

Oponentní posudek  
na doktorskou disertační práci **Mgr. Pavlína Netrdové** s názvem  
**Současné trendy v kvantitativní analýze geografických dat. Možnosti a  
problémy prostorové analýzy.**

### 1. Aktuálnost tématu disertace

Téma disertační práce považuji za vysoce aktuální, protože se zaměřuje na důležité metodické aspekty aplikace kvantitativních metod v geografii a dokumentuje aplikace vybraných prostorových analýz při studiu situace v České republice. Vybrané metody nejsou dosud běžně používány a konkrétní příklady aplikace pro naše podmínky jsou proto velmi cenné. Ukazují potenciál těchto metod, formu jejich aplikace a nabízejí dalším výzkumným pracovníkům i analytickým pracovníkům z praxe vhodný návod.

### 2. Splnění cílů disertace

Cíle disertační práce jsou uvedeny na str. 8 a 9 v úvodu. Prvním cílem je diskutovat a zhodnotit koncepční otázky spojené s problémy a možnostmi aplikace kvantitativních metod, speciálně prostorových analýz, v geografickém výzkumu. Druhým cílem je na příkladech vybraných novějších kvantitativních metod prostorové analýzy doložit význam změn pro kvantitativní geografii a přínos pro geografický výzkum.

Při diskusi k roli kvantitativní geografie, resp. kvantitativních metod uplatňovaných v geografii, je třeba sledovat v první řadě cíle geografického studia a nezaměřovat nástroje, metody a cíle. Proto zcela souhlasím s vyjádřením autorky, že důležitá je především hypotéza, koncept, který je následně verifikován vhodně volenými metodami. Techniky data miningu mají také svůj nepopiratelný význam, nicméně větší roli by měl hrát první uvedený přístup, který snad můžeme označit jako shora-dolů přístup, zatímco jako zdola-nahoru přístup je možné označit data mining a metody vycházející z explorační analýzy dat (EDA, resp. ESDA).

Autorka diskutuje nomotetický a idiografický přístup v geografii, zdůrazňuje současné možnosti pro překonání antagonistického postavení obou přístupů a při využití GIS propojit oba pohledy. Otázka vztahu prostorových analýz a kvantitativní geografie by si zasloužila větší prostor pro diskusi; bohužel přístupy jednotlivých autorů jsou silně ovlivněny jejich zaměřením a prostředím. Autorka se zabývá rolí geografických informačních systémů a GI-science, přístupy prostorové ekonometrie a geostatistiky. Podrobněji se věnuje aspektům využití inferenčních statistických postupů v socioekonomické geografii a rovněž problémům analýzy agregovaných dat (zejména z pohledu MAUP a ekologické chyby). Dalšími diskutovanými tématy jsou prostorová autokorelace a prostorová nestacionarita. Z hlediska současných trendů autorka vyzdvihuje význam GIS, explorační prostorové analýzy dat a lokálních metod prostorové analýzy. V metodické části je pozornost věnována vybraným pokročilým metodám, které byly uplatněny v prezentovaných článcích. Jsou jimi geografická vážená regrese, prostorová autokorelace pro územní jednotky a Theilův index pro sledování regionálních rozdílů. V řadě článků autorka popisuje a aplikuje vybrané metody a rovněž porovnává jejich přínos s klasickými metodami, což je z hlediska praktického využití i z hlediska splnění cíle disertace důležité.

K některým částem jsou níže uvedeny poznámky a dotazy.

Je zřejmé, že se cíle disertace podařilo splnit.

### 3. Připomínky a dotazy

Nebudu se zabývat strukturou práce, která odpovídá tomuto typu prací, ale soustředím se na poznámky a otázky, které by měla doktorandka zodpovědět v rámci diskuse.

Poznámky:

- a. Celkově bych v 1.a 2.části práce uvítal větší přehled dostupných metod a ještě větší skloubení a propojení jednotlivých použitých metod.
- b. Srovnávání fyzické a socioekonomické geografie je vděčným tématem pro diskusi. K dalším faktorům, které přispívají k odlišnostem uplatňovaných metod a postupů, by bylo možné zařadit roli času, význam kontinuální a diskrétní reprezentace, míru standardizace vstupních dat či racionalitu modelů. Naopak nevnímám významný rozdíl mezi fyzickými procesy a sociálními procesy z hlediska existence či neexistence homogenity a stacionarity (P1, s. 127). Spíše se to v daném kontextu nepoužívá, ale to neznamená, že problém neexistuje.
- c. Autorka uvádí, že kvantitativní statistické hodnocení je v geografii tradičně spojováno s opakovatelností a hledáním pravidelnosti v geografických procesech. Realita je ale přece jen širší; k tradičním (ve smyslu dlouho používaným) metodám patří i třeba gravitační analýza vazeb mezi objekty a zde je kladen důraz právě na hledání odchylek z celkového modelu (např. které dojížděkové proudy mezi obcemi se vymykají z běžných poměrů) a tyto odchylky jsou významnější než popis vlastního modelu.
- d. Je otázkou, zda je zvýšená dostupnost SW (a GISu) něčím specifickým pro prostorové analýzy nebo jde o obecný fenomén. Pozor na správné rozlišování pojmu GIS a SW. GIS běžně označuje geografický informační systém, tedy jistou specifickou variantu informačního systému a ten není zaměnitelný se SW. Podobně na str. 14 bych nechápal GIS jako prostředí (tedy něco ve smyslu ovlivňujícího pozadí), ale stále jako IS.
- e. Na několika místech se opakuje rozbor a diskuse o povaze socioekonomického kvantitativního výzkumu, idiografickém a nomotetickém přístupu. Např. na s.13 se opakují některé věci z úvodu. Na str.33 se opakuje myšlenka „vzhledem k velkému citačními ohlasu .. běžně používané metody v geografickém výzkumu“.
- f. S. 14 Vyjádření, že prostorové analýzy z hlediska uplatnění v GIS jsou „analýzy vhodné pro manipulaci s prostorovými daty“ je třeba chápat s tou výhradou, že ne každá manipulace je analýzou. Rovněž i k některým dalším převzatým klasifikacím je třeba přistupovat kriticky – např. těžko hledám vysvětlení pro striktní oddělení explorační analýzy a prostorové statistiky (prostorové statistické analýzy).
- g. S. 15 Nevnímám tak silně jako autorka závislost prostorových analýz na vývoji GIS, nebo alespoň ne na vývoji geografických informačních systémů, ale na vývoji „GI science“, geoinformatiky, případně geocomputingu.
- h. S. 16 Geostatistika se primárně uplatňovala a uplatňuje v geologii a v hornictví, teprve následně v environmentálních vědách. Nemyslím, že by geostatistické postupy nebylo možné uplatnit v socioekonomické geografii. Naopak, řada metod a postupů by mohla být pro kvantitativní geografii inspirující.
- i. S. 18 Problematika náhodného výběru je jistě důležitá pro realizaci metod zjišťujících statistickou inferenci, ale jsou i jiné významné koncepční otázky, spojené např. s volbou nulové hypotézy (vůči čemu porovnávat a testovat?) či naší schopností identifikovat ze zkoumání výsledků příčinné procesy. Realizace náhodného výběru a ověřování modelu nepovažují za zásadní rozdíl mezi fyzicko-geografickým a socio-ekonomickým přístupem ke studiu jevů. Obecně validace modelu za použití nezávislých dat (tj. dat nepoužitých při tvorbě modelu) je velmi žádoucí, ať již je to model přírodního objektu nebo socioekonomického jevu.
- j. S. 23 Domnívám se, že Simpsonův paradox souvisí především s ekologickou chybou, resp. cross-level fallacy, více než s nestacionaritou.
- k. S. 32 GWR je pouze 1 z metod geografického rozšíření regresních modelů. Bylo by vhodné uvést i další modely, které rozšiřují regresní modely o autokorelační prvky (zejména z prostorové ekonometrie).
- l. S. 35 Autorka naznačuje, že geostatistika se vyvinula nezávisle na autokorelaci, ale to tak nelze chápat. Koncept autokorelace je zcela zásadním v teorii regionalizované proměnné a v jiných principech geostatistiky.
- m. S. 38 Nesouhlasím, že je LISA lokálním ekvivalentem Moranova I kritéria. Toto kritérium se sice často v rámci LISA používá, ale není důvodu, proč by měla pracovat LISA jen s touto metodou.
- n. P2, s. 779 Není jasné, proč se použilo sousedství typu věž a královna až do 3.řádu. Je evidentní, že mezi věží a královnou je v případě nepravidelných polygonů téměř nulový rozdíl. Má to smysl rozlišovat jen v případě pravidelné mřížky.
- o. P3, s. 247 Myslím si, že není velkou výhodou, že Česko má tak velký počet obcí na tak málo obyvatel. Část obcí má jen malou datovou základnu a je příliš variabilní a náchylná na výrazné změny při malém ovlivnění. Není to pak příliš stabilní a věrohodný zdroj informací z pohledu statistického zpracování. V podstatě se ze statistického hlediska jedná o skrytou heteroskedasticitu.
- p. P3, s. 257 V interpretaci je podle mne skryt problém. Autorka tvrdí, že celá struktura Česka je spíše nodálního charakteru než že by sledovala nějaké rozvojové osy. Jenže tento závěr je zřejmě ovlivněn použitou metodou – vážení sousedství musí nutně vést ke zdůraznění spíše koncentrických struktur a potlačení liniových vazeb (pokud existují).

- q. P4, s. 351 Autorka demonstruje nízkou variabilitu ve 2 kartogramech, nicméně je potřebné upozornit na možnosti jiného třídního dělení, které by poskytly pestřejší výsledek (kvantilová metoda, metoda přirozených zlomů apod.).

Dotazy:

1. Současný vývoj na poli kvantitativní geografie není spojován pouze s lokálními analýzami, jakkoliv tvoří významnou část vývoje. Mohla by se doktorandka pokusit uvést některé další významné směry?
2. Jak doktorandka chápe heterogenitu dat?
3. Jak doktorandka posuzuje možnost použít regiony vymezené pomocí LISA pro Theilův index a prohloubit tak naše znalosti o homogenitě regionů vymezených pomocí LISA a odlišnostem mezi nimi?
4. Proč se doktorandka nepokusila aplikovat uvedené metody (LISA, Theil atd.) na jiné měřítko (než obec a kraje) v souvislosti s diskusí o ecological fallacy?
5. Hodnotila doktorandka Moranovo I kritérium jiného než 1. řádu (resp. až 1.-3.řádu) a pokud ne, proč?

#### **4. Původní přínos pro další rozvoj vědy a praxi**

Hlavní přínos práce vidím především v diskusí k vybraným metodickým aspektům zpracování socioekonomických dat jako je používání agregátních dat, problém ekologické chyby, hodnocení prostorové autokorelace a zkoumání prostorové nestacionarity. Zde se ukazuje dosud chybějící obecnější systémový pohled na analýzy prostorových jevů. Je výborné, že autorka svým dílem přispívá k budování tohoto potřebného systémového pojetí.

Z hlediska praxe je velmi vítaná aplikace vybraných metod pro konkrétní příklady na území Česka. Poskytuje konkrétní návody a ukázky, jak správně postupovat při řešení vybraných prostorových analýz.

#### **5. Formální úprava disertační práce a jazyková úroveň**

Práce po formální stránce splňuje požadavky kladené na tento typ prací, našel jsem pouze asi 3 překlepy (16 aplikovanými věda, 22 studiích), tedy je napsána velmi pečlivě. Stylisticky je práce velice zdařilá a čtivá.

Několik připomínek mám pouze k anglickým překladům použitým v příloze 1 a 4: P1, s.130 quartic znamená kvadratický, nikoliv kvartický; P4, s. 350 „cartogram“ není dobrým překladem českého kartogramu, lépe použít „choropleth map“; P4, 356 „ecological mistake“ nahradit „ecological fallacy“.

#### **6. Závěr**

Závěrem lze konstatovat, že předložená disertační práce jednoznačně dokládá schopnosti autorky k samostatné tvůrčí činnosti, splňuje všechny požadavky kladené na disertační práci a proto ji **doporučuji k obhajobě.**

Navrhuji, aby po úspěšném obhájení byl podán návrh na udělení vědecko-akademické hodnosti Ph.D. Mgr. Pavlíně Netrdové.

V Ostravě 18.8.2010

Doc. Dr. Ing. Jiří Horák