhodnocen jako signifikantní se silným koeficientem determinace v hodnotě 0,782. Použití metody GWR model výrazně zlepšilo. Koeficient determinace vzrostl na hodnotu 0,865. Koeficient determinace nabývá hodnot od 0 do 1 a platí, že vyšší hodnoty znamenají větší úspěšnost regrese. Použití GWR namísto lineárního modelu přineslo pokles residuální chyby o 160 000 dolarů. Koeficienty ploch vytvořené za použití GWR byly užitečné pro identifikaci prostorových struktur ve sledované oblasti.

Obrázek 9: Prodejní ceny domů modelovaných pomocí metody GWR v ArcGISu



Zdroj: Legg & Bowe (2009)

Hodnoty koeficientů ukázaly, že ceny rostou s rozlohou pozemku. Dále koeficienty ukázaly, že rostoucí obytná plocha přispívá méně k růstu ceny než plocha pozemku. Tyto struktury mohou být vypovídající o stáří domů ve venkovských oblastech. Domy blíže k centru bývají starší, naopak domy na periferii bývají novostavby, avšak ceny pozemků nedosahují takových hodnot, jako pozemky umístěné blíže k centru. Novější domy mají vyšší prodejní cenu za metr čtvereční a jsou také zatíženy vyšší sazbou daně z nemovitosti, než starší domy. Daňové sazby ve městě jsou přibližně dvakrát tak vysoké, než daně na venkově, což může developery odradit od výstavby v městských jádrech (Legg & Bowe, 2009).

Dalším příkladem využití GWR může být práce autorů Nissi, Sarra, & Palermi (2013), kteří aplikovali GWR ke zjištění prostorové nestacionarity vztahu mezi koncentrací radonu v obydlích a radioaktivním obsahem v půdě v italské Provincie of L'Aquila. Jako data posloužila měření radonu v 455 obydlích. Autoři použili v GWR adaptivní vážené schéma a společnou bisquare funkci pro výpočet vah. Po přezkoumání kolinearit zůstal v modelu pouze uran jako vysvětlující proměnná. Ostatní prediktory nebyly signifikantní. Lokální koeficienty determinace ukazovaly hodnoty od 0,13 do 0,47, což znamenalo posun v porovnání s metodou OLS.

2.9 Výstupy geograficky vážené regrese

Za hlavní výstupy geograficky vážené regrese lze považovat soubor lokálně specifických parametrických odhadů, které mohou při zmapování a zanalyzování přinést informace o vztazích a jejich prostorové nestacionaritě. V geograficky váženém formátu lze uvést kteroukoliv formu regrese, takže v geograficky vážené regresi lze:

- hodnotit místní standardní chyby,
- vypočítat míru lokálního vlivu,
- získat místní *t*-statistiky,
- provádět testy pro vyhodnocení významu prostorových změn v odhadech místních parametrů,
- provádět testy pro vyhodnocení, zda poskytuje lepší výsledky lokální nebo globální model, což představuje rozdíly ve stupních volnosti (Fotheringham S. , Brunsdon, Charlton, & Rogerson, 2009).

3 Metodika

3.1 Tvorba databáze

Pro účely práce je nezbytné vytvořit databázi obsahující níže vypsané proměnné. Data byla získána ze stránek ČSÚ a ze sčítání lidu 2011. Další data jsem získal z databáze ArcČR® 500 ve verzi 3.2. Následně jsem upravoval data ve vztahu k sledovaným jednotkám v programu ArcMap. Pro tvorbu a správu databáze jsem použil programy MS Excel a ArcMap 10.0.

3.2 Vstupní proměnné

Za závisle proměnnou regresního modelu byl zvolen ukazatel hrubé míry migračního salda. Jako prediktory v modelu posloužily ukazatele nezaměstnanosti, počtu bytů, hustoty zalidnění, hustoty silniční sítě a koeficientu ekologické stability. Rovněž jako prediktory posloužily ukazatele průměrného věku, indexu ekonomického zatížení a podílu produktivního věku, které byly vyřazeny z důvodu vysoké hodnoty VIF.

3.2.1 Hrubá míra migračního salda

Rozdíl mezi počtem přistěhovalých a počtem nastěhovalých na určitém území se nazývá migrační saldo. Hrubá míra migračního salda je migrační saldo na 1000 obyvatel středního stavu a lze jej zapsat ve tvaru:

$$HMMS = \frac{I_t - E_t}{P_t} \cdot 1000, \quad (\text{XV.})$$

kde:

- I_t je počet přistěhovalých v časovém intervalu t ,
- E_t je počet vystěhovalých v časovém intervalu t ,
- P_t je střední stav počtu obyvatel.

3.2.2 Nezaměstnanost

Data pro registrovanou nezaměstnanost byla čerpána ze zdrojů ČSÚ (2015), jehož metodika platná od 1. 7. 2004 popisuje registrovanou nezaměstnanost jako „*podíl, kde v čitateli je počet dosažitelných neumístěných uchazečů o zaměstnání a ve jmenovateli součet počtu zaměstnaných z VŠPS¹, pracujících cizinců registrovaných na úřadech práce nebo s platným povolením k zaměstnávání či živnostenským oprávněním a počtu*

¹ Výběrové šetření pracovních sil

dosazitelných neumístěných uchazečů o zaměstnání (klouzavý roční průměr)“ (Český statistický úřad, 2015).

3.2.3 Počet bytů

Poměrový ukazatel počtu bytů ukazuje, kolik bytů připadá na 1000 obyvatel na konkrétním území. Bytový fond je tvořen rodinnými i bytovými domy.

3.2.4 Hustota zalidnění

Stavová veličina hustoty zalidnění je ukazatelem, který zachycuje koncentraci obyvatelstva na konkrétním území k určitému okamžiku. V roce 2011 byla hustota zalidnění 133 obyvatel na čtvereční kilometr.

3.2.5 Hustota silniční sítě

Poměrový ukazatel hustoty silniční sítě zachycuje hustotu pozemních komunikací na čtvereční kilometr. Za zdroj dat posloužila databáze ArcČR 500, stav k 1. 1. 2012. Za vstupní data pro silniční síť byly definovány dálnice, rychlostní silnice, silnice I., II. a III. třídy a neevidované silnice.

3.2.6 Koeficient ekologické stability

Koeficient ekologické stability je poměrovým ukazatelem, který zachycuje poměr ploch stabilních a nestabilních ekosystémů na určitém území. Výpočet KES je založen na selekci krajinného typu a jeho zařazení do stabilního, či nestabilního ekosystému (Míchal, 1994).

Vzorec pro výpočet je zapsán ve tvaru:

$$KES = \frac{LP + VP + TTP + Pa + Mo + Sa + Vi}{OP + AP + Ch}, \quad (XVI.)$$

kde:

- v čitateli stabilních ekosystémů *LP* – lesní půda, *VP* – vodní plochy, *TTP* – trvalé travní porosty, *Pa* – pastviny, *Mo* – mokřady, *Sa* – sady, *Vi* – vinice,
- ve jmenovateli nestabilních ekosystémů *OP* – orná půda, *AP* – antropogenizované plochy, *Ch* – chmelnice.

3.3 Řešená území

Regresní modely budou řešeny na úrovních obcí, ORP, okresů a krajů. Ačkoliv území hlavního města Prahy² nelze legislativně zařadit do všech řešených územních úrovní, pro úplnost tomu tak v regresních modelech bylo. Pro administrativní dělení České republiky byla použita data ArcČR® 500 ve verzi 3.2 od společnosti ARCDATA PRAHA s.r.o., která jsou distribuována zdarma.

3.3.1 Obce

Vznik a zánik obcí řeší Zákon o obcích³, kdy obec může vzniknout třemi způsoby a to: sloučením dvou nebo více obcí, které spolu soused, oddělením části obce od obce a zrušením nebo změnou vojenského újezdu. Obec je nejmenší základní územní jednotkou České republiky na úrovni statistické jednotky LAU 2. K 1. 1. 2013 měla ČR 6253 obcí.

3.3.2 Obce s rozšířenou působností

Obce s rozšířenou působností jsou také označovány jako obce III. typu, které vykonávají státní správu v přenesené působnosti. ORP vznikly k 1. 1. 2003 po zrušení okresních úřadů, kdy převzaly část agendy okresních úřadů. Celkem je v ČR na 205 obcí s rozšířenou působností.

3.3.3 Okresy

Okres je označením územně správní jednotky na statistické jednotce LAU 1. Okresní úřady zanikly k 31. 12. 2002 během reformy územní správy. Od roku 1996 má ČR 76 okresů.

3.3.4 Kraje

Kraje jsou vyšším územně správním celkem. Vznikly k 1. 1. 2000 a získaly pravomoc na základě zákona o krajích⁴. Kraje jsou na úrovni statistické jednotky NUTS 3. Celkem existuje 13 krajů.

² Působnost hlavního města Prahy řeší zákon č. 131/2000 Sb., o hlavním městě Praze

³ Zákon č. 128/2000 Sb.

⁴ Zákon č. 129/2000 Sb.

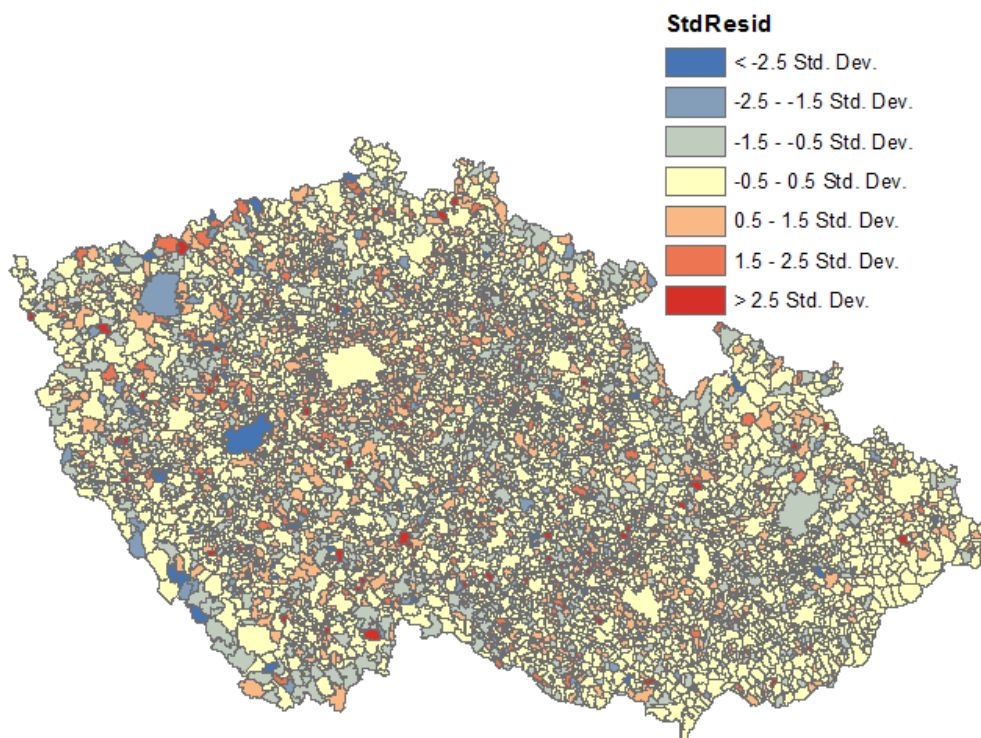
4 Aplikace vybraných metod

4.1 Regresní analýza

4.1.1 Obce – globální model

Pro globální regresní model bylo užito běžné metody nejmenších čtverců. Jak je zachyceno v Tabulce 2, hodnota koeficientu determinace je 0,1669 a korigovaného koeficientu determinace je 0,1663, což jsou dosti malé hodnoty. Přidáním další proměnné by bylo možné zlepšit výkonnost modelu, nicméně charakter a dostupnost dat toto řešení neumožňuje. Výkonnost daného modelu snižuje vysoký počet území nejednotné velikosti, kde jsou zahrnuty i obce o několika málo obyvatelích. Nejmenší jednotkou z hlediska počtu obyvatel je obec Vysoká Lhota s 18 obyvateli. Jakákoliv změna počtu obyvatel jmenované obce bude znamenat vysokou procentuální změnu oproti jednotkám s vyšším počtem obyvatel a tím výrazněji ovlivní regresní model.

Obrázek 10: Standardizovaná residua globálního modelu na úrovni obcí



Zdroj: Vlastní zpracování v ArcGIS 10.0

S daným problémem souvisí pojem MAUP (Modifiable Areal Unit Problem) – problém modifikovatelné územní jednotky. Zkreslení, které významně ovlivňuje výsledky statistických hypotéz. MAUP může mít formu úspor z rozsahu a zóny účinnosti. Úspory z rozsahu vykazují rozdílné výsledky, jsou-li pro analýzu použita stejná data, avšak dojde k propojení jednotek. To bude velmi dobře patrné v modelech na jiných úrovních členění území.

Tabulka 1: Souhrn výsledků globálního modelu na úrovni obcí

Var	Coef	StdError	t-Statistic	Prob	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	2,8180	1,0851	2,5971	0,0094*	1,4727	1,9135	0,0557	
HUST_ZAL	-0,0020	0,0023	-0,8547	0,3927	0,0014	-1,4405	0,1498	2,5092
NEZAM	-0,0667	0,0609	-1,0967	0,2728	0,0957	-0,6973	0,4857	2,4998
BYTY1K	1,3492	0,0385	35,0060	0,0000*	0,0942	14,3289	0,0000*	1,8388
KES	-0,2979	0,1025	-2,9066	0,0037*	0,1745	-1,7076	0,0878	1,5657
SIL_KM2	0,9631	1,0435	0,9229	0,3561	1,1489	0,8383	0,4019	1,1006

Zdroj: Vlastní zpracování

Regresní koeficient pro každou vysvětlující proměnnou reflektuje sílu a typ vztahu vysvětlující proměnné na závislé proměnné. Pokud koeficient dosahuje negativní hodnoty, tak i daný vztah je negativní. Koeficienty hustoty zalidnění, nezaměstnanosti a koeficientu ekologické stability dosahují záporných hodnot, to znamená, že čím vyšších hodnot dosahují, tím nižší jsou hodnoty migračního salda v globálním modelu na úrovni obcí.

V tomto modelu na úrovni obcí byly proměnné, až na hustotu zalidnění a nezaměstnanost, vyhodnoceny jako statisticky signifikantní. Rozptyl inflačního faktoru VIF měří redundanci mezi vysvětlujícími proměnnými. Je doporučeno odstranit proměnné s hodnotami vyššími než 7,5. Proto byly odstraněny proměnné průměrného věku, indexu ekonomického zatížení a podíl produktivního věku, které tuto hodnotu překračovaly.

Měřítka významnosti modelu jsou F -statistika a Waldova statistika, které jsou zaznamenány v Tabulce 2. V tomto modelu jsou obě statistiky signifikantní. F -statistika je však důvěryhodná jen za předpokladu, kdy Koenkerova statistika není signifikantní.

Tento případ nenastal, proto je třeba přijmout hodnotu Waldovo statistiku pro určení signifikance modelu.

Tabulka 2: Diagnostika globálního modelu na úrovni obcí

Diagnostika	Hodnota	Definice
AIC	58150,4710925000	Akaikeho informační kritérium - relativní míra slouží k porovnávání výkonnosti modelu
R2	0,1669845549	koeficient determinace
AdjR2	0,1663176090	korigovaný koeficient determinace
F-Stat	250,3719593310	F-statistika slouží k vyhodnocení celkové významnosti modelu
F-Prob	0,000000000*	p-hodnota hodnocení celkové významnosti modelu
Wald	229,4926257760	Waldova statistika pro vyhodnocení celkové robustní významnosti modelu
Wald-Prob	0,000000000*	p-hodnota Waldovo statistiky
K(BP)	218,1469653870	Koenkerova studentizovaná Breusch-Paganova Statistika
K(BP)-Prob	0,000000000*	p-hodnota Koenkerovy statistiky
JB	37436,4595223000	Jarqueův-Beraův test normality reziduí
JB-Prob	0,000000000*	p-hodnota Jarqueova-Beraova testu
Sigma2	641,4623610590	odhad reziduálního rozptylu

Zdroj: Vlastní zpracování

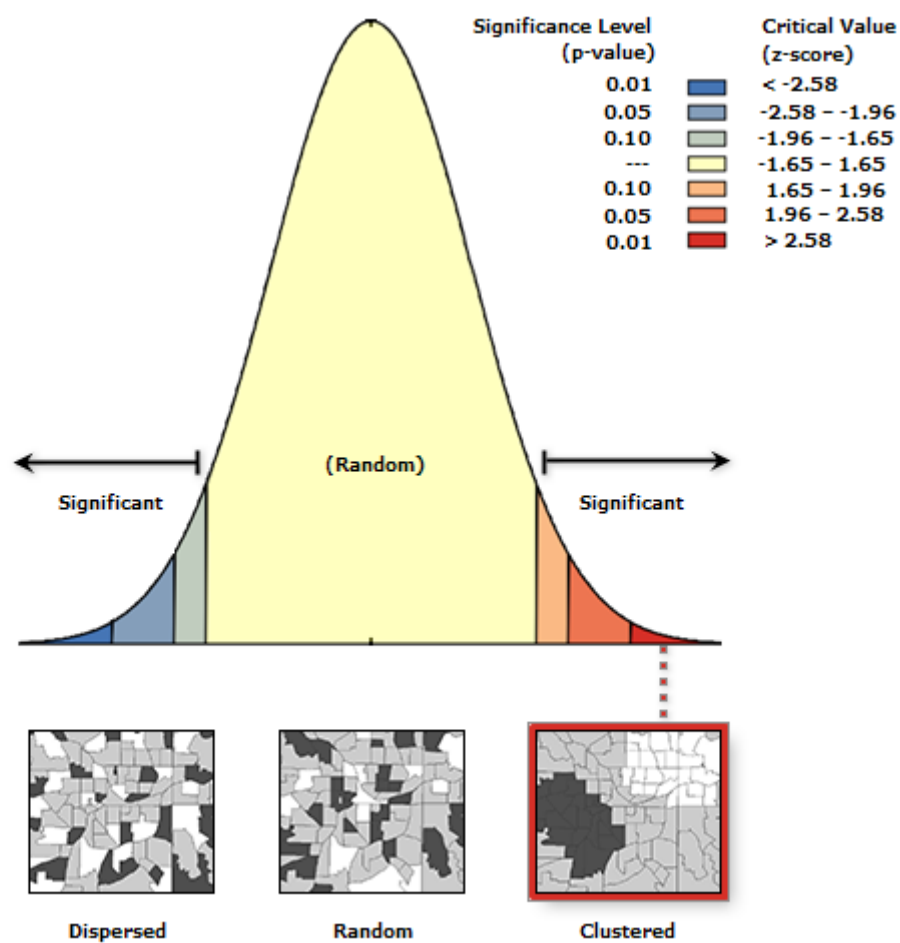
Pro posouzení stacionarity modelu se používá Koenkerova studentizovaná Breuschova-Paganova statistika, která testuje, zda vysvětlující proměnné v modelu mají konzistentní vztah k závisle proměnné v geografickém a datovém prostoru. Pokud je model konzistentní v geografickém prostoru, tak se prostorové procesy vysvětlující proměnné chovají ve všech studovaných oblastech stejně, procesy jsou tedy stacionární. Pokud je model konzistentní v datovém prostoru, variace ve vztazích mezi predikovanými hodnotami a každou vysvětlující proměnnou jsou neměnné se změnami vysvětlujících proměnných, neexistuje heteroskedasticita⁵ v modelu. Nulová hypotéza předpokládá stacionaritu modelu. Na 95% hladině spolehlivosti p-hodnota menší než 0,05 naznačuje statisticky významnou heteroskedasticitu či nestacionaritu. V případě tohoto modelu je

⁵ Heteroskedasticita je porušením předpokladu, že všechny chyby mají stejnou odchylku. Předpoklad lze zapsat ve tvaru $Var(\varepsilon_i) = \sigma^2$ pro všechna $i = 1, 2, \dots, n$. Heteroskedasticita se objevuje, když rozdílné chyby v pozorování mají rozdílné odchylky, např. $Var(\varepsilon_i) = \sigma_i^2$ (Pendakur, 2010).

Koenkerova statistika signifikantní, proto s velkou pravděpodobností půjde model zkvalitnit použitím geograficky vážené regrese.

Pro ověření normálního rozdělení reziduí slouží Jarqueův-Beraův test. Nulová hypotéza tohoto testu vychází z předpokladu, že rezidua mají normální rozdělení. Hodnota vychází menší než 0,05 na 95% hladině významnosti – byl porušen předpoklad normality reziduí. V ideálním případě by bylo vhodné model doplnit o další vhodnou vysvětlující proměnnou.

Obrázek 11: Grafický výstup testování prostorové autokorelace na úrovni obcí



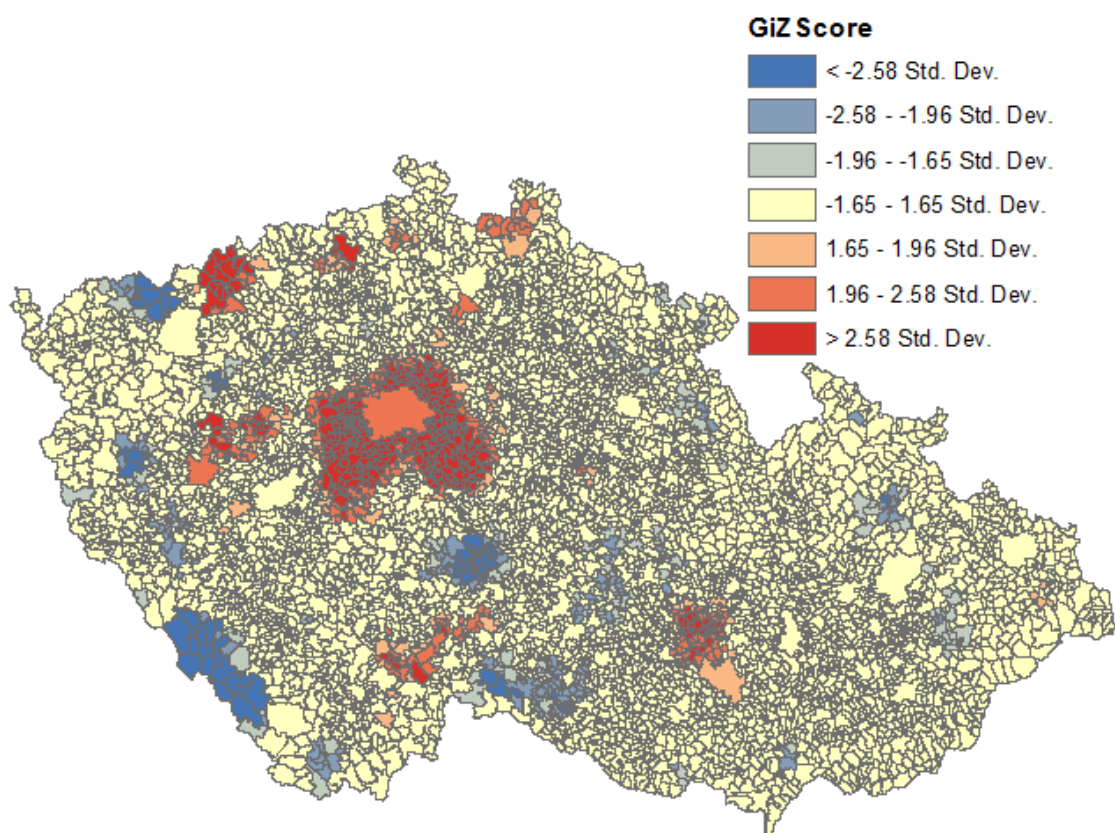
Zdroj: Vlastní zpracování v ArcGIS 10.0

Pro uskutečnění testu prostorové autokorelace standardizovaných reziduí jsem zvolil běžnou metodu Moranova I , které nabylo hodnoty $I = 0,016165$. Výsledná hodnota signalizuje významnou pozitivní prostorovou autokorelací. Standardizovaná rezidua mají

tendenci ke shlukování podobných hodnot v prostoru. Grafický výstup testu prostorové autokorelace je zobrazen na Obrázku 11.

V ideálním případě by model měl mít náhodnou distribuci nadhodnocených a podhodnocených odhadů. Na této úrovni testovaného modelu bylo prokázáno shlukování standardizovaných reziduí a model by mohl být vylepšen přidáním jedné nebo více vysvětlujících proměnných. Pro další prokázání shlukování je možné použít hot spots analýzu.

Obrázek 12: Hot spots analýza standardizovaných reziduí na úrovni obcí



Zdroj: Vlastní zpracování v ArcGIS 10.0

Grafický výsledek hot spots analýzy znázorňuje Obrázek 12. Pro hot spots analýzu standardizovaných reziduí byla stanovena prahová vzdálenost 15 kilometrů, aby každý polygon obce sousedil s alespoň jedním dalším polygonem. Podle grafického znázornění jednoznačně ke shlukování standardizovaných reziduí dochází a potvrzuje to i výsledek Moranova *I*. Hlavní koncentrace shluků vysokých hodnot je viditelná kolem Prahy,

Chomutova a Havlíčkova Brod. Nízké hodnoty v NP Šumava, na Táborsku, Novobystřicku a Karlovarsku.

4.1.2 Obce – lokální model

Statistická signifikantnost Koenkerova testu u globálního modelu obcí je dobrým předpokladem pro zlepšení modelu použitím geograficky vážené regrese. Pro každý model je použito různých typů jádra a metody šířky pásma. Výsledky jednotlivých modelů jsou zachyceny v Tabulce 3. Použitím GWR došlo ke zkvalitnění modelu. Pokud porovnáme lokální modely s globálním, došlo k poklesu Aikakeho informačního kritéria a zároveň vzrostl koeficient determinace.

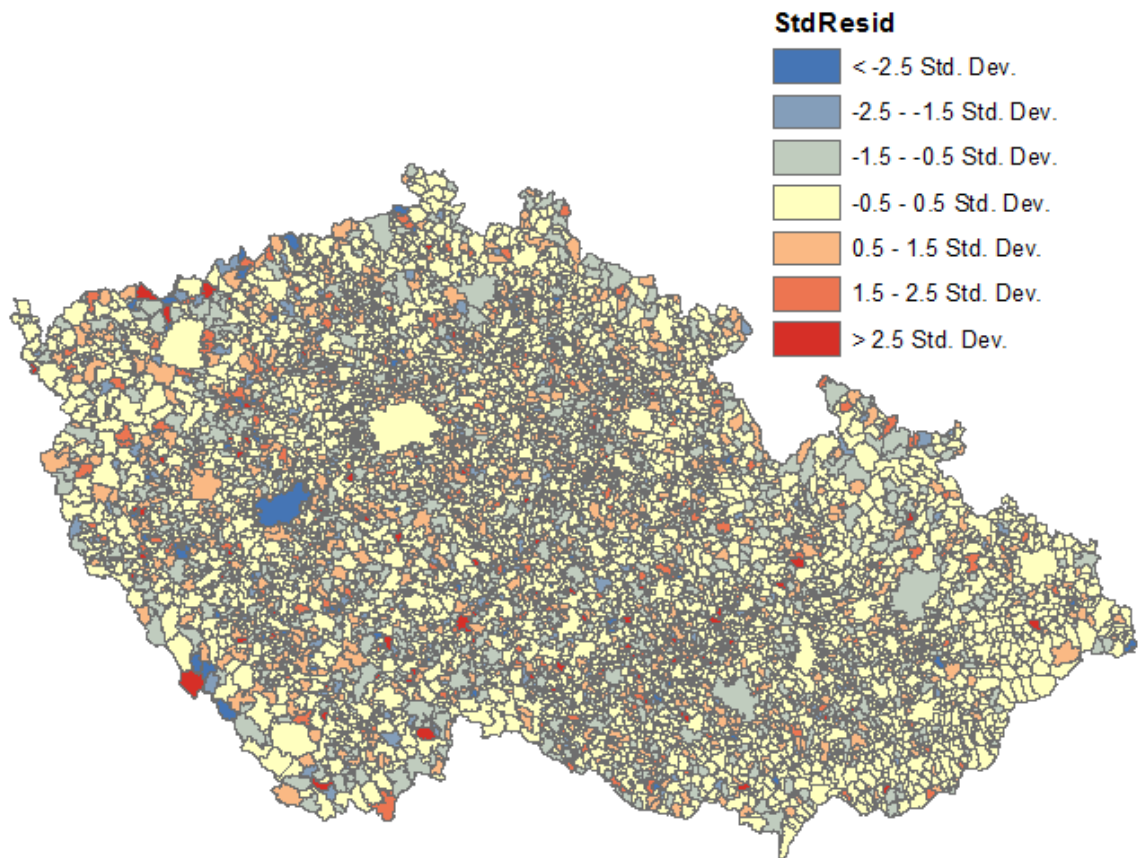
Tabulka 3: Diagnostika lokálního modelu na úrovni obcí

	1	2	3	4
Diagnostika	Fixed - AIC	Adaptive - AIC	Fixed - CV	Adaptive - CV
Bandwidth	19263,899		295617,602	
ResidualSquares	3428208,788	3420049,958	3930830,710	3819017,014
EffectiveNumber	507,048	537,171	40,267	110,329
Sigma	24,430	24,465	25,157	24,988
AICc	57927,672	58012,249	58101,283	58033,426
R2	0,287	0,289	0,183	0,203
R2Adjusted	0,224	0,222	0,177	0,188
Neighbors	30,000	233,000	30,000	1000,000

Zdroj: Vlastní zpracování

Při komparaci modelů lze dojít závěru, že nejlepší z lokálních modelů je model číslo 1 s fixním jádrem, určený na základě porovnání hodnot Aikakeho informačního kritéria a korigovaného koeficientu determinace.

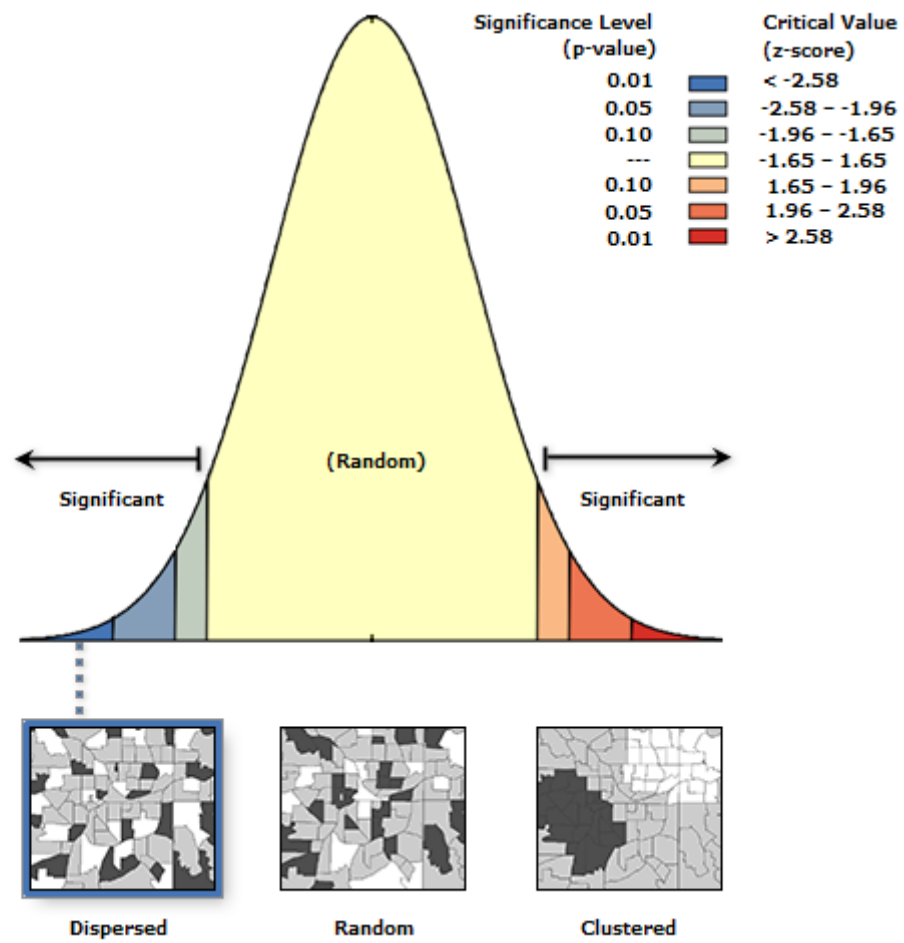
Obrázek 13: Standardizovaná residua lokálního modelu 1s fixním jádrem na úrovni obcí



Zdroj: Vlastní zpracování v ArcGIS 10.0

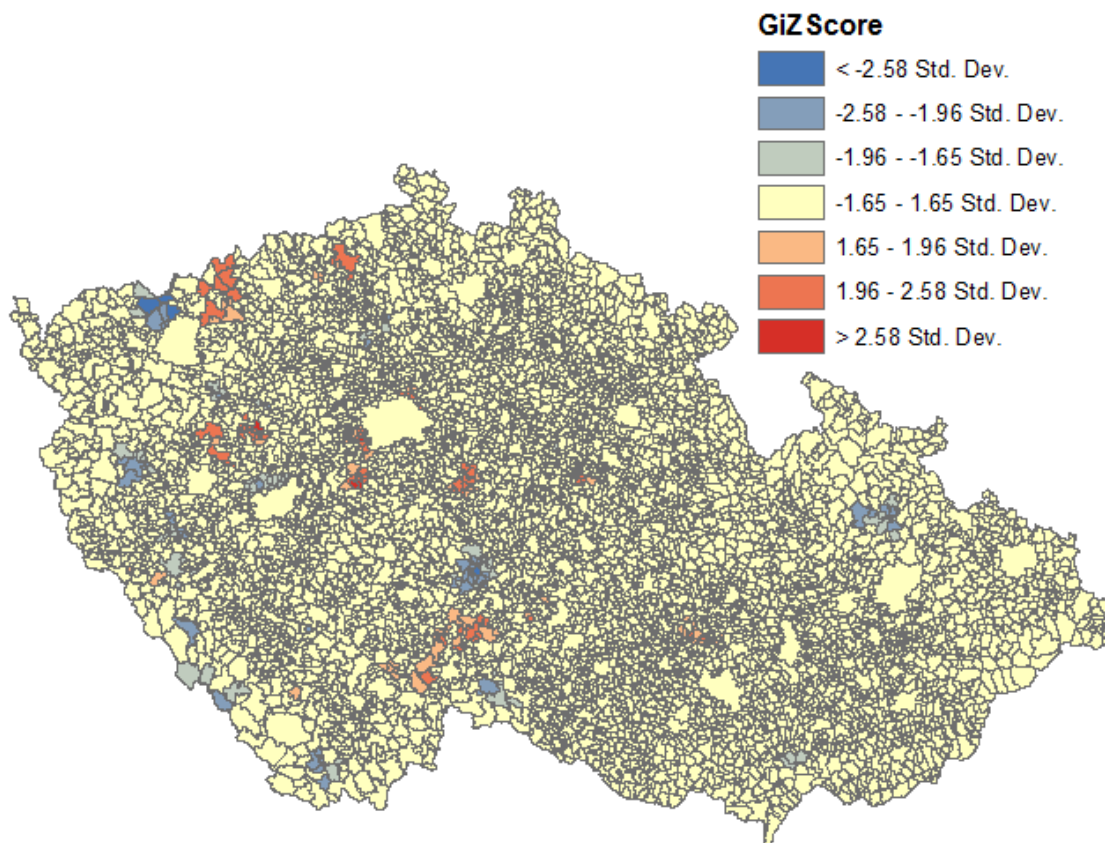
Uskutečněním testu prostorové autokorelace pomocí Moranova I byla zjištěna hodnota $I=-0,011614$. Výsledná negativní hodnota představuje rozptýlené prostorové uspořádání standardizovaných reziduí. Grafický výstup testu prostorové autokorelace je zobrazen na Obrázku 14. Prokázané rozptýlení signalizuje, že by model měl být vylepšen přidáním další proměnné.

Obrázek 14: Grafický výstup testování prostorové autokorelace lokálního modelu 1 s fixním jádrem na úrovni obcí



Zdroj: Vlastní zpracování v ArcGIS 10.0

Obrázek 15: Hot spots analýza standardizovaných reziduí modelu 1s fixním jádrem na úrovni obcí

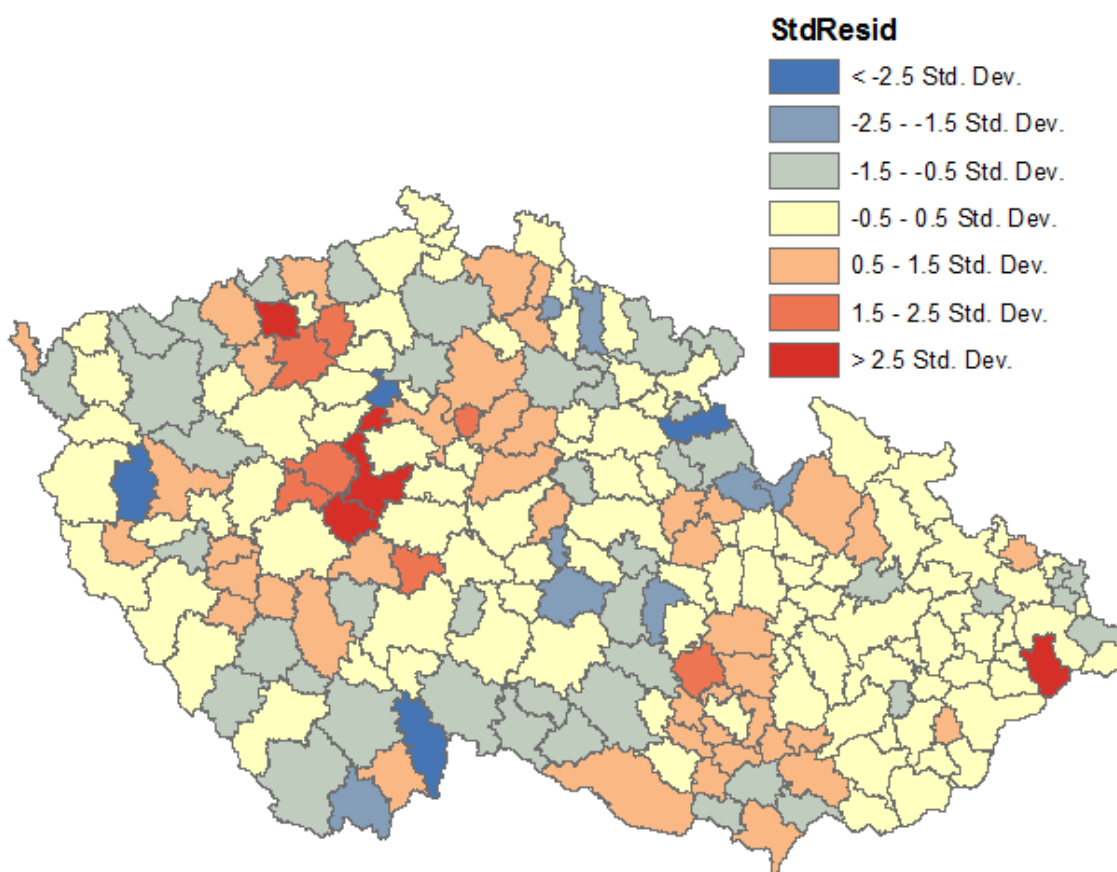


Zdroj: Vlastní zpracování v ArcGIS 10.0

4.1.3 Obce s rozšířenou působností – globální model

Pro model na úrovni SO ORP je koeficient determinace 0,553 a korigovaný koeficient determinace 0,542 (viz Tabulka 5). Oproti modelu s obcemi došlo k významnému nárůstu výkonnosti modelu, kdy již nedochází k tak výraznému zkreslení vlivem změn u malých jednotek.

Obrázek 16: Standardizovaná residua globálního modelu na úrovni ORP



Zdroj: Vlastní zpracování v ArcGIS 10.0

Až na podíl bytů na 1000 obyvatel vyšly koeficienty proměnných záporně, podle modelu tedy jejich růst znamená pokles migračního salda. Jako statisticky signifikantní vysvětlující proměnné byly vyhodnoceny hustota zalidnění a byty na 1000 obyvatel.

Tabulka 4: Souhrn výsledků globálního modelu na úrovni ORP

Var	Coef	StdError	t-Statistic	Prob	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	-4,6442	1,9655	-2,3629	0,0191*	1,8497	-2,5108	0,0128*	
HUST_ZAL	-0,0041	0,0016	-2,6614	0,0084*	0,0010	-4,1231	0,0001*	1,0244
NEZAM	-0,1223	0,1015	-1,2049	0,2297	0,0903	-1,3544	0,1771	1,1589
BYTY1K	2,1433	0,1630	13,1517	0,0000*	0,2063	10,3888	0,0000*	1,1527
KES	-0,2033	0,1927	-1,0548	0,2928	0,1992	-1,0204	0,3088	1,0595
SIL_KM2	3,7983	2,0989	1,8097	0,0718	2,1601	1,7584	0,0802	1,4376

Zdroj: Vlastní zpracování

Jak je patrné z Tabulky 5, hodnoty F-statistiky a Waldovy statistiky jsou signifikantní. Koenkerova statistika je také signifikantní, proto je pro určení signifikantnosti modelu třeba přijmout hodnotu Waldovy statistiky. Opět byl porušen předpoklad normality reziduí, model není vhodně sestaven.

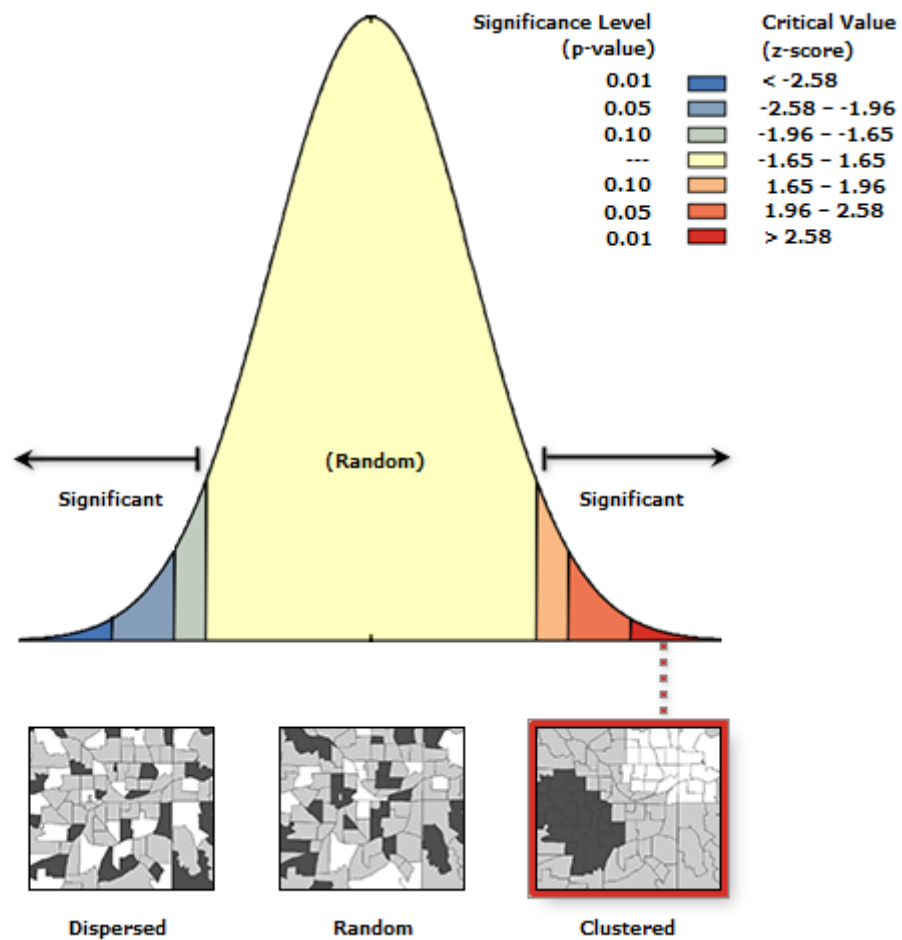
Tabulka 5: Diagnostika globálního modelu na úrovni ORP

Diagnostika	Hodnota	Definice
AIC	1121,01917723000	Akaikeho informační kritérium - relativní míra slouží k porovnávání výkonnosti modelu
R2	0,55375157382	koeficient determinace
AdjR2	0,54259536317	korigovaný koeficient determinace
F-Stat	49,63617046850	F-statistika slouží k vyhodnocení celkové významnosti modelu
F-Prob	0,000000000*	p-hodnota hodnocení celkové významnosti modelu
Wald	166,25487154600	Waldova statistika pro vyhodnocení celkové robustní významnosti modelu
Wald-Prob	0,000000000*	p-hodnota Waldovo statistiky
K(BP)	22,29663284830	Koenkerova studentizovaná Breusch-Paganova Statistika
K(BP)-Prob	0,00045979167*	p-hodnota Koenkerovy statistiky
JB	21,64278080560	Jarqueův-Beraův test normality reziduí
JB-Prob	0,00001996778*	p-hodnota Jarqueova-Beraova testu
Sigma2	13,13486276170	odhad reziduálního rozptylu

Zdroj: Vlastní zpracování

Moranovo I nabylo hodnoty 0,101658. Výsledná hodnota signalizuje významnou pozitivní prostorovou autokorelaci. Standardizovaná rezidua mají tendenci ke shlukování podobných hodnot v prostoru. Výstup testu prostorové autokorelace zachycuje Obrázek 17.

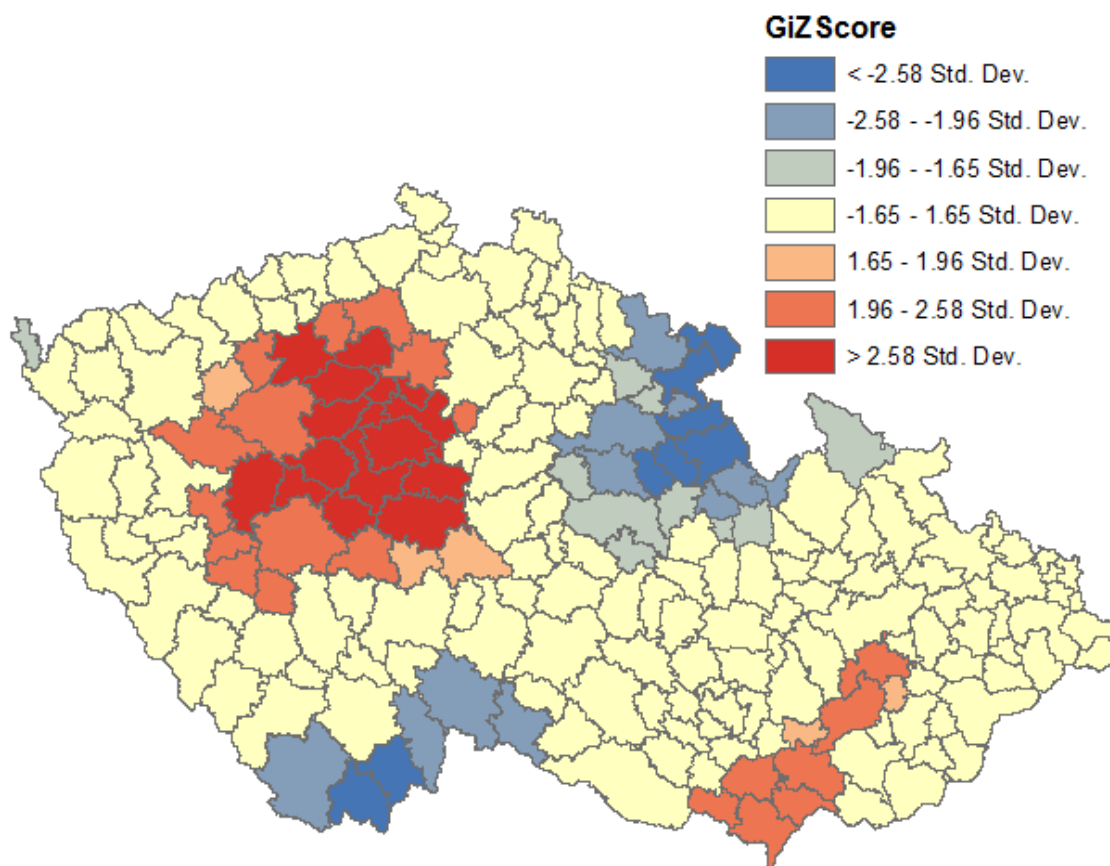
Obrázek 17: Grafický výstup testování prostorové autokorelace na úrovni ORP – globální model



Zdroj: Vlastní zpracování v ArcGIS 10.0

Pro hot spots analýzu byla zvolena prahová vzdálenost 80 kilometrů, která měla zajistit, aby každá jednotka měla alespoň jednoho souseda. Grafický výstup zachycuje Obrázek 18. Shluky nízkých hodnot jsou přítomny na území Královéhradecka, Pardubicka, jižní poloviny jižních Čech. Shluky vysokých hodnot pak na území Středočeského kraje a východní části Jihomoravského kraje.

Obrázek 18: Hot spots analýza standardizovaných reziduí globálního modelu na úrovni ORP



Zdroj: Vlastní zpracování v ArcGIS 10.0

4.1.4 Obce s rozšířenou působností – lokální model

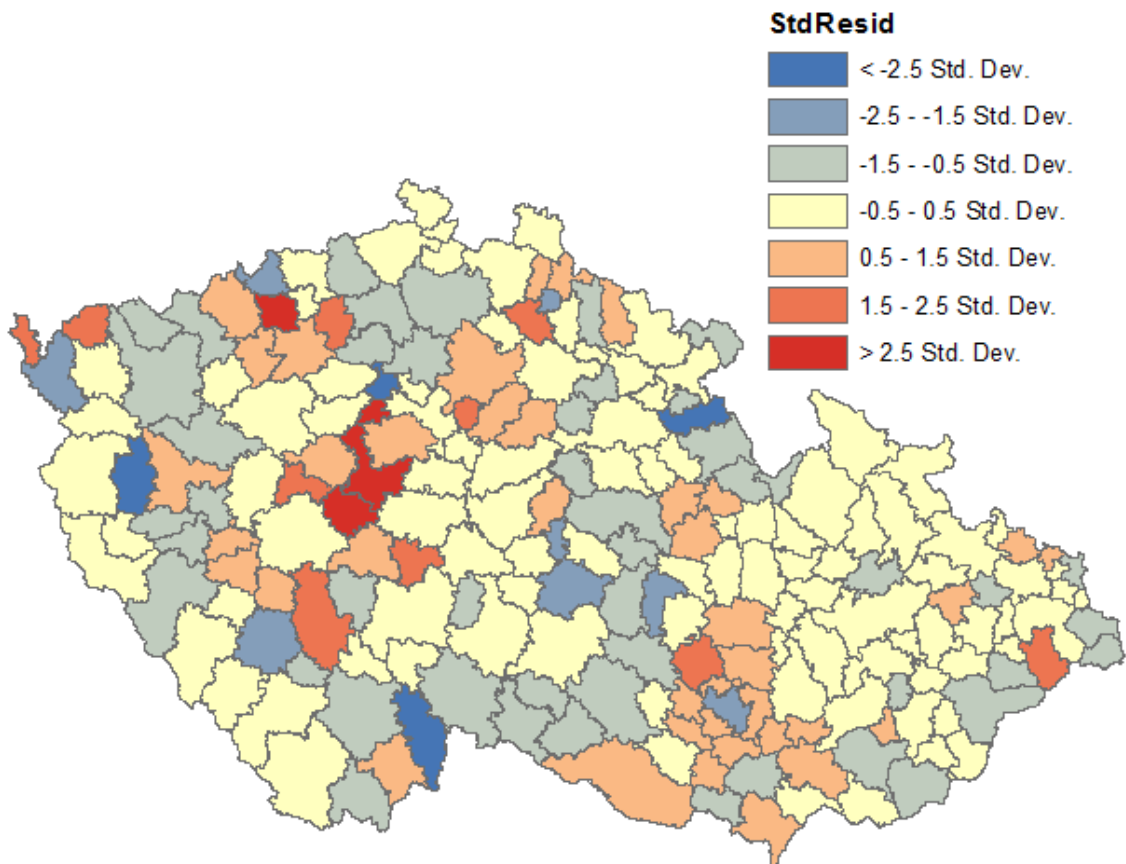
Použitím GWR na úrovni obcí s rozšířenou působností došlo ke zkvalitnění modelu. Aikakeho informačního kritérium a korigovaný koeficient determinace ukazuje na nejlepší výsledky u modelu 1 s fixním jádrem. Korigovaný koeficient determinace tedy vzrostl z 0,543 na 0,602, což jednoznačně mluví ve prospěch použití GWR.

Tabulka 6: Diagnostika lokálního modelu na úrovni ORP

	1	2	3	4
Diagnostika	Fixed - AIC	Adaptive - AIC	Fixed - CV	Adaptive - CV
Bandwidth	88156,814		228656,072	
ResidualSquares	1972,303	2176,166	2481,851	2314,262
EffectiveNumber	33,336	22,153	11,925	17,430
Sigma	3,380	3,440	3,576	3,503
AICc	1110,981	1111,600	1119,753	1115,095
R2	0,665	0,630	0,578	0,607
R2Adjusted	0,602	0,588	0,554	0,573
Neighbors	30,000	143,000	30,000	172,000

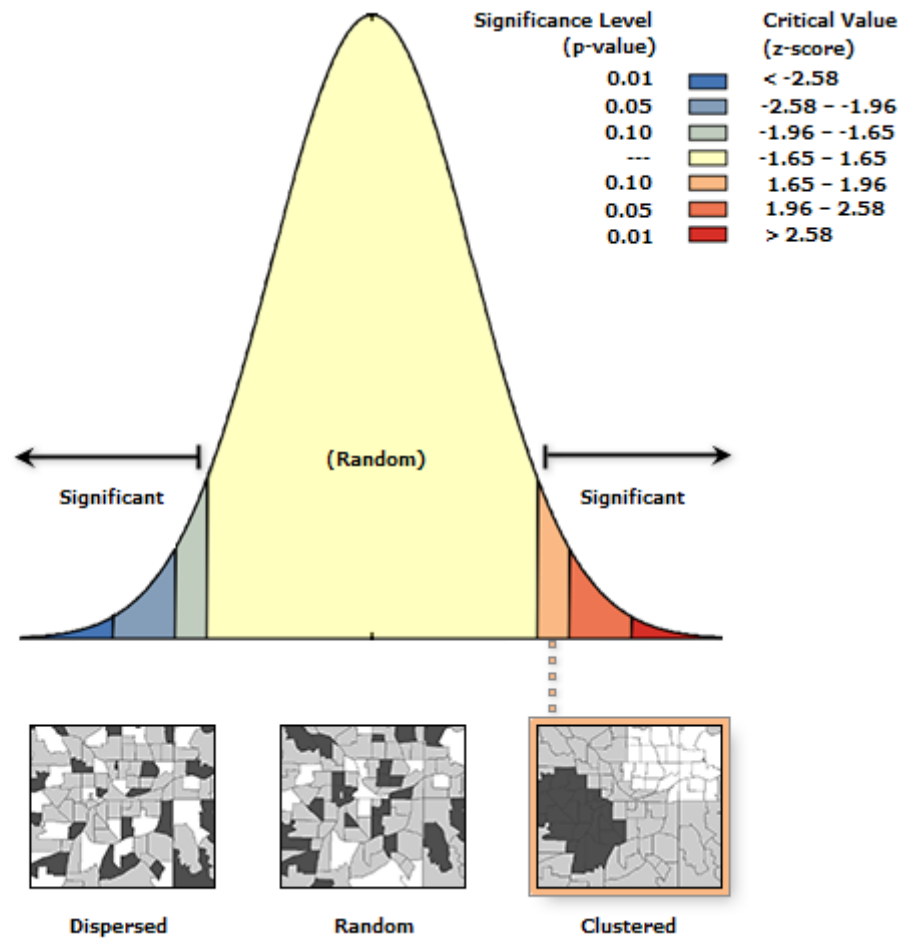
Zdroj: Vlastní zpracování

Obrázek 19: Standardizovaná residua lokálního modelu 1s fixním jádrem na úrovni ORP



Zdroj: Vlastní zpracování v ArcGIS 10.0

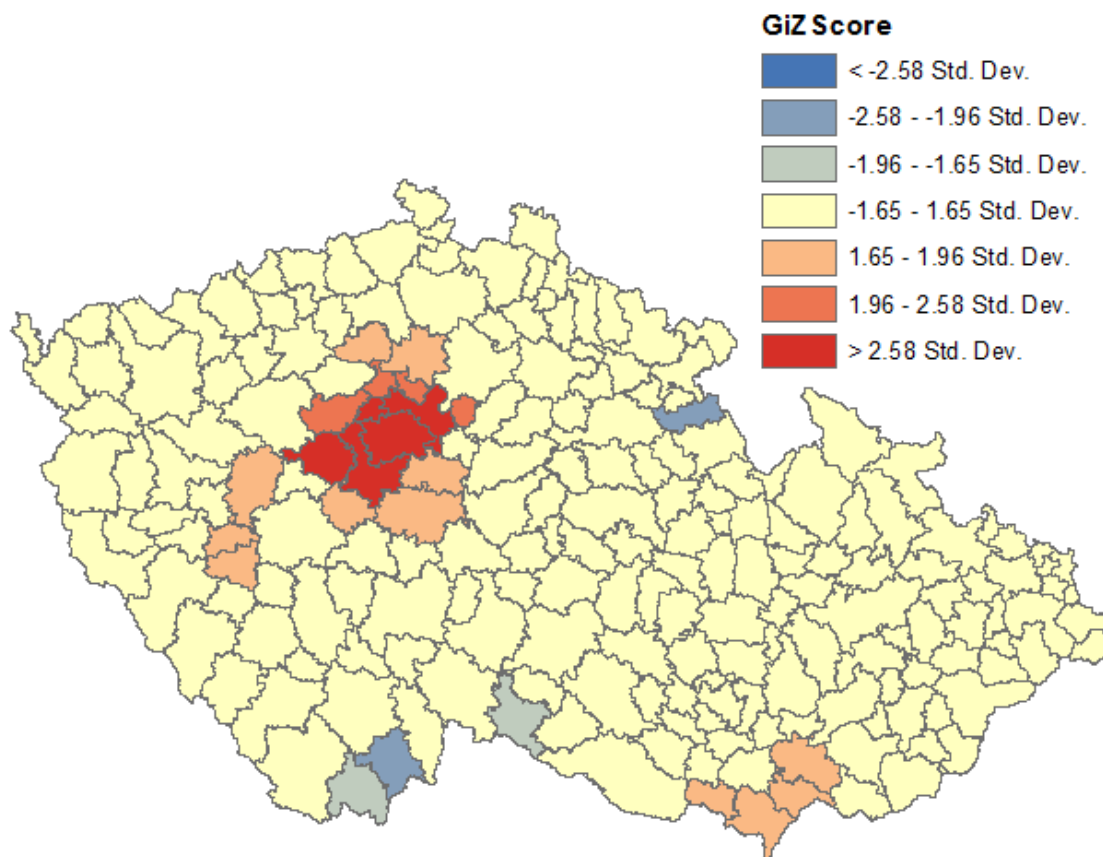
Obrázek 20: Grafický výstup testování prostorové autokorelace lokálního modelu 1 s fixním jádrem na úrovni ORP



Zdroj: Vlastní zpracování v ArcGIS 10.0

Kladná hodnota Moranova I ukazuje na významnou pozitivní prostorovou autokorelaci. Standardizovaná rezidua mají tendenci se shlukovat v podobných hodnotách v prostoru. Hot spots analýza s výstupem na Obrázku 21 potvrzuje závěr z testu prostorové autokorelace. Shluky vysokých hodnot jsou viditelné na území Středočeského kraje a Břeclavsku. Shluky nízkých hodnot jsou k vidění na Českokrumlovsku.

Obrázek 21: Hot spots analýza standardizovaných reziduí modelu 1s fixním jádrem na úrovni ORP

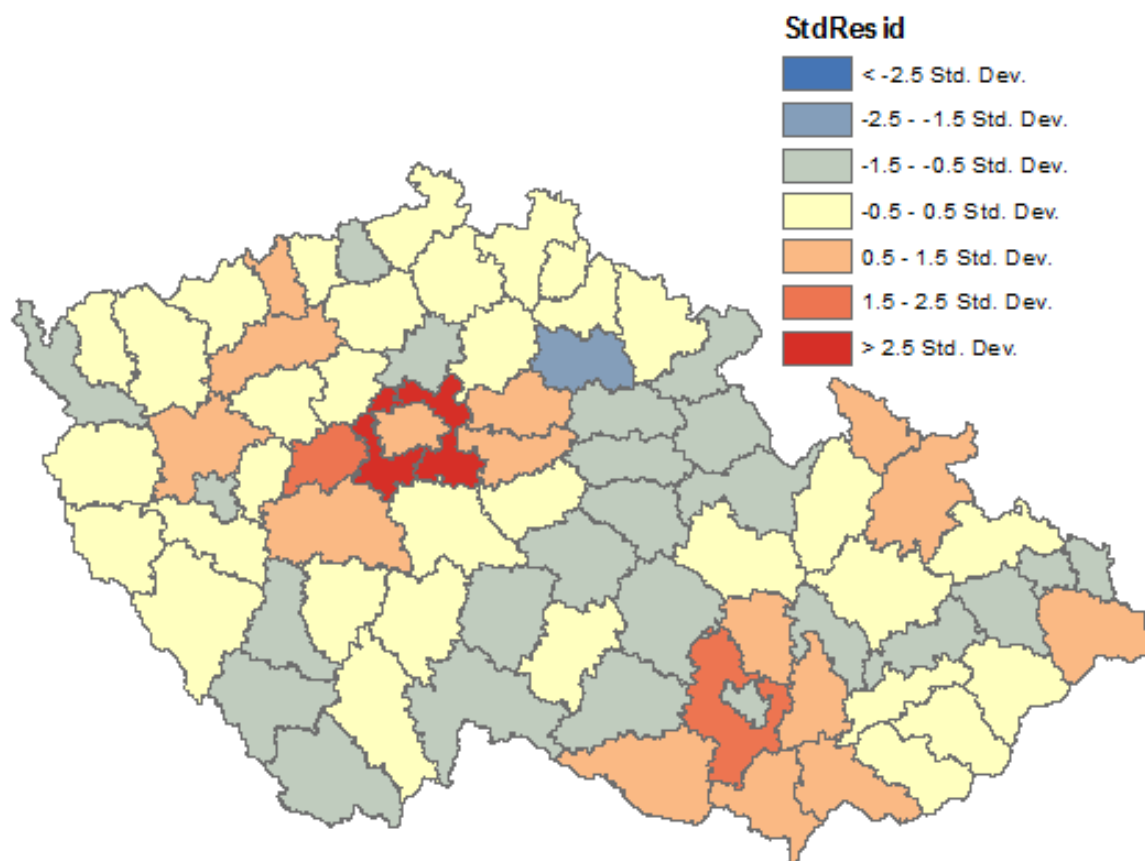


Zdroj: Vlastní zpracování v ArcGIS 10.0

4.1.5 Okresy – globální model

U stanoveného regresního modelu pro okresy vyšel koeficient determinace 0,319 a korigovaný koeficient determinace 0,271 jak ukazuje Tabulka 9. V porovnání s modelem ORP došlo k poklesu výkonnosti modelu, avšak model je výkonnější než v případě obcí. Hodnoty koeficientů vysvětlujících proměnných z Tabulky 7 jsou, až na hustotu silniční sítě, negativní – růst jejich hodnot by v modelu vedl k poklesu vysvětlované proměnné. Jako statisticky významné vysvětlující proměnné byly vyhodnoceny ukazatele hustoty zalidnění, nezaměstnanosti a hustoty silniční sítě.

Obrázek 22: Standardizovaná residua globálního modelu na úrovni okresů



Zdroj: Vlastní zpracování v ArcGIS 10.0

Tabulka 7: Souhrn výsledků globálního modelu na úrovni okresů

Var	Coef	StdError	t-Statistic	Prob	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	2,678	4,116	0,651	0,517	2,814	0,952	0,344	
HUST	-0,004	0,002	-2,516	0,014	0,002*	-2,738	0,008*	1,045
NEZAM	-0,699	0,213	-3,277	0,002	0,240*	-2,911	0,005*	1,135
BYTY1K	-0,001	0,001	-0,878	0,383	0,000	-4,703	0,000*	1,009
KES	-0,290	0,387	-0,747	0,457	0,248	-1,167	0,247	1,140
SIL_KM2	11,883	4,758	2,497	0,015	4,722*	2,516	0,014*	1,920

Zdroj: Vlastní zpracování

Tabulka 8: Hodnoty korelačních koeficientů mezi ukazateli na úrovni okresů

	NEZAM	HUST_ZAL	BYTY1K	KES	SIL_KM2
NEZAM		-0,148	0,0825	0,325	-0,341
HUST_ZAL	-0,148		-0,0557	-0,178	0,57
BYTY1K	0,0825	-0,0557		0,0255	-0,0722
KES	0,325	-0,178	0,0255		-0,456
SIL_KM2	-0,341	0,57	-0,0722	-0,456	

Zdroj: Vlastní zpracování

Tabulka 9: Diagnostika globálního modelu na úrovni okresů

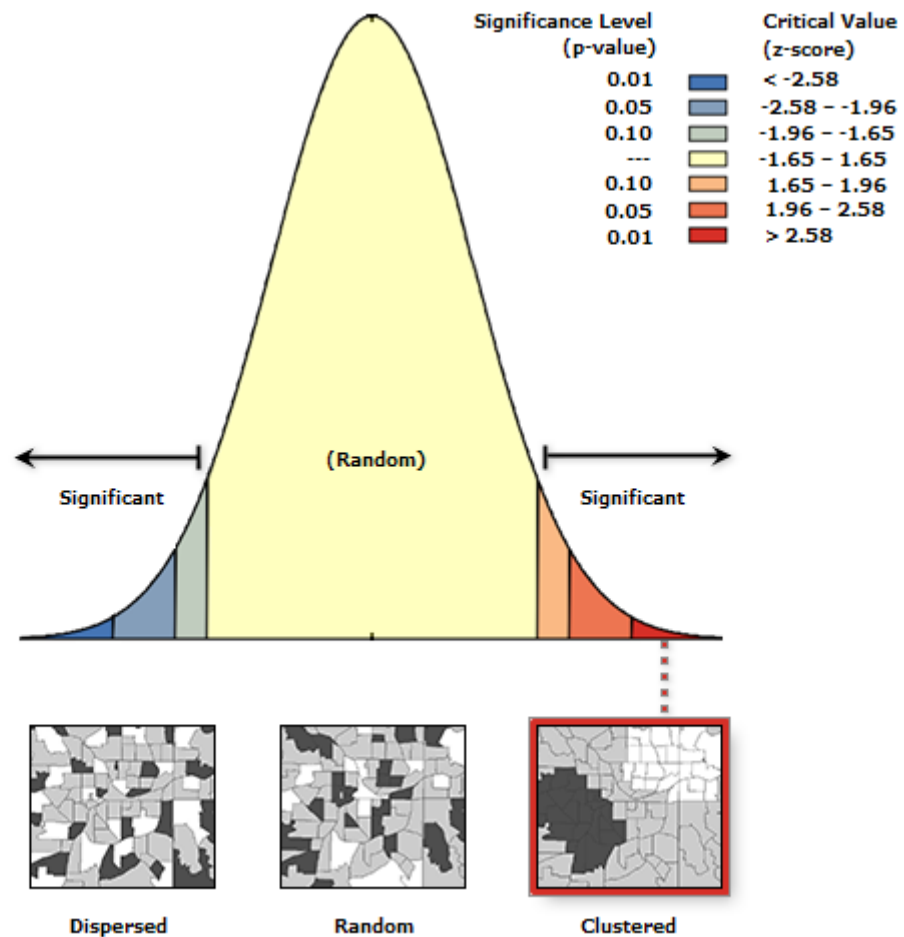
Diagnostika	Hodnota	Definice
AIC	451,76999252400	Akaikeho informační kritérium - relativní míra slouží k porovnávání výkonnosti modelu
R2	0,31887204386	koeficient determinace
AdjR2	0,27090528638	korigovaný koeficient determinace
F-Stat	6,64777151188	F-statistika slouží k vyhodnocení celkové významnosti modelu
F-Prob	0,00003951678*	p-hodnota hodnocení celkové významnosti modelu
Wald	44,62046884010	Waldova statistika pro vyhodnocení celkové robustní významnosti modelu
Wald-Prob	0,00000001733*	p-hodnota Waldovy statistiky
K(BP)	9,67275423852	Koenkerova studentizovaná Breusch-Paganova Statistika
K(BP)-Prob	0,08505673585	p-hodnota Koenkerovy statistiky
JB	159,50763241000	Jarqueův-Beraův test normality reziduí
JB-Prob	0,00000000000*	p-hodnota Jarqueova-Beraova testu
Sigma2	19,19302789480	odhad reziduálního rozptylu

Zdroj: Vlastní zpracování

F-statistika a Waldova statistika byla vyhodnocena jako signifikantní. Koenkerova statistika není signifikantní, proto je pro určení signifikantnosti modelu možné přijmout hodnotu *F*-statistiky. Stejně jako v předchozích dvou modelech byl porušen předpoklad

normality reziduí, což znamená, že model není vhodně sestaven a bylo by vhodné přidat další proměnnou.

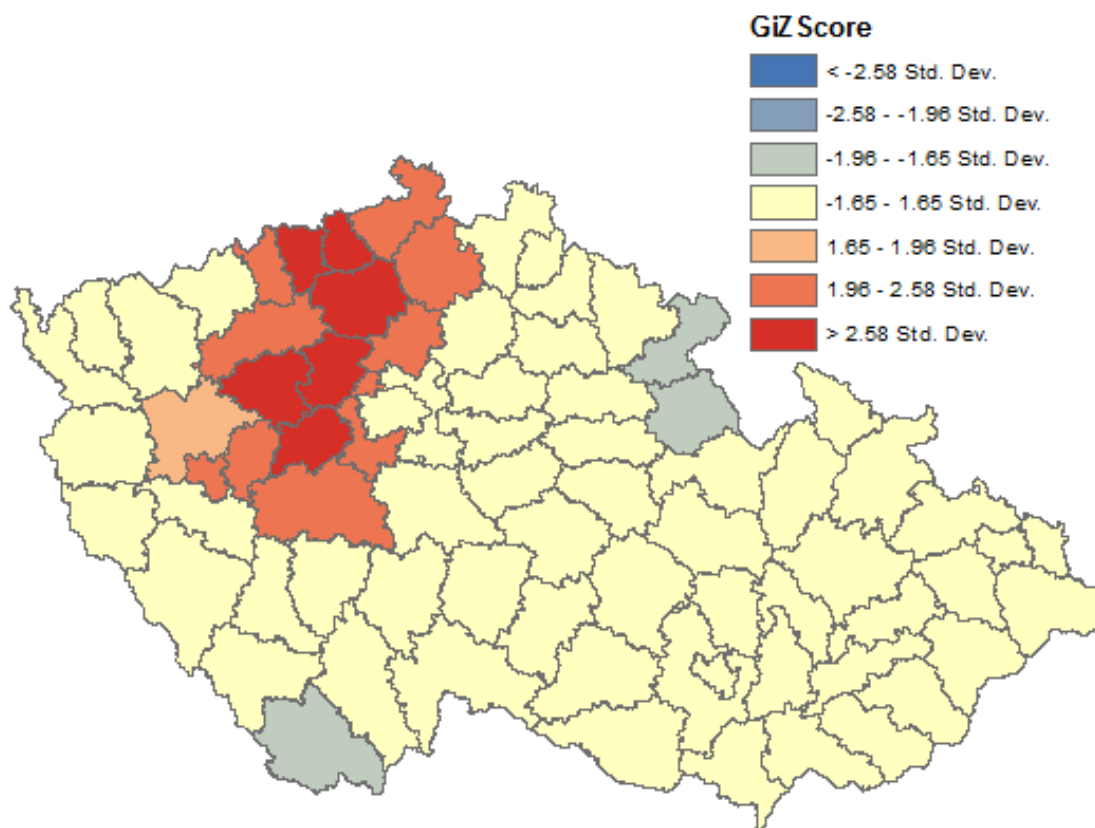
Obrázek 23: Grafický výstup testování prostorové autokorelace na úrovni okresů



Zdroj: Vlastní zpracování v ArcGIS 10.0

Hodnota Moranova I nabyla hodnoty 0,136548, což signalizuje pozitivní prostorovou autokorelaci. Z grafického výstupu testu na Obrázku 24 je patrné, že standardizovaná rezidua mají tendenci ke shlukování v prostoru. Stejný závěr byl i u předchozích dvou modelů.

Obrázek 24: Hot spots analýza standardizovaných reziduí na úrovni okresů



Zdroj: Vlastní zpracování v ArcGIS 10.0

Na Obrázku 24 je zřejmý předchozí závěr – standardizovaná rezidua mají tendenci se v prostoru shlukovat. Jako prahová vzdálenost byla zvolena hodnota 100 kilometrů. Shluky nízkých hodnot představují okresy Český Krumlov, Náchod a Trutnov. Vysoké shluky představuje území na severovýchod od Prahy.

4.1.6 Okresy – lokální model

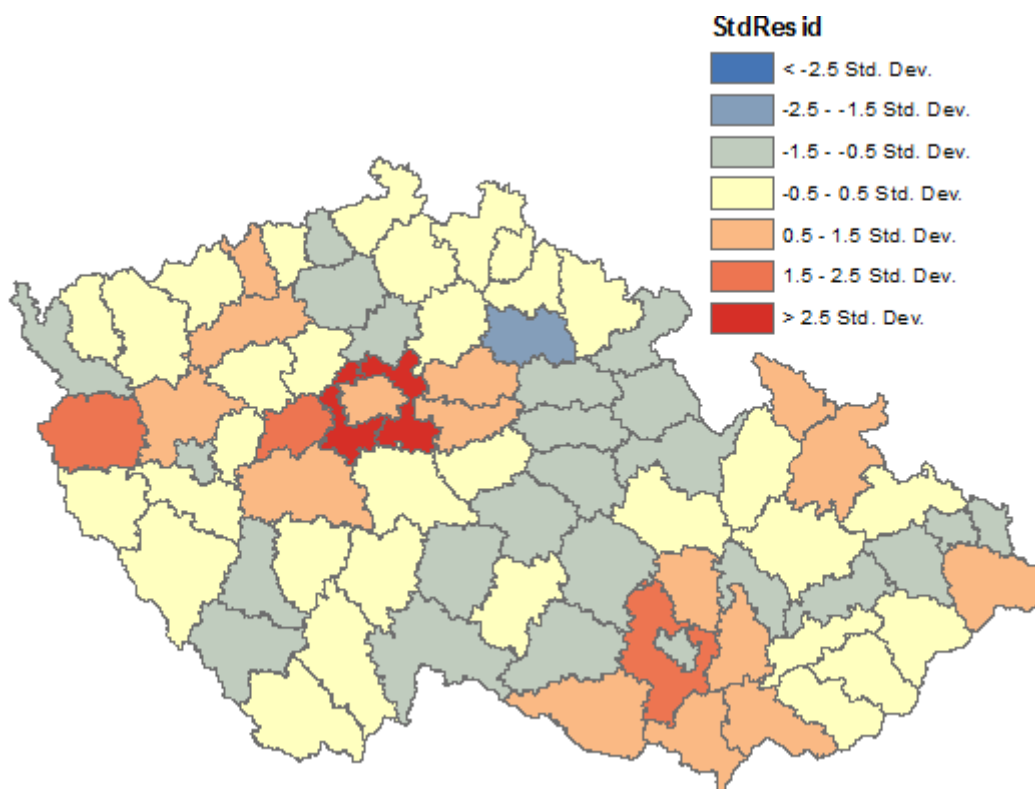
Použití GWR na úrovni okresů přineslo jen nepatrné zlepšení přesnosti modelu. Bylo možné sestavit pouze první model s fixním jádrem z důvodu vykazování chyby multikolinearity. Korelační koeficient u ukazatelů hustoty zalidnění a hustoty silniční sítě na kilometr čtvereční dosahuje vysoké hodnoty 0,57, což značí přímou lineární závislost (viz Tabulka 8). Hodnoty VIF, jak je zachycuje Tabulka 7, jsou však nízké. Proto v rámci srovnání globálního a lokálního modelu počet prediktorů ponechám a jako jediný oponent poslouží model 1.

Tabulka 10: Diagnostika lokálního modelu na úrovni okresů

	1	2	3	4
Diagnostika	Fixed - AIC	Adaptive - AIC	Fixed - CV	Adaptive - CV
Bandwidth	350947,441		x	
ResidualSquares	1305,870	x	x	x
EffectiveNumber	8,094	x	x	x
Sigma	4,353	x	x	x
AICc	455,002	x	x	x
R2	0,347	x	x	x
R2Adjusted	0,280	x	x	x
Neighbors	30	x	30	x

Zdroj: Vlastní zpracování

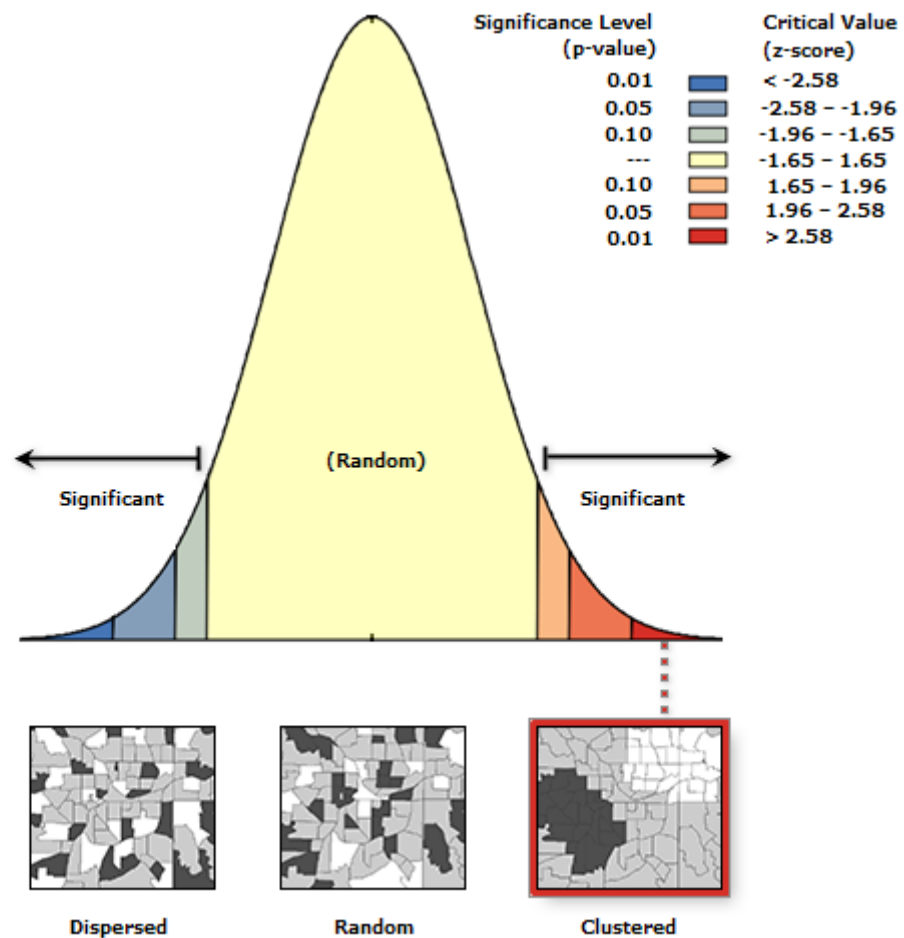
Obrázek 25: Standardizovaná residua lokálního modelu 1 s fixním jádrem na úrovni okresů



Zdroj: Vlastní zpracování v ArcGIS 10.0

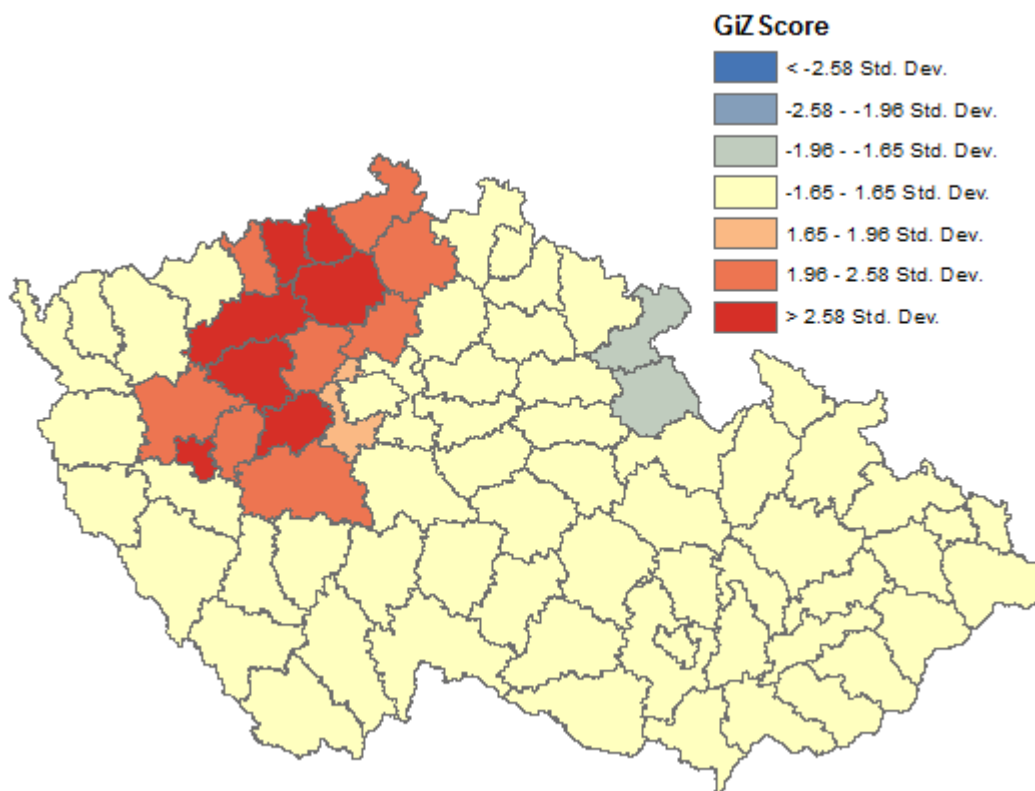
Uskutečněním testu prostorové autokorelace pomocí Moranova I byla zjištěna hodnota $I=0,122775$. Pozitivní hodnota znamená shlukované prostorové uspořádání standardizovaných reziduí. Grafický výstup testu prostorové autokorelace je zobrazen na Obrázku 14.

Obrázek 26: Grafický výstup testování prostorové autokorelace lokálního modelu 1 s fixním jádrem na úrovni okresů



Zdroj: Vlastní zpracování v ArcGIS 10.0

Obrázek 27: Hot spots analýza standardizovaných reziduí modelu 1s fixním jádrem na úrovni okresů

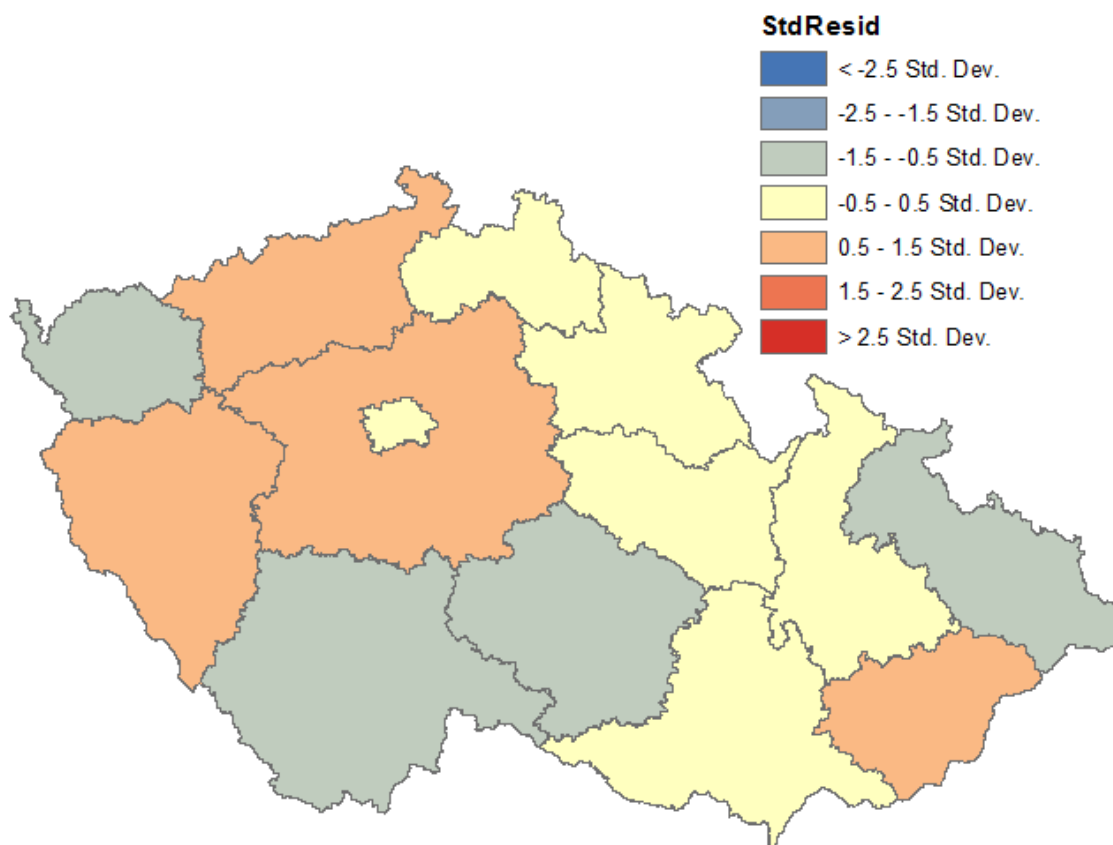


Zdroj: Vlastní zpracování v ArcGIS 10.0

4.1.7 Kraje – globální model

Model krajů vyšel z pohledu výkonnosti ze sestavených modelů nejlépe. Koeficient determinace vyšel 0,867 a korigovaný koeficient determinace 0,807 (viz Tabulka 11). Kladné koeficienty vyšly u proměnných hustota zalidnění, nezaměstnanost a byty na 1000 obyvatel. Růst těchto proměnných v modelu znamená pokles migračního salda. Záporná hodnota vyšla pouze u ukazatele ekologické stability. Signifikantní jsou ale jenom byty, takže by stála za úvahu diskuse použitých proměnných na této úrovni. Zřejmě by na úrovni krajů migraci lépe vysvětlovaly nějaké další proměnné.

Obrázek 28: Standardizovaná residua globálního modelu na úrovni krajů



Zdroj: Vlastní zpracování v ArcGIS 10.0

Tabulka 11: Souhrn výsledků globálního modelu na úrovni krajů

Var	Coef	StdError	t-Statistic	Prob	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	-11,1281	5,2246	-2,1299	0,0657	4,3894	-2,5352	0,0349*	
NEZAM	0,0953	0,2590	0,3679	0,7225	0,1976	0,4820	0,6426	2,5092
HUST_ZAL	0,0003	0,0016	0,1589	0,8777	0,0012	0,2240	0,8284	7,9528
BYTY1K	0,4523	0,0947	4,7790	0,0014*	0,0609	7,4325	0,0000*	2,0117
KES	-0,1527	0,8929	-0,1711	0,8685	0,5363	-0,2848	0,7831	1,6411
SIL_KM2	6,5226	6,1997	1,0521	0,3235	5,0122	1,3014	0,2294	7,4329

Zdroj: Vlastní zpracování

Vysvětlující proměnné hustoty zalidnění a hustoty silnic na čtvereční kilometr vykazují vysoké hodnoty VIF (variance inflation factor), jak ukazuje Tabulka 11. Vysoké hodnoty ukazují na multikolinearitu. Hodnoty nad 7,5 by měly být vyřazeny z modelu. Tabulka

12 ukazuje, že mezi ukazateli hustoty zalidnění a hustoty silnic na čtvereční kilometr existuje signifikantní přímá lineární závislost. To znamená, že se změnou míry zalidnění ve většině případů stejným způsobem dochází ke změně hustoty silniční sítě.

Tabulka 12: Hodnoty korelačních koeficientů mezi ukazateli na úrovni krajů

	NEZAM	HUST_ZAL	BYTY1K	KES	SIL_KM2
NEZAM		-0,605	-0,467	0,314	-0,625
HUST_ZAL	-0,605		0,00859	-0,469	0,908
BYTY1K	-0,467	0,00859		-0,331	0,195
KES	0,314	-0,469	-0,331		-0,554
SIL_KM2	-0,625	0,908	0,195	-0,554	

Zdroj: Vlastní zpracování

Jako signifikantní byla vyhodnocena F -statistika i Waldova statistika. Koenkerova statistika nebyla vyhodnocena jako statisticky signifikantní, proto lze přijmout hodnotu F -statistiky.

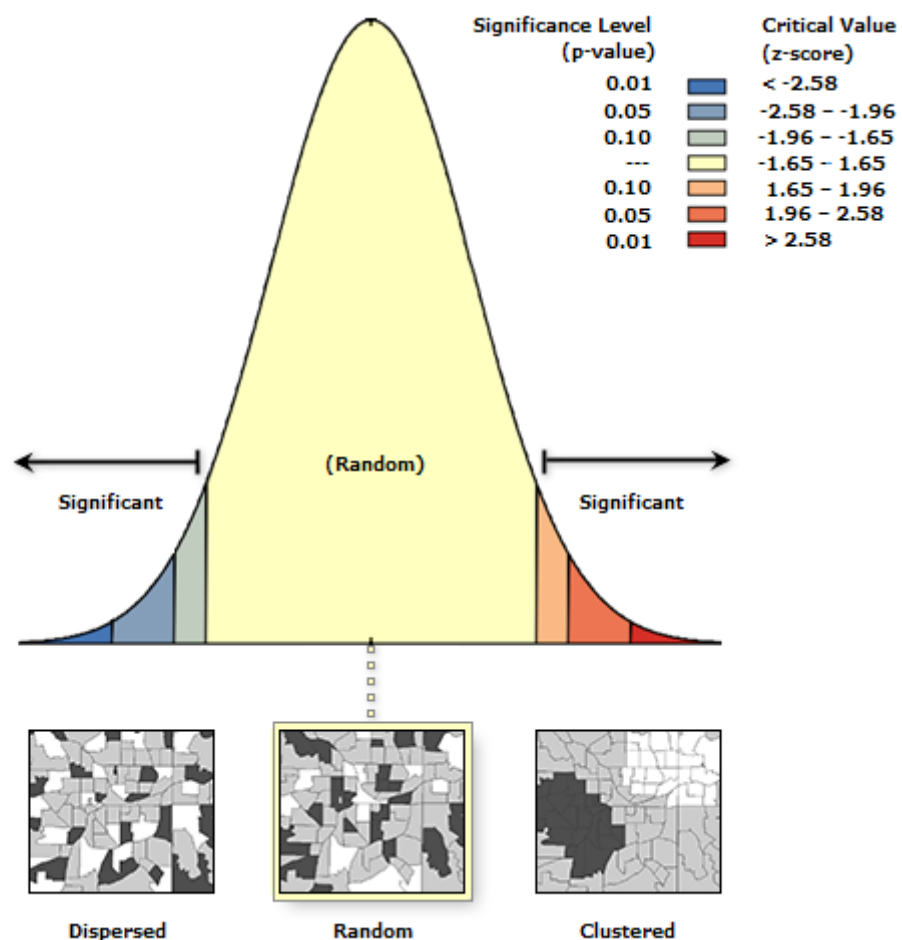
Tabulka 13: Diagnostika globálního modelu na úrovni krajů

Diagnostika	Hodnota	Definice
AIC	51,76963536900	Akaikeho informační kritérium - relativní míra slouží k porovnávání výkonnosti modelu
R2	0,88291231452	koeficient determinace
AdjR2	0,80973251110	korigovaný koeficient determinace
F-Stat	12,06497248180	F-statistika slouží k vyhodnocení celkové významnosti modelu
F-Prob	0,00146358700*	p-hodnota hodnocení celkové významnosti modelu
Wald	271,69301493300	Waldova statistika pro vyhodnocení celkové robustní významnosti modelu
Wald-Prob	0,000000000*	p-hodnota Waldovo statistiky
K(BP)	3,96905134012	Koenkerova studentizovaná Breusch-Paganova Statistika
K(BP)-Prob	0,55388037577	p-hodnota Koenkerovy statistiky
JB	0,89993928623	Jarqueův-Beraův test normality reziduí
JB-Prob	0,63764750832	p-hodnota Jarqueova-Beraova testu
Sigma2	1,75492647626	odhad reziduálního rozptylu

Zdroj: Vlastní zpracování

Při testování prostorové autokorelace vyšla hodnota Moranova $I = -0,108691$, negativní prostorová autokorelace. Standardizovaná rezidua jsou náhodně rozptýlená. To pouze v tomto modelu na úrovni krajů, modely na jiných úrovních měly tendenci ke shlukování. P -hodnota Jarqueova-Beraova testu nebyla vyhodnocena jako signifikantní, nebyl porušen předpoklad normality reziduí. Ani tento model není vhodně sestaven, pro jeho sestavení bylo použito málo vstupů a pouze jediná proměnná byla vyhodnocena jako statisticky signifikantní (viz Tabulka 11).

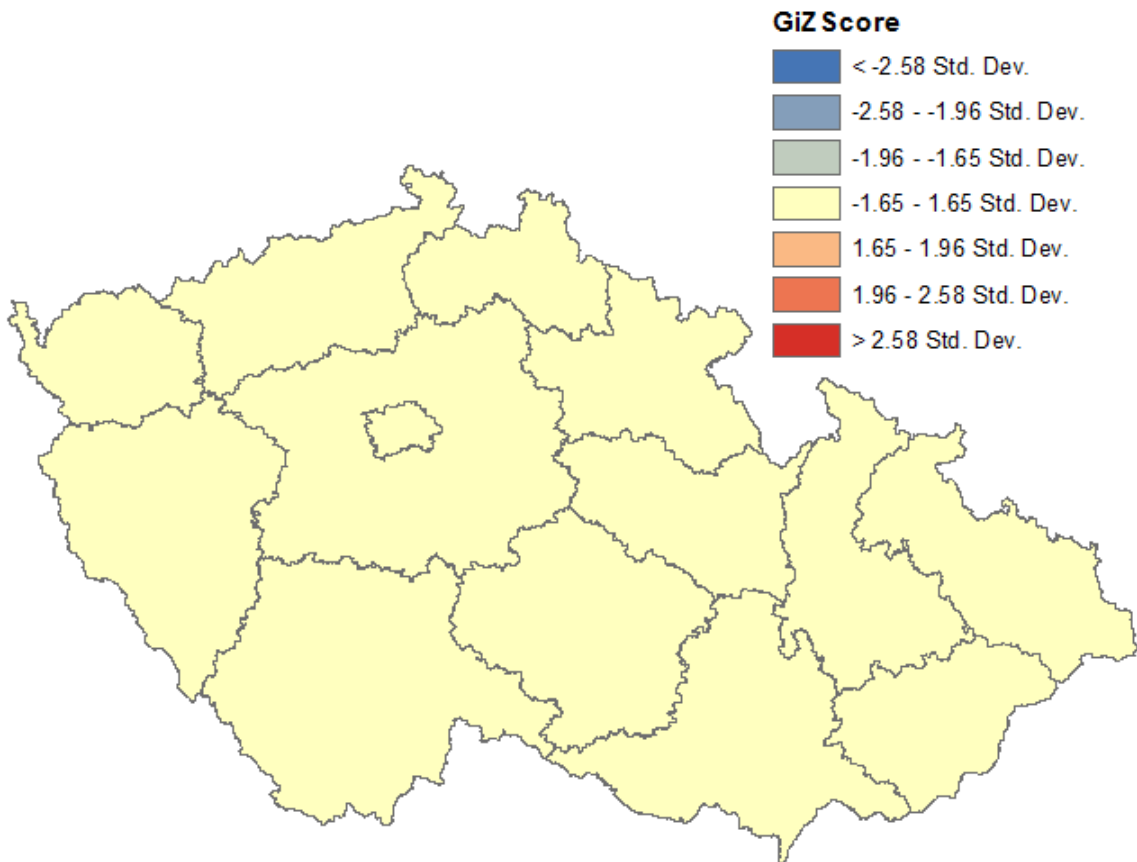
Obrázek 29: Grafický výstup testování prostorové autokorelace na úrovni krajů



Zdroj: Vlastní zpracování v ArcGIS 10.0

Obrázek 29 a 30 zachycují výsledky testů shlukování standardizovaných reziduí. Rezidua jsou shlukována náhodně a hot spots analýza vyhodnotila, že neexistují žádné polygony nízkých ani vysokých hodnot. Výstup přesto může být diskutabilní, neboť pro hot spots analýzu je doporučeno použití minimálně třiceti vstupů.

Obrázek 30: Hot spots analýza standardizovaných reziduí na úrovni krajů

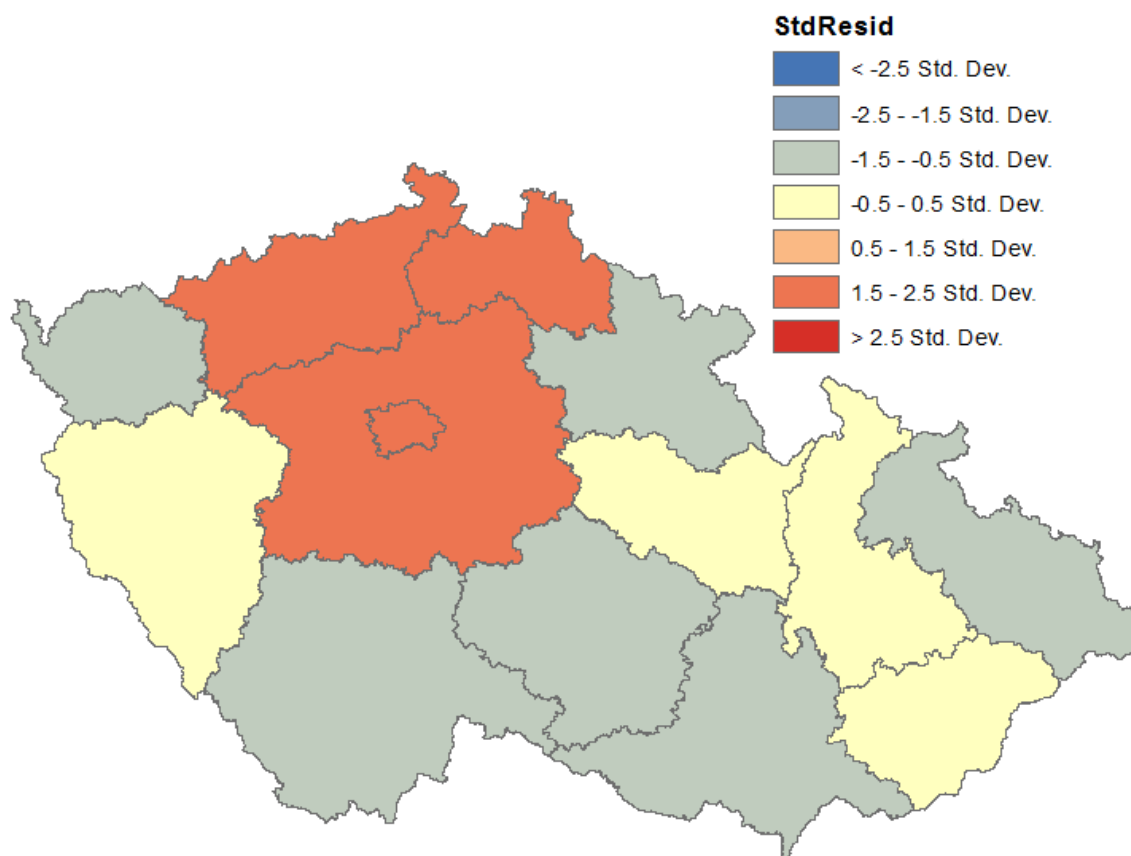


Zdroj: Vlastní zpracování v ArcGIS 10.0

4.1.8 Kraje – lokální model

Pro sestavení lokálního modelu na úrovni krajů bylo třeba vyřadit proměnné hustoty zalidnění a hustoty silniční sítě s vysokým VIF. Korelační koeficient těchto proměnných dosáhl hodnoty 0,908. Na úrovni krajů geograficky vážená regrese zlepšení nepřinesla. Nejlepšího výsledku dosáhl nejspíše modely 1 a 3 s fixním jádrem. Pro výběr posloužilo Aikakeho informačního kritérium a korigovaný koeficient determinace.

Obrázek 31: Standardizovaná residua lokálního modelu s fixním jádrem na úrovni krajů



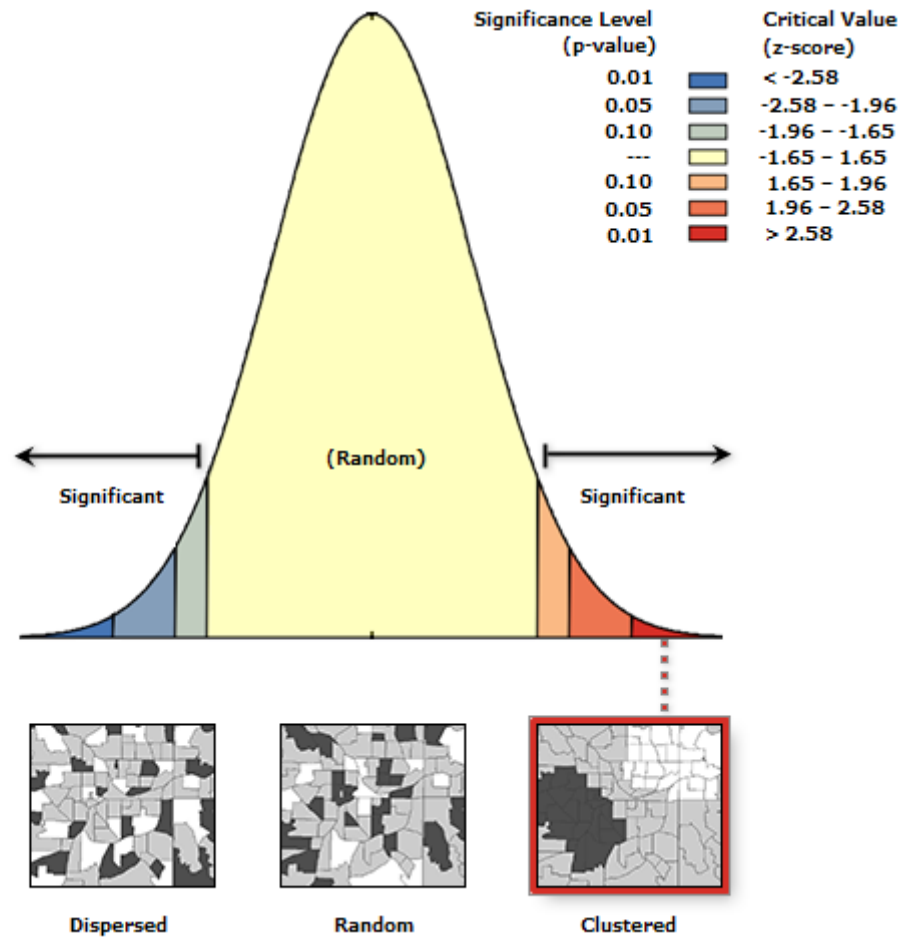
Zdroj: Vlastní zpracování v ArcGIS 10.0

Tabulka 14: Diagnostika lokálního modelu 1 na úrovni krajů

	1	2	3	4
Diagnostika	Fixed - AIC	Adaptive - AIC	Fixed - CV	Adaptive - CV
Bandwidth	4390966,462		4390966,462	
ResidualSquares	21,892	13,645	21,892	13,645
EffectiveNumber	4,008	7,276	4,008	7,276
Sigma	1,480	1,424	1,480	1,424
AICc	63,512	74,043	63,512	74,043
R2	0,817	0,886	0,817	0,886
R2Adjusted	0,762	0,780	0,762	0,780
Neighbors	30	14	30	14

Zdroj: Vlastní zpracování

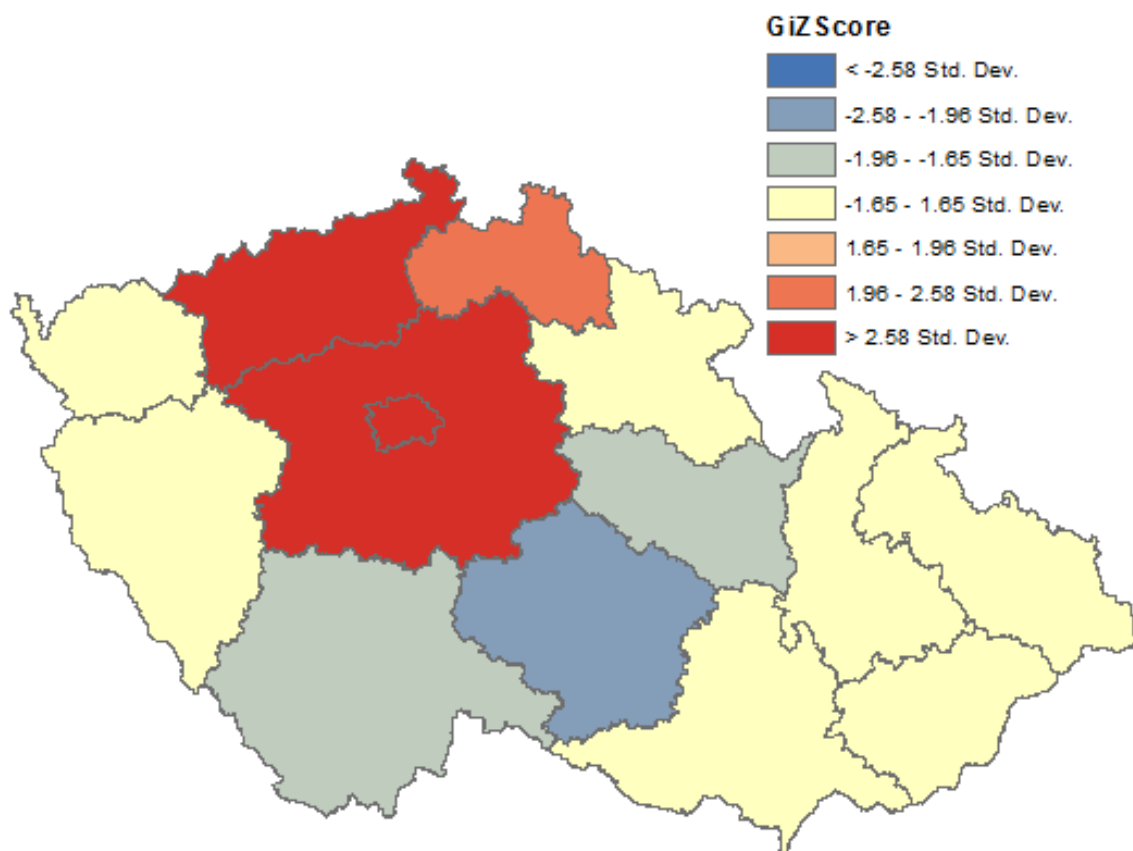
Obrázek 32: Grafický výstup testování prostorové autokorelace lokálního modelu 1 s fixním jádrem na úrovni krajů



Zdroj: Vlastní zpracování v ArcGIS 10.0

Kladná hodnota 0,849 Moranova I prokázala významnou pozitivní autokorelaci. Standardizovaná rezidua mají tendenci ke shlukování podobných hodnot v prostoru. Ke stejnému závěru vede i hot spots analýza, která jednoznačně prokazuje shlukování (viz Obrázek 33). Shluk vysokých hodnot představuje Praha, Středočeský kraj a severní Čechy. Nízké hodnoty pak jižní a západní Čechy. Výsledky opět nejsou optimální, neboť pro hot spots analýzu je doporučeno minimálně 30 vstupů.

Obrázek 33: Hot spots analýza standardizovaných reziduí modelu 1 s fixním jádrem na úrovni krajů



Zdroj: Vlastní zpracování v ArcGIS 10.0

5 Závěr

Cílem této práce byla aplikace geograficky vážené regrese při modelování regionálních rozvojových procesů.

První stanovená hypotéza tvrdí, že použití geograficky vážené regrese na prostorových datech může vést k lepším výsledkům v případě, že jsou porušeny předpoklady pro použití klasického regresního modelu. Tuto hypotézu se v práci podařilo potvrdit, byť ne v plném rozsahu testování.

Jedním z typických znaků prostorových dat je prostorová autokorelace. Dalo se proto předpokládat, že použití GWR, jež dovoluje odhadovat lokální podobu pohyblivých regresních parametrů, zlepší výkonnost modelu. Pro komparaci regresních modelů jsem zvolil ukazatele Akaikeho informačního kritéria (AIC) a korigovaného koeficientu determinace (AdjR2).

Na úrovni obcí jednoznačně vykazoval GWR model lepší výsledky než OLS model. Korigovaný koeficient determinace byl pro OLS pouze 0,166 proti 0,224 u GWR. Rovněž hodnota Akaikeho informačního kritéria mluví ve prospěch použití GWR, kdy u GWR činila 57927,67 a u OLS 58150,47. Ke stejnému závěru lze dojít i u obcí s rozšířenou působností, kde opět byl lokální model výkonnější. (AIC 1121 OLS a 1110 GWR, AdjR2 0,543 OLS a 0,602 GWR).

U modelu okresů již tolik patrný rozdíl ve výkonnosti nebyl. Hodnota AIC vyšla lépe u OLS (451,76 vs. 455,00 GWR). Hodnota korigovaného koeficientu determinace však vyšla lépe u lokálního modelu, a to 0,28 proti 0,27 u globálního. Rovněž u GWR byla hodnota Moranova *I* menší (0,122 vs. 0,137), což znamená menší tendenci ke shlukování standardizovaných reziduí. Proto bych se přiklonil k modelu GWR jako více vhodnému.

Pouze na úrovni krajů nedošlo použitím GWR k nárůstu výkonnosti modelu a jako vhodnější metoda se ukázala OLS. (AIC 51,77 OLS a 63,512 GWR, AdjR2 0,809 OLS a 0,762 GWR). V tomto modelu figurovalo pouze 8 územních jednotek, což se zřejmě negativně projevilo na výsledcích GWR. Rovněž sestavený model měl problémy s multikolinearitou a bylo nutné z modelu některé z proměnných odstranit jako redundantní. V tomto ohledu by bylo vhodné model jinak sestavit s důrazem na dostupnost dat na všech řešených územních úrovních. V době dělání práce však nebyla k mání lepší socioekonomická data, která by se dala vztáhnout k regionálnímu rozvoji.

Druhá hypotéza souvisí s problémem modifikovatelné územní jednotky (MAUP) – se zmenšováním počtu jednotek do vyšších celků dochází ke zvyšování kvality modelu. Daný problém souvisí s tím, že na nejmenší úrovni jednotek může docházet k velkým výkyvům. Tyto výkyvy by se měly sdružováním do vyšších celků snižovat. Tento předpoklad byl vyvrácen u modelu okresů, kde došlo k poklesu korigovaného koeficientu determinace při přechodu z úrovně ORP. U globálního modelu to bylo z 0,542 na 0,271, u lokálního modelu pak z 0,602 na 0,280. Druhou hypotézu proto lze zamítnout.

I. Seznam použitých zdrojů

- Bogard, M. (2011). *Econometric sense: November 2011*. Načteno z Econometric Sense: http://econometricsense.blogspot.cz/2011_11_01_archive.html
- Český statistický úřad. (2015). *Zaměstnanost a nezaměstnanost podle výsledků VŠPS - Metodika*. Načteno z Český statistický úřad: http://notes.czso.cz/csu/redakce.nsf/i/zam_vsps
- Environmental Systems Research Institute. (2015). *Esri - GIS Mapping Software, Solutions, Services, Map Apps, and Data*. Načteno z Esri Support: <http://support.esri.com/en/>
- Fortheringham, S., & Charlton, M. (2009). *Geographically weighted regression - White paper*. Načteno z http://gwr.nuim.ie/downloads/GWR_WhitePaper.pdf
- Fotheringham, S., Brunson, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Hoboken: 2002.
- Fotheringham, S., Brunson, C., Charlton, M., & Rogerson, P. (2009). *The SAGE handbook of spatial analysis*. London: SAGE Publications.
- Klufová, R., Rost, M., & Klicnarová, J. (2012). *Modelování regionálních procesů*. Praha: Alfa nakladatelství.
- Kyratso, M., & Yiorgos, P. (2004). *Defining a geographically weighted regression model of urban evolution*. Načteno z <http://www-sre.wu-wien.ac.at/ersa/ersaconfs/ersa04/PDF/507.pdf>
- Legg, R., & Bowe, T. (2009). *Applying geographically weighted regression to a real estate problem*. Načteno z Esri - GIS Mapping Software, Solutions, Services, Map Apps, and Data: http://www.esri.com/news/arcuser/0309/files/re_gwr.pdf
- Matthews, S. A., & Yang, T.-C. (2012). *Mapping the results of local statistics: Using geographically weighted regression*. Načteno z <http://www.demographic-research.org/volumes/vol26/6/26-6.pdf>
- Míchal, I. (1994). *Ekologická stabilita*. Brno: Veronica, ekologické středisko ČSOP.
- Netrdová, P., & Nosek, V. (2009). Přístupy k měření významu geografického rozměru společenských nerovnoměrností. *Sborník české geografické společnosti*.

- Nissi, E., Sarra, A., & Palermi, S. (2013). *Application of GWR for assessing spatial nonstationarity of residential radon concentration*. Načteno z Società italiana di statistica: <http://new.sis-statistica.org/wp-content/uploads/2013/10/CO09-Application-of-GWR-for-assessing-spatial-nonstationarity.pdf>
- Oliver, L. (2001). *The Modifiable Areal Unit Problem*. Načteno z Department of Geography: http://ibis.geog.ubc.ca/courses/geog570/talks_2001/scale_maup.html
- Openshaw, S. (1983). *The modifiable areal unit problem*. Norwick: Geo Books.
- Pendakur, K. (2010). *Heteroskedasticity and correlations across errors*. Načteno z University Simon Fraser: <http://www.sfu.ca/~pendakur/teaching/buec333/Heteroskedasticity%20and%20Correlations%20Across%20Errors.pdf>
- United States Geological Survey. (2007). *What is a GIS*. Načteno z U. S. Geological Survey: http://webgis.wr.usgs.gov/globalgis/tutorials/what_is_gis.htm

II. Summary

This thesis deals with the modelling of applying techniques of ordinary least squares method and geographically weighted regression for all administrative divisions of the Czech Republic using ArcGIS software. Spatial data are data containing geographic information and they are often handled by using geographic information system. Spatial data also have specific characteristics for processing which requires use of spatial statistical methods. There are often used methods that may not be suitable for geoprocessing.

In the practical part application of linear regression and geographically weighted regression is demonstrated. Dependent variable is the net migration rate. Predictors are unemployment, housing density, road density and coefficient of ecological stability. The multicollinearity problem was found during work with ArcGIS, so some variables had to be removed. All calculations were performed using the ArcGIS with database compiled from publicly available sources.

The first hypothesis claims that the use of geographically weighted regression for spatial data can lead to better results if assumptions for the use of classical regression model are violated. This hypothesis was confirmed. There is a comparison of models using Akaike information criterion and adjusted coefficient of determination. These indicators identifies GWR models as more powerful in all forms of administrative structure except regions (NUTS 3).

The second hypothesis is focused on modifiable areal unit problem. It claims that reducing the number of units to bigger area units leads to quality improvement of model. This assumption was refuted by the model of districts, which was less powerful than the previous model of the ORP. The second hypothesis was rejected.

Keywords: Ordinary least squares, geographically weighted regression, spatial statistics, spatial autocorrelation, regional development.

III. Seznam obrázků a tabulek

Seznam obrázků

Obrázek 1: Základní složky GIS

Obrázek 2: Vztah GIS k ostatním informačním technologiím

Obrázek 3: Znázornění metody nejmenších čtverců

Obrázek 4: Schéma fungování funkce prostorových vah

Obrázek 5: Prostorově vážená funkce s fixní šířkou jádra

Obrázek 6: Prostorově vážená funkce s adaptivní šířkou jádra

Obrázek 7: Základní typy prostorového shlukování dat

Obrázek 8: MAUP – scale effects a zoning effect

Obrázek 9: Prodejní ceny domů modelovaných pomocí metody GWR v ArcGISu

Obrázek 10: Standardizovaná residua globálního modelu na úrovni obcí

Obrázek 11: Grafický výstup testování prostorové autokorelace na úrovni obcí

Obrázek 12: Hot spots analýza standardizovaných reziduí na úrovni obcí

Obrázek 13: Standardizovaná residua lokálního modelu 1s fixním jádrem na úrovni obcí

Obrázek 14: Grafický výstup testování prostorové autokorelace lokálního modelu 1 s fixním jádrem na úrovni obcí

Obrázek 15: Hot spots analýza standardizovaných reziduí modelu 1s fixním jádrem na úrovni obcí

Obrázek 16: Standardizovaná residua globálního modelu na úrovni ORP

Obrázek 17: Grafický výstup testování prostorové autokorelace na úrovni ORP – globální model

Obrázek 18: Hot spots analýza standardizovaných reziduí globálního modelu na úrovni ORP

Obrázek 19: Standardizovaná residua lokálního modelu 1s fixním jádrem na úrovni ORP

Obrázek 20: Grafický výstup testování prostorové autokorelace lokálního modelu 1 s fixním jádrem na úrovni ORP

Obrázek 21: Hot spots analýza standardizovaných reziduí modelu 1s fixním jádrem na úrovni ORP

Obrázek 22: Standardizovaná residua globálního modelu na úrovni okresů

Obrázek 23: Grafický výstup testování prostorové autokorelace na úrovni okresů

Obrázek 24: Hot spots analýza standardizovaných reziduí na úrovni okresů

Obrázek 25: Standardizovaná residua lokálního modelu 1 s fixním jádrem na úrovni okresů

Obrázek 26: Grafický výstup testování prostorové autokorelace lokálního modelu 1 s fixním jádrem na úrovni okresů

Obrázek 27: Hot spots analýza standardizovaných reziduí modelu 1 s fixním jádrem na úrovni okresů

Obrázek 28: Standardizovaná residua globálního modelu na úrovni krajů

Obrázek 29: Grafický výstup testování prostorové autokorelace na úrovni krajů

Obrázek 30: Hot spots analýza standardizovaných reziduí na úrovni krajů

Obrázek 31: Standardizovaná residua lokálního modelu s fixním jádrem na úrovni krajů

Obrázek 32: Grafický výstup testování prostorové autokorelace lokálního modelu 1 s fixním jádrem na úrovni krajů

Obrázek 33: Hot spots analýza standardizovaných reziduí modelu 1 s fixním jádrem na úrovni krajů

Seznam tabulek

Tabulka 1: Souhrn výsledků globálního modelu na úrovni obcí

Tabulka 2: Diagnostika globálního modelu na úrovni obcí

Tabulka 3: Diagnostika lokálního modelu na úrovni obcí

Tabulka 4: Souhrn výsledků globálního modelu na úrovni ORP

Tabulka 5: Diagnostika globálního modelu na úrovni ORP

Tabulka 6: Diagnostika lokálního modelu na úrovni ORP

Tabulka 7: Souhrn výsledků globálního modelu na úrovni okresů

Tabulka 8: Hodnoty korelačních koeficientů mezi ukazateli na úrovni okresů

Tabulka 9: Diagnostika globálního modelu na úrovni okresů

Tabulka 10: Diagnostika lokálního modelu na úrovni okresů

Tabulka 11: Souhrn výsledků globálního modelu na úrovni krajů

Tabulka 12: Hodnoty korelačních koeficientů mezi ukazateli na úrovni krajů

Tabulka 13: Diagnostika globálního modelu na úrovni krajů

Tabulka 14: Diagnostika lokálního modelu 1 na úrovni krajů