

PAVLÍNA SPURNÁ

GEOGRAFICKY VÁŽENÁ REGRESE: METODA ANALÝZY PROSTOROVÉ NESTACIONARITY GEOGRAFICKÝCH JEVŮ

P. Spurná: *Geographically weighted regression: method for analysing spatial non-stationarity of geographical phenomenon.* – Geografie–Sborník ČGS, 113, 2, pp. 125–139 (2008). – The article deals with one of the new quantitative method used in geography, geographically weighted regression (GWR). This method is based on the premise that relationships between variables might not be constant across the study area and explores this phenomenon called spatial non-stationarity. Using the GWR technique to study voting behaviour in Czechia in the parliamentary election in 2002, it is evident that there is a significant difference between the linear regression and GWR models. The examples highlight the relevance and usefulness of GWR and show how it can improve geographical research and potentially also our understanding of geographical processes.

KEY WORDS: geographically weighted regression – quantitative methods – spatial analysis – parliamentary elections.

1. Úvod

Vývoj využití kvantitativních metod v geografii úzce souvisí se změnami paradigmatických směrů a reflektuje obecně vývoj v rámci geografie, potažmo celé vědě (Poon 2004). Období kvantitativní revoluce vycházející z filozofie pozitivismu a glorifikující možnosti kvantitativního přístupu bylo překonáno a nastala doba někdy až příliš kritického pohledu na využití kvantitativních metod v geografickém výzkumu. Z mnoha odborných publikací a článků však vyplývá, že vývoj na poli kvantitativních metod jde neustále dopředu. Vždyť rostoucí technické možnosti usnadňují výpočty složitějších modelů a v mnoha případech umožňují alespoň částečné řešení problémů souvisejících s kvantitativní analýzou geografických dat. Vznikají tak i nové metody, které lze vzhledem k jejich povaze právem označit za geografické.

V posledních dvaceti letech a především v posledním desetiletí došlo v rámci kvantitativní geografie k většímu množství výrazných změn. Nejedná se přitom pouze o vývoj nových metod, ale spíše o obecné změny metodologického charakteru, tedy ve způsobu přístupu ke geografické realitě a využitelnosti kvantitativních metod ve výzkumu. Na kvantitativní geografii lze z tohoto pohledu nahlížet jako na relativně rychle se vyvíjející disciplínu, která si zaslouží větší pozornost, než je jí všeobecně přisuzována. Kvantitativní geografii definují Fotheringham a kol. (2000) jako obor zabývající se analýzami číselných prostorových dat a tvorbou a testováním matematických modelů procesů, které mají významné prostorové vazby, přičemž cílem všech těchto činností je přispět k hlubšímu porozumění těmto procesům. Mezi hlavní trendy, které charakterizují současný přístup ke kvantitativním analýzám, patří

zaměření na explorační povahu analýz, metody lokální analýzy a důraz na vizualizaci dat v souvislosti s geografickými informačními systémy (Fotheringham 1997b, 1998, 1999). Metody tzv. moderní kvantitativní geografie se liší od tradičních metod skutečností, že byly vyvinuty s přihlédnutím ke specifickým vlastnostem geografických dat, zejména jejich prostorové povaze.

Hlavním cílem příspěvku je představit možnosti, jež nabízí aplikace jedné z novějších kvantitativních metod, která svým charakterem zcela odpovídá současným trendům v rámci moderní kvantitativní geografie. Jedná se o metodu s názvem geograficky vážená regrese (geographically weighted regression, dále jen GWR) vyvinutou v 90. letech 20. století. Ta umožňuje celkem jednoduchým způsobem analyzovat na lokální úrovni nejrůznější prostorové vztahy. Vzhledem k relativně velkému citačnímu ohlasu, který vzbudila v geografické literatuře (Huang, Leung 2002; Kyratso, Yiorgos 2004; Lloyd, Shuttleworth 2005; Mittal a kol. 2004), má GWR předpoklady stát se jednou z běžně používaných metod geografického výzkumu. Kromě diskuze vhodnosti použití této nové metody a jejího metodologického založení je náplní příspěvku také příklad aplikace metody GWR v rámci geografické analýzy volebního chování obyvatel v Česku. Důvodem pro zvolení problematiky regionální diferenciací a podmíněností volebního chování jako ukázkového příkladu byla zejména dostupnost dat potřebných pro prováděné analýzy na úrovni obcí a dále relativně četná literatura věnovaná analýze volebního chování pomocí novějších kvantitativních metod (například Jones a kol. 1992, Jones 1997, Kouba 2007). Získané výsledky pomocí metody GWR jsou porovnány se závěry získanými pomocí běžně používané metody mnohonásobné lineární regresní analýzy, z jejíhož metodologického rámce GWR vychází. Tento postup umožnil zhodnocení jak možností, tak i omezení této metody.

2. Problémy agregátní analýzy prostorových dat

Aplikace statistických metod v geografii je spjata s množstvím více či méně závažných metodologických obtíží a problémů. Kromě skutečnosti, že prosté přenášení statistických metod vypracovaných pro jiné vědní disciplíny vede často k velkému zjednodušení geografické reality, komplikuje použití statistických metod v geografii samotná prostorová povaha geografických dat. Vedle obtíží spojených se základními problémy použití statistických metod jako je výběr vhodných proměnných, nepravá korelace (pseudokorelace) či identifikace kauzálního vztahu, které jsou společné mnoha empirickým vědám, se tak při analýze geografických dat setkáme i s dalšími problémy souvisejícími se samotnou specifičností předmětu geografického zkoumání. Právě prostorová povaha geografických dat způsobuje četné specifické potíže, které v mnohém ztěžují použití klasických statistických metod.

Jako nejvýznamnější problémy či specifika analýzy geografických dat lze označit závislost výsledků analýz na agregaci dat neboli způsobu vymezení prostorových jednotek ve spojení s ekologickou chybou, prostorovou autokorelací a prostorovou nestacionaritou (Fotheringham 1997a, Robinson 1998, Spurná 2006). Anselin (1988) uvádí, že agregovaná prostorová data jsou charakteristická prostorovou závislostí a heterogenitou, přičemž právě tyto prostorové efekty mají za následek nemožnost aplikace standardních statistických a ekonometrických metod a vyžadují vývoj specializovaného souboru technik.

Identifikace uvedených problémů statistické analýzy geografických dat a obtíží s aplikací převzatých metod, které vycházejí z méně komplexních věd,

je významná pro změnu pohledu na využití kvantitativních metod v geografickém výzkumu. Do popředí jsou nově stavěny otázky zaměřující se více na hledání odlišností než podobností ve zkoumaných vztazích, důraz je kladen spíše na explorační povahu statistických analýz než na „jednobarevné“ výsledky statistického usuzování. Vznikla například celá skupina metod lokální analýzy dat, které jsou ovlivněny právě vědomím o důležitosti prostorové nestacionarity.

2.1 Prostorová nestacionarita

Prostorová heterogenita, pro kterou se v geografické literatuře při studiu sociálních procesů lokalizovaných v prostoru používá také pojem prostorová nestacionarita (Fotheringham 1997b, Fotheringham a kol. 2002), se týká nestability zkoumaných jevů a vztahů v prostoru. Zatímco při studiu fyzických procesů lze předpokládat určitý stupeň homogenity či stacionarity (například obecně platný vzorec pro výpočet velikosti gravitační síly), u sociálních procesů tomu tak není. Procesy studované v geografii často probíhají v několika různých geografických oblastech, které jsou charakteristické různou mírou příznivosti pro daný proces (Undin, Unwin 1998). Při použití regresní analýzy tak můžeme získat jiný funkční vztah mezi dvěma proměnnými v závislosti na charakteru určitého regionu atd. Například vztah mezi cenou bytu a hustotou zalidnění v jeho okolí může být prostorově velmi odlišný, neboť na cenu bytu má vliv množství dalších charakteristik, které uvedený vztah modifikují.

Důvody, které způsobují prostorovou nestacionaritu, lze shrnout do tří okruhů (Fotheringham a kol. 2000). První příčina prostorové nestacionarity je ryze statistického charakteru a nevztahuje se k samotnému analyzovanému procesu. Při analýze vztahu dvou proměnných oddělené za dvě či více prostorových jednotek nelze očekávat, že získané regresní koeficienty budou zcela totožné. Existence relativně malých rozdílů je tedy přímým následkem rozdílných souborů dat, která byla v analýze použita. Další dva důvody prostorové nestacionarity se týkají již samotného zkoumaného vztahu a z hlediska geografického výzkumu jsou nesrovnatelně důležitější, neboť zároveň odpovídají dvěma protichůdným přístupům k používání kvantitativních metod v geografii.

Druhým důvodem prostorové nestacionarity je pravá podstata analyzovaných vztahů či závislostí, které se za určitých okolností mohou v prostoru skutečně odlišovat. Často se v této souvislosti hovoří o kontextuálních faktorech ovlivňujících chování jedinců a dalších aktérů. Vliv kontextuálních faktorů je dobře dokumentován zejména ve volební geografii (Kostelecký, Čermák 2004). Uvedená myšlenka, že se chování jednotlivců skutečně může měnit v prostoru, je v souladu s postmoderními přístupy v geografii, které zdůrazňují důležitost místa a lokálních podmínek jako rámce umožňujícího pochopení určitého chování (viz Holt-Jensen 1999). Z tohoto pohledu je identifikace lokálních odchylek v analyzovaném vztahu velmi užitečná, neboť poskytuje vodítko, kam zaměřit kvalitativní výzkum.

Třetí důvod prostorové nestacionarity odpovídá pozitivistické škole myšlení a je v rozporu s předcházejícím důvodem, neboť předpokládá, že o zkoumaných vztazích můžeme formulovat obecná tvrzení platná pro jakékoliv místo a čas (Fotheringham a kol. 2000). Nedostatečná přesnost a nesprávnost navrhovaných modelů, z nichž jsou vztahy odvozovány a které vedou k velkému zjednodušení reality, nám pouze nedovoluje tyto pravidelnosti nalézt. Identifikace lokálních odchylek v analyzovaném vztahu je na rozdíl od předešlého přístupu užitečná k hlubšímu porozumění nedostatkům navrhovaného modelu a má vý-

zkumníka upozornit na proměnné, které do modelu nezahrnul a jejichž uvažování zlepší statistický model. Uvedený pozitivistický přístup k prostorové nestacionaritě je častý v ekonomické literatuře (Anselin 1988).

V konkrétním geografickém výzkumu je nutné uvažovat všechny tři zmíněné příčiny prostorové nestacionarity. V případě zjištění výrazných lokálních odchylek od obecně navrženého modelu, které nelze přičíst statistické chybě, je teprve na základě dalších znalostí o zkoumaném jevu a jeho podstatě možné se přiklonit k jednomu z nich. V každém případě samotné zjištění míry prostorové stability navrženého modelu obohatí naše znalosti o zkoumaném jevu či procesu a může pomoci odhalit jeho vlastní povahu včetně příčin.

Ze statistického pohledu lze na problematiku prostorové nestacionarity nahlížet jako na problém třetí proměnné, kterou je v tomto případě umístění v prostoru (Hendl 2004, Disman 2005). Například pro popsání vztahu mezi dvěma proměnnými lze rozdělit data dle vybraného regionálního členění, sestavit parciální vztahy a porovnat vztahy nalezené v jednotlivých regionech se vztahem původním. Možností, jež mohou nastat, je více (viz Hendl 2004), přičemž pro analýzu dat je nejdůležitější tzv. Simpsonův paradox, který označuje obrácení závislosti na lokální úrovni, když parciální vztahy jsou silnější než vztah původní. I když je Simpsonův paradox obvykle demonstrován na neprostorových datech, kde dochází k agregaci subpopulací, příklady z geografické literatury (Fotheringham a kol. 2002, Spurná 2006) dokazují, že platí rovněž pro agregovaná prostorová data, a poukazují tak na nebezpečí spojené s analýzou dat za větší území. Může se totiž stát, že nalezený závěr bude pravým opakem reálných vztahů, které platí na lokální úrovni.

Je zřejmé, že popsaná podrobnější analýza dat je přesnější a přínosnější než pouhé zkoumání původního vztahu bez sledování prostorového efektu. Základní přístupy napomáhající k odhalení prostorových změn ve zkoumaném vztahu, a tedy postupy sloužící k zohlednění prostorové nestacionarity včetně konkrétních příkladů uvádí Spurná (2006). Již znázornění regresních reziduí a následný výpočet jejich prostorové autokorelace z velké části odhalí přesnost a spolehlivost regresního modelu. Výsledky dalších obvyklých postupů, mezi něž patří například provedení výpočtu regresního modelu zvlášť pro jednotlivá diskrétní území nebo využití metody regrese pohyblivých oken (Spurná 2006), závisí do velké míry na předem provedeném vymezení území, za něž jsou regresní modely kalibrovány. O překonání těchto nedostatků se snaží metody lokální analýzy zohledňující prostorové efekty na zkoumané vztahy, především metoda geograficky vážené regrese.

3. Metoda geograficky vážené regrese

Metoda geograficky vážené regrese, patřící k exploračním technikám prostorové analýzy dat, se snaží přispět alespoň k částečnému vyřešení problému prostorové nestacionarity. Z tohoto pohledu lze metodu GWR označit jako metodu analýzy prostorově se měnících vztahů, což je i v podtitulu knihy vydané o této metodě jejími tvůrci (Fotheringham a kol. 2002). Pro praktické použití GWR byl jejími autory vyvinut uživatelsky jednoduše ovladatelný speciální software GWR 3.0¹, jehož výstupy lze snadno propojit nejen se standardními statistickými programy jako je například SPSS, ale také se softwary GIS jako ArcMap.

¹ GWR 3.0 je zpoplatněný software pro výpočet modelů GWR a dalších geograficky vážených statistik vyvinutý v roce 2003. Bližší informace viz oficiální internetové stránky o GWR <http://ncg.nuim.ie/ncg/GWR/index.htm>.

Kromě schopnosti analyzovat lokální povahu vztahů ve vícerozměrných datových souborech vykazuje metoda GWR ještě další klady. První výhodou této metody lze spatřovat ve skutečnosti, že je založena na tradičním regresním rámci, který je většinou geografů důvěrně známý. Za další výhodou lze považovat jednoznačný způsob zahrnutí lokálních prostorových vztahů do regresního rámce. Tyto skutečnosti podporují myšlenku o budoucí vysoké využitelnosti této metody v geografickém výzkumu.

3.1 Teoreticko-metodologické základy GWR

Metoda GWR rozšiřuje tradiční regresní rámec o pohyblivé regresní parametry, když umožňuje odhadovat jejich lokální podobu. Obecný regresní model, u něhož předpokládáme prostorovou neměnnost regresních parametrů, může být tedy přepsán do nového tvaru: obecný regresní model:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i,$$

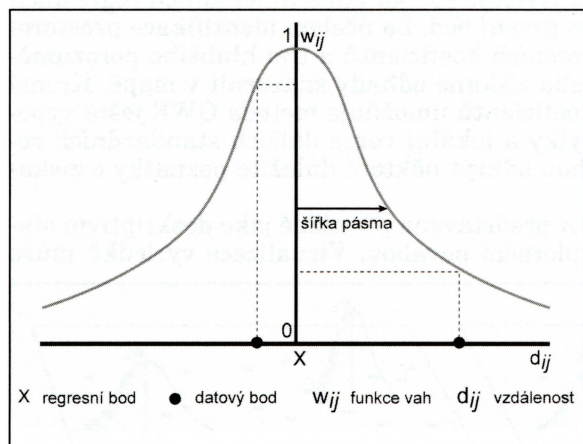
model GWR:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i,$$

kde (u_i, v_i) označuje souřadnice i -tého bodu v prostoru a regresní koeficient $\beta_k(u_i, v_i)$ vyjadřuje hodnotu spojitě funkce $\beta_k(u, v)$ v bodě i . Metoda GWR tedy předpokládá možnost existence prostorových odlišností ve vztazích dvou a více proměnných a poskytuje způsob, jak tyto odchylky měřit. V rámci GWR je regresní analýza provedena pro každý regresní bod zvlášť, čímž jsou získány lokální regresní parametry. Vynesením výsledných odhadů lokálních regresních parametrů do mapy je následně přehledně znázorněn charakter zkoumaného vztahu. Metodologické základy metody GWR včetně potřebného složitějšího statistického aparátu jsou podrobně popsány ve Fotheringham a kol. (2002).

Zde je pozornost soustředěna zejména na aplikační možnosti metody GWR, tedy vstupní charakteristiky modelu, které může uživatel softwaru GWR 3.0 zvolit, a základní výstupy získané pomocí metody GWR.

K výpočtu modelu GWR pro určitý regresní bod se používá spojitě vážící schéma neboli prostorové jádro (viz obr. 1), které zajišťuje vhodné vážení datových bodů na základě vzdálenosti od regresního bodu. Ve své podstatě se nejedná o nic jiného než o funkční vyjádření klesající intenzity prostorových interakcí s rostoucí vzdáleností. Charakter prostorové-



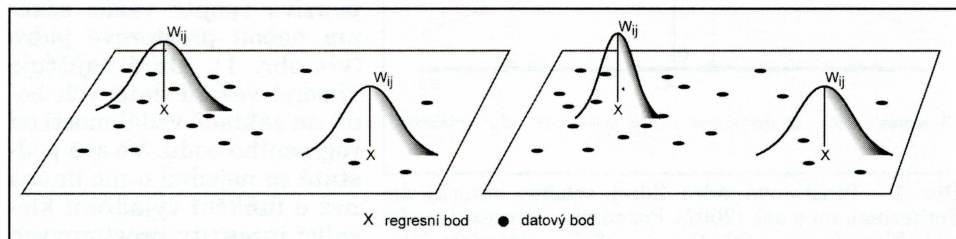
Obr. 1 – Prostorové jádro. Zdroj: schéma autorky dle Fotheringham a kol. (2002). Poznámka: Regresní model je kalibrován na základě geograficky vážených datových bodů (např. obcí), které se nacházejí ve vymezeném území okolo regresního bodu.

ho jádra daný zvolením typu vážící funkce a určením šířky pásma samozřejmě ovlivňuje výsledky GWR, a proto je nutné jeho výběru věnovat pozornost. Fotheringham a kol. (2002) přitom uvádí, že ve skutečnosti jsou výsledky GWR relativně necitlivé k výběru typu vážící funkce, ale jsou ovlivněny zejména volbou šířky pásma b , která je klíčovým problémem kalibrace modelu GWR. Pokud je například zvoleno b příliš velké, váhy všech datových bodů se budou přibližovat hodnotě jedna a všechny odhadované parametry budou víceméně shodné. Model odvozený pomocí GWR tak bude z velké části shodný s obecným regresním modelem. Naopak když bude b zvoleno příliš malé, odhady parametrů se stanou více závislé na datových bodech v blízkosti regresního bodu, což bude mít za následek rostoucí nepřesnost modelu. K určení optimální hodnoty šířky pásma existuje množství statistických kritérií a procedur (viz Fotheringham a kol. 2002). Nejčastěji se při aplikaci metody GWR setkáme s použitím kvartického typu prostorového jádra definovaného pomocí kvartické funkce $w_{ij} = [1 - (d_{ij}/b)^2]^2$ pokud $d_{ij} < b$ a $w_{ij} = 0$ v ostatních případech (Fotheringham a kol. 2002). Jeho výhodou je, že vytváří spojitou vážící funkci podobnou Gaussově křivce až do zvolené vzdálenosti b od regresního bodu, přičemž vzdálenějším datovým bodům přiřazuje nulovou váhu.

Při kalibraci modelu GWR můžeme použít dva základní typy prostorových jader, fixní a adaptivní. Jak vyplývá z jejich názvu a schematického zobrazení na obrázku 2, zatímco u fixního jádra zůstává šířka pásma stálá, u adaptivního se mění v závislosti na hustotě datových bodů v okolí regresního bodu (větší je v oblastech s řidším výskytem datových bodů a naopak menší v místech s jejich vysokou hustotou). Při použití adaptivního jádra tak čelíme problému jednak v oblastech řídkého výskytu datových bodů, kde by model získaný použitím fixního jádra vykazoval velké směrodatné odchylky a v extrémním případě by nemohl být ani kalibrován, a jednak v oblastech s hustým výskytem datových bodů, kde by použití fixního jádra širšího, než je nezbytné nutné, mohlo způsobit větší zkresení získaných odhadů regresních parametrů. Způsobů, jak přistupovat k vytvoření adaptivního prostorového jádra, existuje více (Fotheringham a kol. 2002).

Výsledkem kalibrace modelu GWR je soubor odhadů lokálních regresních koeficientů $\beta_k(u, v)$ pro každý regresní bod. Za účelem identifikace prostorových odchylek v odhadech regresních koeficientů a tím hlubšího porozumění zkoumané závislosti je potřeba získané odhady znázornit v mapě. Kromě odhadů lokálních regresních koeficientů umožňuje metoda GWR ještě vypočítat lokální směrodatné odchylky a lokální verze dalších standardních regresních diagnostik, které mohou odkrýt některé důležité poznatky o získaném modelu.

Až doposud byla metoda GWR představena převážně jako deskriptivní statistická metoda s výraznou explorační povahou. Vizualizace výsledků může



Obr. 2 – Použití fixního a adaptivního prostorového jádra. Zdroj: schéma autorky dle Fotheringham a kol. (2002).

odhalit některé zajímavé odlišnosti ve vztazích geografických dat, přičemž ale nevypovídá nic o významnosti zjištěných rozdílů. Odhady regresních parametrů získané pomocí GWR však mohou být také statisticky testovány, zda vykazují signifikantní prostorové odlišnosti. Na základě měření prostorové variability jednotlivých regresních koeficientů je tak v podstatě testována stacionarita zkoumaného jevu. Přitom pokud sledovaná variabilita postačuje k zamítnutí nulové hypotézy tvrdící, že regresní koeficient je v prostoru neměnný, pak je statisticky potvrzena prostorová nestacionarita. Ke statistickému testování lze v rámci GWR opět využít několika přístupů (viz Fotheringham a kol. 2002). K nejpoužívanějším patří metoda Monte Carlo založená na náhodných permutacích datových bodů a opakování procedury GWR pro každé význlké přeuspořádání hodnot. Porovnáním získaných výstupů na základě experimentálního rozložení lze poté rozhodnout o statistické významnosti prostorových odlišností hodnot určitého regresního koeficientu a tím i o prostorové nestacionaritě zkoumaného vztahu.

3.2 Příklad aplikace GWR ve volební geografii

Konkrétní aplikace, na které je metoda GWR představena, se týká vícerozměrné analýzy územních souvislostí volebních výsledků dvou politických stran ODS a KSCM v parlamentních volbách v roce 2002. Na základě analýzy agregátních dat jsou hledány významné závislosti mezi volebním výsledkem těchto dvou politických stran a základními charakteristikami populací na úrovni obcí a městských částí a jsou studovány možné faktory, které regionální rozdíly ve volební podpoře jednotlivým stranám vysvětlují. Pozornost je přitom zaměřena na komparativní aplikaci metody GWR a běžně používané metody mnohonásobné lineární regrese s cílem srovnání výstupů obou přístupů.

Mnohonásobná regresní analýza dovoluje odhalit souvislosti mezi vysvětlovanou proměnnou a vysvětlujícími proměnnými, přičemž výsledkem je jedno tvrzení o vztahu zkoumaných proměnných pro celé sledované území. Z tohoto pohledu se jedná o globální analýzu (Fotheringham a kol. 2002, Spurná 2006). Metoda GWR jako metoda lokální analýzy vychází z předpokladu, že výše zmíněný výsledek neadekvátně reflektuje vztahy mezi proměnnými, které se proměňují v prostoru.

Mezi základní přístupy, jakými je možno studovat příčiny existujících regionálních rozdílů v politické orientaci obyvatel jednotlivých regionů, patří kompoziční a kontextový přístup (Kostecký, Čermák 2004). Prováděná analýza vychází z názoru, že volební chování je determinováno jak postavením jedince ve struktuře společnosti, tak prostorovým kontextem, ve kterém se jedinec rozhoduje. Jelikož faktorů ovlivňujících rozhodování jedince při volbách je velmi mnoho, tak i soubor agregátních charakteristik, které asociují s volebními výsledky, je značně rozsáhlý. Analýze volebního chování na agregátní úrovni za okresy se věnoval ve svých studiích Kostecký (1992, 2001), jehož volba a výběr proměnných byly podkladem a inspirací pro tuto analýzu. Vysvětlujícími proměnnými jsou ukazatele charakterizující jednotlivé obce, tedy prostředí, ve kterém voliči žijí a ve kterém se rozhodovali při volbách. Jedná se jak o ukazatele popisující strukturu populace jednotlivých obcí z hlediska demografického, národnostního, náboženského či ekonomického, tak charakteristiky prostředí obce jako takové, které nemají přímou vazbu na strukturu populace, nicméně volební rozhodování potenciálně významně ovlivňují. Datovou základnu tvoří výsledky Sčítání lidu, domů a bytů 2001 (SLDB 2001) na úrovni obcí, u Prahy, Brna, Ostravy a Plzně pak městských

Tab. 1 – Regresní modely volebních výsledků ODS a KSČM a vysvětlujících proměnných získané metodou mnohonásobné lineární regrese a GWR

Vysvětlující proměnné	Mnohonásobná lineární regrese			Geograficky vážená regrese				
	r ²	Beta	sig. 1	r ²	b _{min}	b _{medián}	b _{max}	sig. 2
ODS	0,369			0,492				
PODNIK		0,306	0,000		-0,014	0,409	0,792	0,060
VERICI		-0,336	0,000		-0,299	-0,067	0,275	0,000
VS		0,211	0,000		-0,416	0,628	1,505	0,000
NEZAM		-0,150	0,000		-0,585	-0,130	0,542	0,000
KSČM	0,318			0,494				
NEZAM		0,284	0,000		-0,217	0,228	0,841	0,000
VERICI		-0,296	0,000		-0,396	-0,206	0,006	0,000
VS		-0,228	0,000		-2,093	-0,767	0,159	0,000
PODNIK		-0,200	0,000		-0,725	-0,295	0,185	0,000

Zdroj: výpočet autorky v programu SPSS 12.0 a GWR 3.0; SLDB 2001, Volby 2002

Poznámka: r² – koeficient determinace; sig. 1 – statistická významnost; Beta – standardizovaný beta koeficient; b_{min}, b_{medián}, b_{max} – minimální, mediánová a maximální hodnota odhadu regresního koeficientu; sig. 2 – statistická významnost prostorové nestacionarity regresního koeficientu; proměnná PODNIK značí podíl zaměstnavatelů a osob samostatně výdělečně činných na celkovém počtu ekonomicky aktivních obyvatel v obci, VERICI podíl věřících obyvatel v obci, VS podíl obyvatel s vysokoškolským vzděláním na populaci ve věku 15 let a více v obci a NEZAM podíl nezaměstnaných obyvatel v obci.

částí, ve stavu k 1.1.2003. Dohromady je tedy analyzováno 6 364 jednotek, přesněji 6 245 obcí a 119 městských částí.

Pro zhodnocení podmíněnosti volebních výsledků bylo nakonec na základě korelační a shlukové analýzy provedené s cílem eliminace multikolinearity vybráno čtrnáct proměnných. Pro více vysvětlovaných proměnných je analýza provedena v diplomové práci Spurné (2006), kde je také uveden seznam všech proměnných včetně jejich konstrukce a konkrétní postup výběru. Po stanovení proměnných jsou provedeny statistické analýzy pomocí dvou srovnávaných metod, přičemž nejprve je použita lineární regresní metoda postupné regrese („stepwise“). Základní parametry obou regresních modelů získaných pomocí metody lineární regrese i GWR jsou uvedeny v tabulce 1.

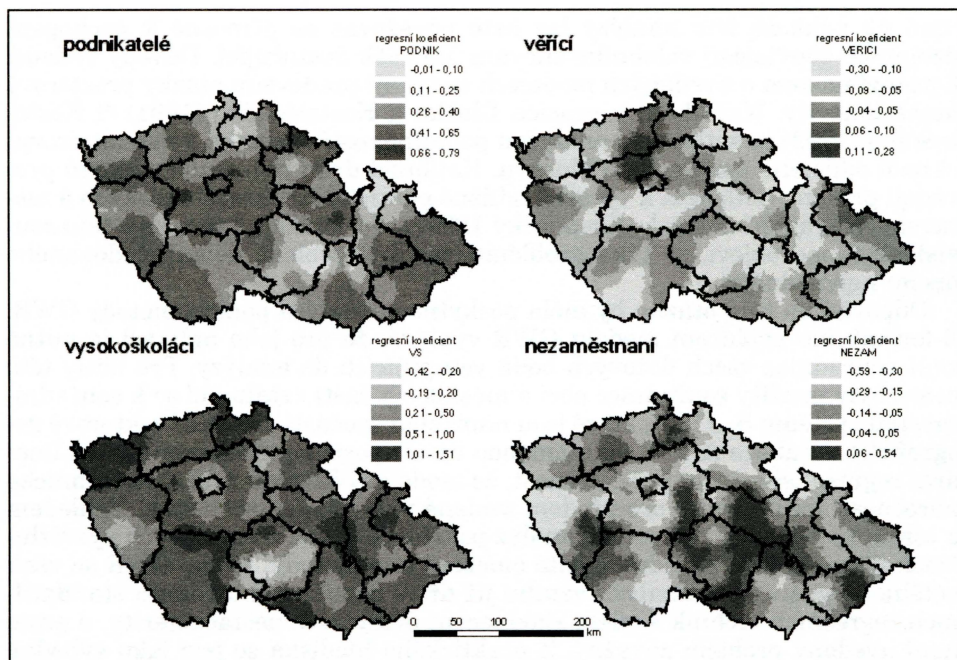
Modely volebního výsledku ODS a KSČM získané pomocí mnohonásobné lineární regrese „stepwise“ jsou tvořeny stejnými čtyřmi proměnnými, přičemž však všechny proměnné vstupující do modelů vykazují podstatně odlišný vliv na volební výsledek těchto dvou politických stran. Celkové procento vysvětlitelné variability je u obou modelů relativně vysoké, přibližně třetinu variability volebního výsledku se modely podařilo vysvětlit. Z regresního modelu pro ODS vyplývá, že ODS byla tím úspěšnější, čím větší podíl obyvatel obce tvořili podnikatelé a vysokoškoláci a naopak čím menší podíl obyvatel se hlásil k náboženskému vyznání a čím méně bylo nezaměstnaných. Nejvýznamnější charakteristikou pro vysvětlení variability volebního výsledku ODS je územní rozložení podnikatelů, přesněji zaměstnavatelů a osob samostatně výdělečně činných, které vysvětluje téměř 18 % celkové variability. Pro vysvětlení variability volebních výsledků KSČM je nejdůležitější podíl nezaměstnaných obyvatel v obci, který vysvětluje celkovou variabilitu téměř z 15 %.

Uvedenými poznámkami a podrobnějšími komentáři k získaným regresním modelům by ve většině případů končil tradiční geografický kvantitativní výzkum pomocí mnohonásobné lineární regrese. Nové poznatky a závěry vytvo-

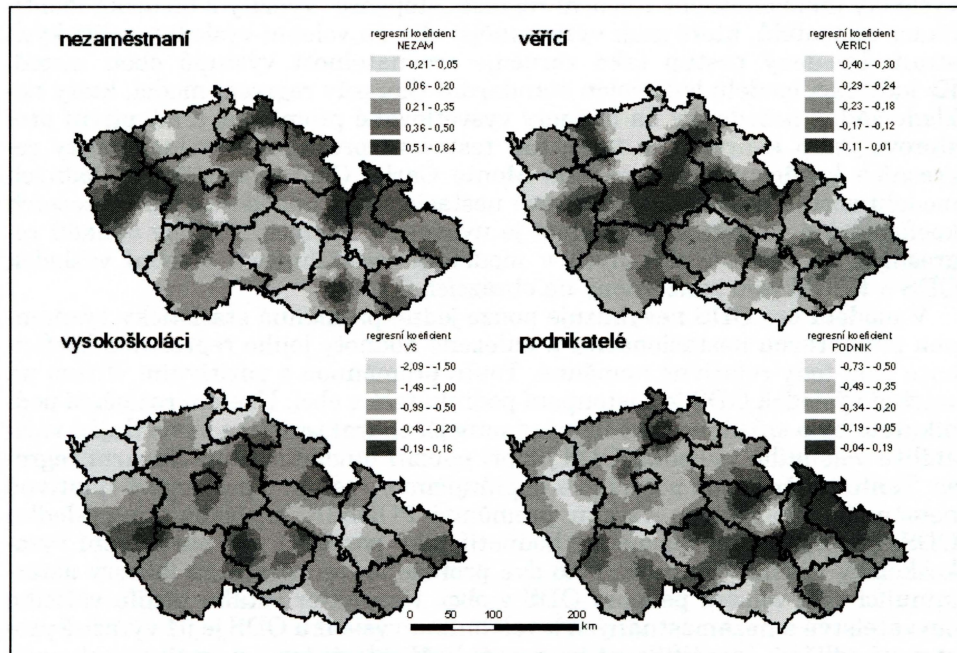
řené na základě této analýzy lze jistě považovat za přínosné k pochopení územních souvislostí volebního chování, ne však dostačující. Důvody vedoucí k pochybnostem o uvedených modelech se týkají především otázky prostorové nestacionarity. Například v pracích Blažka a Kosteleckého (1991) či Kosteleckého (1992) se setkáme s analýzou podmíněnosti výsledků voleb na úrovni okresů odděleně pro Čechy a Moravu. Kalibraci dvou regresních modelů provádějí přitom „vzhledem k výrazně odlišné politické struktuře v českých a moravských krajích“ (Blažek, Kostelecký 1991, s. 7). Otázka, která se v této souvislosti ihned objeví, se týká problému, jestli je zvolené dělení sledovaného území nejvhodnější.

Odpověď na tuto otázku by měla poskytnout analýza pomocí metody GWR. Z formálního vyjádření modelu GWR vyplývá, že pro jeho aplikaci je nutné znát souřadnice všech datových bodů vstupujících do analýzy. Pro účely této práce byly použity souřadnice obcí a městských částí vztahující se k souřadnicovému systému S—JTSK, které jsou například součástí digitální vektorové geografické databáze ArcČR 500. Změnou oproti postupu u mnohonásobné lineární regrese „stepwise“ je skutečnost, že modely GWR jsou z důvodů technické náročnosti kalibrovány pro předem zvolené vysvětlující proměnné. Vzhledem k tomu, že výsledky regresních analýz podstatně závisí na zvolených vysvětlujících proměnných, se jedná o jisté omezení této metody. Model GWR se však většinou využívá pro analýzu vztahu již dříve zkoumaného pomocí standardních regresních technik s cílem zhodnocení prostorové nestacionarity, a proto není uvedený problém závažný. Z praktického hlediska se jeví jako výhodné zahrnout do modelu GWR ty vysvětlující proměnné, které byly získány pomocí mnohonásobné lineární regrese „stepwise“, či kalibrovat větší počet modelů GWR s různými vysvětlujícími proměnnými a následně je porovnat. Zde jsou výsledky mnohonásobné lineární regrese „stepwise“ využity z hlediska identifikace ukazatelů, které mají významnější vliv na volební výsledky politických stran. Uvedený postup také zaručuje srovnatelnost výstupů obou metod. Ke kalibraci modelů byl zvolen standardní Gaussův regresní model, který neklade žádné požadavky na hodnoty vysvětlované proměnné, a adaptivní prostorové jádro kvartického typu, pro testování prostorové nestacionarity regresních koeficientů pak metoda Monte Carlo. Charakteristika výsledných modelů včetně posouzení prostorové nestacionarity odhadovaných regresních koeficientů a rozpětí jejich hodnot je uvedena v tabulce 1. Mapy odhadů regresních koeficientů získaných v modelech vysvětlujících volební výsledek ODS a KSČM jsou znázorněny na obrázcích 3 a 4.

V modelu pro ODS nevykazuje pouze jedna proměnná statisticky významnou prostorovou nestacionarit u a nalezené hodnoty jejího regresního koeficientu jsou tedy relativně neměnné. Touto proměnnou s pozitivním vlivem na volební výsledek ODS je zastoupení podnikatelů v obci. Územní rozložení podnikatelů bylo shledáno jako nejvýznamnější charakteristika vysvětlující variabilitu volebního výsledku ODS již při použití mnohonásobné lineární regrese. Tento závěr je zde potvrzen s doplňujícím tvrzením o prostorově relativně neměnném charakteru vztahu proměnné PODNIK k volebnímu výsledku ODS. Podobně jednoznačně lze hodnotit také pozitivní vliv zastoupení vysokoškolsky vzdělaných osob. Tyto dvě proměnné lze označit za faktory determinující výši volební podpory ODS v obci. Charakter vztahu podílu věřícího obyvatelstva a nezaměstnaných k volebnímu výsledku ODS je již výrazně prostorově odlišný. Identifikovat lze rozsáhlejší oblasti jak s negativní, tak pozitivní územní souvislostí mezi uvedenými dvěma proměnnými a volebním výsledkem ODS.



Obr. 3 – Analýza volebního výsledku ODS v parlamentních volbách v roce 2002 pomocí GWR. Zdroj: výpočet autorky v programu GWR 3.0; SLDB 2001, Volby 2002



Obr. 4 – Analýza volebního výsledku KSČM v parlamentních volbách v roce 2002 pomocí GWR. Zdroj: výpočet autorky v programu GWR 3.0; SLDB 2001, Volby 2002

Oproti modelu pro ODS vyplývá z map odhadů regresních koeficientů získaných v modelu vysvětlujícím volební výsledek KSČM relativní shodnost charakteru zkoumaného vztahu v prostoru. Přestože i v tomto případě vykazují regresní koeficienty prostorovou nestacionaritu (tab. 1), jedná se v naprosté většině pouze o rozdíly v síle vztahu než o změny z pozitivního charakteru vztahu na negativní a naopak. Lze se domnívat, že určitá jednotnost územních souvislostí volební podpory KSČM a vysvětlujících proměnných vypovídá o zřetelnější vyhraněnosti voličů této strany. Obecně lze shrnout, že regresní vztah vysvětlující územní souvislosti volební podpory KSČM nalezený pomocí mnohonásobné lineární regrese lze přijmout za obecně platný.

Analýza volebního chování pomocí metody GWR poskytla nový pohled na analyzované vztahy a především možnost jejich podrobnějšího studia. Kromě proměnné PODNIK u modelu pro ODS byla u všech dalších potvrzena jejich statisticky významná prostorová nestacionarita z hlediska charakteru působení na vysvětlovanou proměnnou. Z tohoto pohledu je možné hodnotit použití modelu GWR jako vhodné, ne-li nutné, neboť bez znalosti rozdílného působení určitých vysvětlujících proměnných bychom mohli snadno sklouznout k závěru, že výsledky nalezené pomocí tradiční mnohonásobné lineární regrese jsou obecně platné, což by bylo nesprávné. Vhodnost modelu GWR dokazují i hodnoty koeficientu determinace, neboť u obou modelů kalibrovaných pomocí metody GWR došlo ke zvýšení procenta vysvětlené variability modelem téměř na 50 %.

Vizuální analýza výsledků získaných pomocí metody GWR vypovídá mnohé nejen o celkové variabilitě hodnot odhadů regresních koeficientů, a tedy proměnlivosti charakteru vztahů v prostoru, ale i o existenci rozdílných regionů z hlediska zkoumaných vztahů. Vymežit tak v mnoha případech lze oblasti, které jsou charakterizovány stejnými či podobnými vztahy mezi proměnnými, nebo naopak regiony s překvapivě odlišným působením vysvětlujících faktorů. Kalibrace regresního modelu zvláště pro Čechy a Moravu je z tohoto pohledu velmi zjednodušující. Výsledky získané metodou GWR také nepřímou dokládají, že k vysvětlení volebního chování obyvatel nestačí vycházet pouze z kompozičního přístupu. Nutné je uvažovat také kontextuální přístup, který klade důraz na prostorový kontext, v němž se jedinec rozhoduje. Na závěr je potřeba uvést, že metoda GWR byla použita v exploračním duchu a hypotézám vyplývajícím ze závěrů analýz by bylo jistě přínosné se věnovat v dalším výzkumu.

4. Závěr

Metoda GWR patří do skupiny metod lokální analýzy dat, které se obecně snaží zachytit procesy odehrávající se v prostoru v podrobnějším měřítku, přičemž se pozornost soustřeďuje především na identifikaci rozdílů v prostoru spíše než na hledání generalizovaných podobností. Společně s metodami lokální analýzy, které v současnosti nabývají na významu, patří k pozitivům GWR možnost snadné identifikace odchylek a výjimek ve zkoumaných vztazích, zahrnutí prostorového aspektu do analýzy a především možnost zviditelnění výsledků analýz v mapové podobě. Zdůraznit je nutné také explorační povahu metody GWR, jejíž závěry mají být spíše námětem pro další hlubší analýzu. Důležité tedy není samotné zamítnutí či potvrzení testované hypotézy, ale získání nového pohledu na zkoumaná data, který nám může odhalit nečekané skutečnosti. Do oblastí, v nichž byl identifikován určitým způsobem

překvapivý vztah mezi proměnnými, tak může být zaměřen kvalitativní výzkum atd. Nejdůležitější charakteristikou metody GWR, která ovlivňuje možnost jejího širokého využití, je zajištění její prostorové povahy. Využití spojitě vážící funkce k definování velikosti vlivu prostorově blízkých jednotek je relativně jednoduchý způsob, jak docílit kalibraci regresního modelu pro jakýkoliv bod v prostoru. Zdůvodnění tohoto postupu přitom odpovídá předpokladu, že jevy v blízkých jednotkách se ovlivňují více než ve vzdálenějších.

Kromě přesnějšího pohledu na celkový charakter neboli směr a sílu zkoumaného vztahu lze metodu GWR použít ještě k dalším účelům. Porovnat lze například stacionaritu několika různých faktorů ovlivňujících vysvětlovanou proměnnou, vhodnost modelu pro určitá území atd. Při kalibraci modelu GWR je také řešena velmi přínosná otázka vhodné měřítkové úrovně analýzy (Spurná 2006). Důležitým přínosem metody GWR je skutečnost, že umožňuje vymezit oblasti s podobným charakterem zkoumaného vztahu, což může být podkladem pro provádění dalších analýz přímo v určených regionech. Výhodou využití metody GWR pro účely definování relativně homogenních oblastí z hlediska zkoumaného vztahu je například ve srovnání se shlukovou analýzou, která se v geografii k podobným účelům velmi často využívá, zejména její větší statistická srozumitelnost, a tedy i snazší interpretovatelnost, a vymezení územně souvislých regionů.

Přestože byla metoda GWR vyvinuta v nedávné době, a lze ji tak považovat za relativně novou techniku, v odborné literatuře se již objevilo množství jejích empirických aplikací (Brunsdon a kol. 1998; Fotheringham a kol. 2001; Huang, Leung 2002; Kyratso, Yiorgos 2004; Lloyd, Shuttleworth 2005; Mittal a kol. 2004). Jako důvod pro zvolení metody GWR za nástroj analýzy uvádí většina autorů očekávanou prostorovou nestacionaritu ve zkoumaných vztazích, která je téměř bez výjimky potvrzena. Právě prostorová nestacionarita je základním bodem, který musí být při používání statistických metod v geografickém výzkumu zohledněn. Skutečnost, že se charakter určitých jevů a procesů, a tedy i vzájemných vztahů mezi nimi, odlišuje v prostoru, je nepochybnitelná a z pohledu geografa v podstatě klíčová. Ačkoliv byla metoda GWR primárně vyvinuta pro aplikaci v sociální geografii, protože právě při analýze socioekonomických dat lze očekávat geografické změny ve zkoumaných vztazích, s jejím využitím se lze setkat i ve fyzické geografii (Atkinson a kol. 2003; Zhang, Shi 2004).

V rámci kvantitativní geografie existuje v současnosti spousta otázek a výzev, kterým by měla být věnována pozornost. Kromě vývoje nových metod lokální analýzy, které budou přesněji reflektovat vztahy odehrávající se v prostoru, je nutné zmínit především samotné zlepšování již používaných metod. Jako příklad lze uvést rozšíření konceptu geografického vážení na vícerozměrnou diskriminační analýzu (Brunsdon a kol. 2007). S tímto tématem souvisí v neposlední řadě potřeba větší osvěty a zvýšení obecného povědomí o současných trendech a možnostech kvantitativní analýzy v řadách geografů. Potřebná je i širší aplikace nově vznikajících metod umožňující hlubší porozumění jejich možnostem a omezením, která jistě také přispěje k rostoucímu zájmu o tyto metody a může být inspirací pro další výzkumy.

Literatura:

ANSELIN, L. (1988): Spatial econometrics: Methods and models. Kluwer Academic Publishers, Dordrecht, 284 s.

- ATKINSON, P. M., GERMAN, S. E., SEAR, D. A., CLARK, M. J. (2003): Exploring the relations between riverbank erosion and geomorphological controls using geographically weighted logistic regression. *Geographical Analysis*, 35, č. 1, s. 58–82.
- BLAŽEK, J., KOSTELECKÝ, T. (1991): Geografická analýza výsledků parlamentních voleb v roce 1990. *Geografie–Sborník ČGS*, 96, č. 1, s. 1–14.
- BRUNSDON, C., FOTHERINGHAM, S., CHARLTON, M. (1998): Geographically weighted regression – modelling spatial non-stationarity. *The Statistician*, 47, č. 3, s. 431–443.
- BRUNSDON, C., FOTHERINGHAM, S., CHARLTON, M. (2007): Geographically weighted diskriminant analysis. *Geographical Analysis*, 39, č. 4, s. 376–396.
- DISMAN, M. (2005): Jak se vyrábí sociologická znalost – příručka pro uživatele. Karolinum, Praha, 374 s.
- FOTHERINGHAM, A. S. (1997a): Analyzing numerical data. In: Flowerdew, R., Martin, D. (ed.): *Methods in human geography – a guide for students doing a research project*. Longman, Essex, s. 155–171.
- FOTHERINGHAM, A. S. (1997b): Trends in quantitative methods I: stressing the local. *Progress in Human Geography*, 21, č. 1, s. 88–96.
- FOTHERINGHAM, A. S. (1998): Trends in quantitative methods II: stressing the computational. *Progress in Human Geography*, 22, č. 2, s. 283–292.
- FOTHERINGHAM, A. S. (1999): Trends in quantitative methods III: stressing the visual. *Progress in Human Geography*, 23, č. 4, s. 597–606.
- FOTHERINGHAM, A. S., BRUNSDON, C., CHARLTON, M. (2000): *Quantitative geography – Perspectives on spatial data analysis*. SAGE Publications, London, 270 s.
- FOTHERINGHAM, A. S., CHARLTON, M. E., BRUNSDON, C. (2001): Spatial variations in school performance: a local analysis using geographically weighted regression. *Geographical & Environmental Modelling*, 5, č. 1, s. 43–66.
- FOTHERINGHAM, A. S., BRUNSDON, C., CHARLTON, M. (2002): Geographically weighted regression – the analysis of spatially varying relationships. John Wiley & Sons, London, 269 s.
- HENDL, J. (2004): Přehled statistických metod zpracování dat – analýza a metaanalýza dat. Portál, Praha, 584 s.
- HOLT-JENSEN, A. (1999): *Geography – history and concepts*. SAGE Publications, London, 228 s.
- HUANG, Y., LEUNG, Y. (2002): Analysing regional industrialisation in Jiangsu province using geographically weighted regression. *Journal of Geographical Systems*, 4, č. 2, s. 233–249.
- JONES, K. (1997): Multilevel approaches to modelling contextuality: From nuisance to substance in the analysis of voting behavior. In: Westert, G. P., Verhoeff, R. N. (ed.): *Places and people: multilevel modelling in geographical research*. Nederlandse Geographical Studies 227. Urban research centre Utrecht, Utrecht, s. 19–43.
- JONES, K., JOHNSTON, R. J., PATTIE, C. J. (1992): *People, Places and Regions: Exploring the Use of Multi-Level Modelling in the Analysis of Electoral Data*. *British Journal of Political Science*, 22, č. 3, s. 343–380.
- KOSTELECKÝ, T. (1992): Geografické analýzy volebních výsledků jako součást politické geografie. Kandidátská disertační práce. Katedra sociální geografie a regionálního rozvoje PřF UK, Praha, 180 s.
- KOSTELECKÝ, T. (2001): Vzestup nebo pád politického regionalismu? Změny na politické mapě v letech 1992 až 1998 – srovnání České a Slovenské republiky. *Edice Sociological Papers SP 01:9. Sociologický ústav AV ČR*, Praha, 96 s.
- KOSTELECKÝ, T., ČERMÁK, D. (2004): Vliv teritoriálně specifických faktorů na formování politických orientací voličů. *Sociologický časopis*, 40, č. 4, s. 469–487.
- KOUBA, K. (2007): *Prostorová analýza českého stranického systému. Institucionalizace a prostorové režimy*. *Sociologický časopis*, 43, č. 5, s. 1017–1037.
- KYRATSO, M., YIORGOS, P. (2004): Defining a geographically weighted regression model of urban evolution. Application to the city of Volos, Greece. 44th European Congress of the European Regional Science Association: Regions and Fiscal Federalism, University of Porto, 25.–29.8.2004, <http://www.ersa.org/ersaconfs/ersa04/PDF/507.pdf>, 16.12.2005.
- LLOYD, C., SHUTTLEWORTH, I. (2005): Analysing commuting using local regression techniques: scale, sensitivity, and geographical patterning. *Environment and Planning A*, 37, č. 1, s. 81–103.
- MITTAL, V., KAMAKURA, W. A., GOVIND, R. (2004): Geographic patterns in customer service and satisfaction: An empirical investigation. *Journal of Marketing*, 68, č. 3, s. 48–62.

- POON, J. P. H. (2004): Quantitative methods: past and present. *Progress in Human Geography*, 28, č. 6, s. 807–814.
- ROBINSON, G. M. (1998): *Methods and techniques in human geography*. John Wiley & Sons, London, 556 s.
- SPURNÁ, P. (2006): Současné trendy v kvantitativní analýze geografických dat se zaměřením na využití metody geograficky vážené regrese. Diplomová práce. Katedra sociální geografie a regionálního rozvoje PřF UK, Praha, 150 s.
- UNWIN, A., UNWIN, D. (1998): Exploratory spatial data analysis with local statistics. *The Statistician*, 47, č. 3, s. 415–421.
- Výsledky Sčítání lidu, domů a bytů 2001 (SLDB 2001). ČSÚ, Praha.
- Výsledky voleb do Poslanecké sněmovny Parlamentu České republiky v roce 2002 (Volby 2002). ČSÚ, Praha.
- ZHANG, L., SHI, H. (2004): Local modeling of tree growth by geographically weighted regression. *Forest Science*, 50, č. 2, s. 225–244.

S u m m a r y

GEOGRAPHICALLY WEIGHTED REGRESSION: METHOD FOR ANALYSING SPATIAL NON-STATIONARITY OF GEOGRAPHICAL PHENOMENON

In quantitative geography, there have been a large number of major changes within the past decade. These changes often do not necessarily represent the development of new techniques, but they rather reflect general methodological changes, i.e. how quantitative methods are approached and their results interpreted. A much greater emphasis is given to the development of new techniques for spatial analysis which do not conform to the traditional statistical ways of thinking. For example, emphasis is put on exploratory spatial data analysis, GIS-based forms of spatial analysis, methods of local analysis, visualizing data and model outputs. It is focused primarily on spatial data analysis as a part of geographical research, because spatial data have unique qualities which make the use of methods borrowed from non-spatial disciplines highly questionable.

The article deals with one of the new quantitative methods used in geography - geographically weighted regression (GWR). GWR is a method of analysing spatially varying relationships which completely corresponds to the contemporary development in quantitative analysis used in geography. GWR is described first by reviewing the basic methodological framework including some of the related issues, and second by presenting an empirical example of application.

GWR is based on the premise that relationships between variables might not be constant across the study area. In traditional regression analysis, it is assumed that the modelled relationship exists everywhere in the study area – i.e. the regression parameters are fixed over space. In many situations this is not necessarily the case, as mapping of residuals may reveal. This phenomenon called spatial non-stationarity is one of the major geographical analytical issues and GWR is a technique for exploring it. The advantage of GWR is that it is based on the traditional and well-known regression framework. Estimation of parameters is done through implementation of a geographical weighting scheme that is organised in a way that observation nearer the point in space where the parameter estimates is given has a heavier weight in the model than data further away. The resulting regression parameter estimates may be mapped in order to examine local variations and tested for significant spatial non-stationarity.

The empirical quantitative analyses in this article are based on aggregate statistical data referring to voting behaviour in Czechia in the parliamentary election in 2002. First, global multiple linear regression is applied to explore relationships between electoral results of two political parties (ODS and KSCM) and other selected socio-economic variables at municipality level. Second, GWR model is employed for the same input variables in view to proceed to comparative analysis. Although global regression models show some interesting and significant results, they have a shortcoming that they can mask geographical variations in studied relationships. GWR as a method of local analysis explores inaccuracy of global regression models. Empirical examples highlight the relevance and usefulness of GWR and show how it can improve our understanding of geographical processes.

- Fig. 1 – A spatial kernel. Cross – regression point, point – data point, w_{ij} – weight function, d_{ij} – distance. Source: scheme of the author according to Fotheringham et al. (2002). Note: The regression model is calibrated after geographically weighted data points (as municipalities) situated in the territory delimited around the regression point.
- Fig. 2 – Fixed and adaptive spatial kernels. Cross – regression point, point – data point. Source: scheme of the author according to Fotheringham et al. (2002).
- Fig. 3 – GWR model of electoral result of ODS in the parliamentary election in 2002. Above on the left: entrepreneurs, above on the right: believers, below on the left: university graduates, below on the right: unemployed. In keys: regression coefficient. Source: Computation of the author done with GWR 3.0; Czech Statistical Office.
- Fig. 4 – GWR model of electoral result of KSČM in the parliamentary election in 2002. Above on the left: unemployed, above on the right: believers, below on the left: university graduates, below on the right: entrepreneurs. In keys: regression coefficient. Source: Computation of the author done with GWR 3.0; Czech Statistical Office.

Autorka je postgraduální studentkou katedry sociální geografie a regionálního rozvoje Přírodovědecké fakulty UK, Albertov 6, 128 43 Praha 2; e-mail:spurna@natur.cuni.cz.

Do redakce došlo 23. 1. 2008