

Posudek na disertační práci Mgr. Lukáše MARKA
„Prostorové a vícerozměrné statistické analýzy epidemiologických dat“
Univerzita Palackého v Olomouci, Přírodovědecká fakulta,
Katedra geoinformatiky
Školitel: doc. Mgr. Jiří Dvorský, Ph.D.

Disertační práce Mgr. Lukáše Marka „Prostorové a vícerozměrné statistické analýzy epidemiologických dat“ představuje v ucelené formě nové kartografické možnosti a nástroje vhodné k identifikaci a kvantifikaci prostorových vzorů a struktur epidemiologických dat. Pojícím konceptem práce je prostorová analýza infekčního onemocnění - kampylobakteriózy v různých měřítkových rovinách Česka.

Práce se skládá z teoretického zarámování, které je přehledně členěno do pěti hlavních kapitol shodujícími se s vytčenými dílčími cíli práce (v práci značeno jako DC1-DC5 a názorně představeno na obr. 1). V úvodní části práce autor představuje standardní metody prostorové analýzy a přínos moderních informačních technologií v mezioborovém výzkumu a definuje pět hlavních cílů své disertační práce. Mapováním a vizualizací výsledků analýz se zabývaly první a pátý dílčí cíl. Třetím hlavním tématem byla analýza možných vztahů mezi výskytem onemocnění a vnějšími faktory prostředí. V následné části jsou obce Česka klasifikovány do skupin na základě relativního rizika ohrožení nemocí a převládajících socioekonomických, demografických a environmentálních vlastností území. V poslední části je pomocí metody geografického profilování retrospektivně zkoumán možný vliv přítomnosti automatů na prodej čerstvého mléka na lokální zvýšení počtu případů řešeného infekčního onemocnění.

Aktuálnost zvoleného tématu

Zájem autora předložené disertační práce o sledování prostorových aspektů výskytu infekčního onemocnění je možné hodnotit jako počín, který přispívá k šíření mezioborových poznatků. Zvolené téma je mimořádně badatelsky zajímavé a aktuální nejen z hlediska časoprostorové dynamiky šíření infekčního onemocnění, ale také z hlediska upřesňování terminologického systému a diskuse o relevantních teoreticko-metodologických problémech (geo)vizualizace v oblasti prostorové epidemiologie.

Přínos pro další rozvoj vědy

Předložené disertační práce představuje velice cenný materiál pro oblast (geo)vizualizace a užití kartografických metod ve studiu zdravotního stavu obyvatel. Poznatky zde uvedené tak zcela jistě mohou být cennou inspirací nejen pro geografický, ale i široce pojatý mezioborový výzkum. Jednoznačně souhlasím s tvrzením autora na str. 32, že: „(geo)vizualizace je důležitá také pro komunikaci výsledků a jejich předání širokému spektru možných příjemců. Těmi mohou být jak odborníci, tak i laici, což platí obzvlášť v interdisciplinárních oborech, kde spolupracují různě zaměřeni experti, takže je vždy potřeba zvolit kompromis mezi způsobem a podrobností výstupu“.

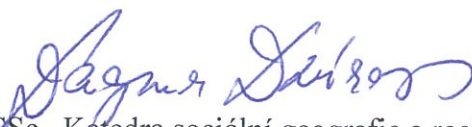
Poznámky/doporučení:

- Práce je napsána velice pěkným vědeckým jazykem, je doplněná o množství doprovodných grafických výstupů. Bohužel zde mám asi hlavní výtku k předložené práci, na úkor množství zařazených velmi zajímavých obrázků je jejich velikost pro

čtenáře nekomfortní, resp. často velmi nedostatečná zejména v oblasti legend (viz např. obr. 9, 10, 24, 25, 28, 29, 33 atd.).

- Základní data předložené práce se týkala výskytu kampylobakterií v České republice v období let 2008 - 2012. V této části postrádám diskusi k lokalizaci užitých dat, kdy se jednalo o místo nákazy a kdy o lokalizaci dle místa bydliště, autor pouze na str. 11 uvádí: „Zkontrolovaným datům byla na základě místa nakažení a/nebo místa bydliště přiřazena poloha.“ V kap. 4.2 „Základní charakteristika a analýza časových řad“ jsem očekávala jasnější přehlednou informaci o struktuře a rozložení datového souboru, který je používán v celé předložené práci, tedy nikoliv odvozování z řady obrázků, které jsou pro svoji malou velikost v předložené podobě téměř nečitelné.
- Na str. 25 je uveden přehled základních pojmů, ale není zde dostatečně popsána metoda standardizace, resp. na jakou standardní populaci byly použité míry aplikovány.
- Domnívám se, že metoda kartografické anamorfózy si v řešené problematice zasloužila větší pozornost, než jak autor vyřešil na straně 42.
- V části 6.1.1. věnované zpracování vybraných charakteristik studia vztahů mezi onemocněním a charakteristikami prostředí je pouze nastíněn konstrukt indexu socioekonomické deprivace (str. 68), což pro další využitelnost práce považuji za limitující, a to zejména v kontrastu detailních popisů vybraných analýz, např. hlavních komponent či logistické regrese (str. 81).
- V některých případech mohl být podán fundovanější komentář, např. k obr. 30, ke kterému autor na str. 89 uvádí: „Obrázek vykresluje ROC křivky pro obě prostorové metody, tedy lokální ordinální regresi a GW DA, v obou případech jsou výrazně nejlépe klasifikovány obce bez výskytu kampylobakterií.“ Možná v diskusi by oponentka uvítala vysvětlení, které jí z obrázku nevyplývá (byť zřejmě na základě její neznalosti).
- Za velmi zajímavý považuji boxplot na str. 98, leč opět pro jeho velikost je získaná hodnota pro čtenáře silně limitující, škoda.
- V poslední části nazvané „Geografické profilování“ bylo vymezeno 18 různorodých oblastí. Autor na str. 103 uvádí: „Z důvodu výpočetně a časově náročného procesu byla data rozdělena do 18 nestejnorodých oblastí“, popis procesu jsem v práci nenašla.

Závěr: Na základě komplexního hodnocení předložené disertační práce Mgr. Lukáše Marka lze konstatovat, že práce je cennou, ucelenou studií poměrně neprobádané oblasti kartograficko-epidemiologického výzkumu. K předložené práci a její obsahové a formální stránce nemám žádné zásadní výhrady, práce je zpracována přehledně a pečlivě, neobsahuje překlepy ani gramatické chyby. Výsledky práce potvrzují velmi dobrou schopnost autora zvládnout moderní teoreticko-metodologické postupy vědecké práce. Vzhledem ke skutečnosti, že autor předložené práce nastudoval množství odborné, zejména zahraniční literatury, prezentoval významné nové teorie a koncepty prostorové analýzy, přehledně je zpracoval a úspěšně aplikoval ve svém výzkumu, doporučuji **disertační práce Mgr. Lukáše Marka přijmout k obhajobě.**



24. května 2015

Doc. RNDr. Dagmar Džurová, CSc., Katedra sociální geografie a regionálního rozvoje,
Přírodovědecká fakulta, Univerzita Karlova v Praze

prof. Mgr. Jaroslav Hofierka, PhD.

Ústav geografie, Prírodovedecká fakulta, Univerzita Pavla Jozefa Šafárika v Košiciach

Oponentský posudok

doktorskej dizertačnej práce

„Prostorové a vícerozměrné statistické analýzy epidemiologických dat“

Mgr. Lukáša Marka

doktoranda Katedry geoinformatiky, Prírodovedeckej fakulty Univerzity Palackého v
Olomouci

1. Zhodnocení významu disertace pro příslušný obor

Predložená dizertačná práca sa venuje aktuálnej, no veľmi náročnej téme komplexnej geopriestorovej analýzy rozšírenia kampylobakteriôzy v Českej republike. Problematika má značne interdisciplinárny charakter, vyžaduje poznatky nielen z geoinformatiky, ale aj epidemiológie, medicínskej geografie a matematickej štatistiky. Práca ponúka nový pohľad na problematiku rozšírenia tohto ochorenia s využitím celej plejády štatistických metód a geoinformatických nástrojov. Má nielen odborný, ale aj praktický význam a pre rozvoj tejto problematiky je jednoznačným prínosom.

2. Vyjádření k postupu řešení problému, použitým metodám a splnění stanoveného cíle

Autor si stanovil jeden hlavný a 5 čiastkových cieľov. Základným cieľom bola komplexná priestorová analýza epidemiologických dát pre obdobie rokov 2008-2012 získaných z epidemiologickej databázy EPIDAT pomocou viacerých geoinformatických a štatistických metód. Čiastkové ciele sa zameriavali na rôzne aspekty tohto hlavného cieľa, od všeobecného popisu priestorového rozšírenia ochorenia na území Českej republiky až po identifikáciu možných faktorov vplývajúcich na výskyt ochorenia a vhodnej časopriestorovej vizualizácie tohto javu.

Na splnenie týchto úloh autor použil široké spektrum metód a softvérových nástrojov. Nosnými boli geoinformatické a kartografické metódy, ktoré boli použité najmä na vyjadrenie priestorového aspektu javu a pri analýze výsledkov. Bolo použité celé spektrum metód priestorovej štatistiky, ktoré boli úplne vyčerpávajúcym spôsobom aplikované na odhalenie väzieb a priestorových súvislostí. Na tento účel autor použil GIS softvér ako ArcGIS, QGIS, GoogleEarth Pro a štatistické nástroje ako R, GeoDA a SaTScan.

Splnenie cieľov autor realizoval veľmi robustným a kvalifikovaným spôsobom. Rozsiahlu databázu takmer 100 tis. záznamov spracoval do geopriestorovej podoby pomocou rozhrania API Mapy.cz, čo bolo základným predpokladom pre všetky ďalšie geopriestorové analýzy. Následne dôkladne analyzoval priestorovú distribúciu výskytu ochorenia, pričom využil množstvo náročných metód a indikátorov, ktorú mu umožnili identifikovať oblasti, kde je zvýšený výskyt ochorenia. Najnáročnejšou časťou bolo hľadanie vzťahov medzi výskytom ochorenia a faktormi, ktoré môžu prispievať k výskytu ochorenia. Tu vychádzal z viac než sto možných premenných, od environmentálnych až po socioekonomické faktory. Dôkladnou analýzou a aj aplikáciou metódy hlavných komponentov zredukoval počet možných faktorov na 11, resp. 5. Tieto premenné vysvetlili výskyt ochorenia na viac ako 50%, čo zodpovedá výsledkom iných, zahraničných autorov. Vo finálnej časti autor využil nástroje softvéru R a

GoogleEarth Pro na časopriestorovú vizualizáciu javu, keďže skúmaný jav má aj zaujímavú dynamiku. Môžem konštatovať, že autor inovatívnym a vyčerpávajúcim spôsobom aplikoval použité metódy a stanovené ciele splnil.

3. Stanovisko k výsledkům disertační práce a k původnímu konkrétnímu přínosu předkladatele disertační práce

Dizertačná práca priniesla viacero zaujímavých výsledkov. V prvom rade ukázala na veľmi široké možnosti využitia geopriestorových metód pri analýze epidemiologických ochorení. Autor predstavil jasnú a ucelenú metodiku na spracovanie takýchto dát a predstavil množstvo metód a nástrojov, ktoré je možné na tento účel použiť. Veľmi zaujímavé je aj kartografické vyjadrenie výskytu ochorení z priestorového a časového hľadiska, ktoré je nepochybne pre epidemiológiu veľmi inšpiratívne. Z hľadiska geoinformatiky považujem za veľmi cennú priestorovú aplikáciu viacerých sofistikovaných štatistických metód či už na globálnej alebo lokálnej úrovni. Za inšpiratívne považujem aj netradičné využitie GoogleEarthPro na vizualizáciu časopriestorových aspektov skúmaného javu. V empirickej rovine autor do značnej miery potvrdil zistenia iných výskumníkov. Identifikoval niekoľko regiónov so zvýšeným výskytom ochorenia a v týchto regiónoch analyzoval jeho možné príčiny.

V práci sa však objavilo aj niekoľko nedostatkov formálneho charakteru, ktoré však zásadne neovplyvňujú celkový prínos práce. Napríklad niektoré obrázky sú horšie čitateľné (napr. Obr. 32). Z obsahového hľadiska sa mi zdá, že autor až príliš zjednodušene analyzoval niektoré premenné prostredia, ktoré môžu vplývať na výskyt ochorenia. Konštatuje, že hydínové (drubčie) mäso býva najčastejším zdrojom nákazy (s. 68), avšak neanalyzuje podrobnejšie prevádzky s chovom hydiny a ani miesta porážky a spracovania tohto mäsa. Počet hydiny a aj iných chovných zvierat bol odhadnutý na základe počtu ekonomických subjektov so živočíšnou výrobou a cez vážený rozklad boli počty zvierat z úrovne okresov rozdelené na úroveň obcí. Voči takto odvodeným premenným boli realizované ďalšie analýzy. Je tu otázka, či je postačujúce dávať do vzťahu výskyt ochorenia a prevádzky chovu zvierat na lokálnej úrovni, alebo by bolo vhodnejšie sledovať priestorový pohyb tohto mäsa a možné efekty na výskyt ochorenia aj mimo lokality prevádzky. Autor sa môže k týmto aspektom vyjadriť počas obhajoby.

4. Případné další vyjádření, zejména vyjádření k uspořádanosti, přehlednosti, formální úpravě a jazykové úrovni disertační práce

Práca má primeranú formálnu úroveň, je písaná zrozumiteľne a s vysokou odbornosťou. Oceňujem kvalitu mnohých kartografických výstupov.

5. Vyjádření k publikacím studenta doktorského studia

Publikačná aktivita dizertanta je na nadpriemernej úrovni. Je autorom alebo spoluautorom niekoľko desiatok prác, z toho 1 práca je časopisecká publikácia evidovaná v databáze WoS a ktorá priamo súvisí s témou práce, okrem toho je spoluautorom ďalších dvoch príspevkov v impaktovaných časopisoch a jednej kapitoly v monografii vydanej vo vydavateľstve Springer. Celkovo prevažujú konferenčné príspevky, avšak viaceré z nich sú evidované v databáze WoS alebo Scopus.

6. Jednoznačné vyjádření, zda předloženou disertační práci doporučuje či nedoporučuje k obhajobě

Autor si v predloženej dizertačnej práci stanovil viacero úloh, ktoré vhodnou aplikáciou metodických postupov splnil. Konkrétny prínos práce je v oblasti metodického postupu pri

časopriestorovej analýze epidemiologických dát, aplikácii sofistikovaných metód priestorovej štatistike a metodologickej analýze faktorov prostredia spôsobujúcich výskyt ochorenia. Práca má jednoznačný metodologický a aj praktický význam. Dôsledná aplikácia objektívnych štatistických metód umožňuje selektovať kľúčové faktory vplývajúce na výskyt ochorenia a teda prijímať kvalifikované epidemiologické a preventívne opatrenia. Má preto aj vysoký potenciál uplatnenia v praxi. Na základe vyššie uvedeného hodnotenia dizertačnej práce konštatujem, že predložená dizertačná práca spĺňa podmienky kladené na tento typ prác a preto doporučujem dizertačnú prácu Mgr. Lukáša Marka prijať k obhajobe.

V Košiciach, 21.5.2015



prof. Mgr. Jaroslav Hofierka, PhD.

Oponentský posudek na dizertační práci

Mgr. Lukáš Marek: Prostorové a vícerozměrné statistické analýzy epidemiologických dat

Katedra geoinformatiky, Přírodovědecká fakulta Univerzity Palackého v Olomouci

Dizertační práce Mgr. Lukáše Marka je věnována problematice prostorové analýzy rozšíření kampylobakterií v České republice v letech 2008-2012. Vzhledem ke komplexnosti analýzy je práce zaměřena na řešení pěti dílčích cílů. Konkrétně se jedná o mapování a popis výskytu kampylobakterií; průzkum, kvantifikaci a vizualizaci prostorových a časoprostorových vzorů ve výskytu kampylobakterií; identifikaci a analýzu vztahů mezi výskytem onemocnění a vnějšími environmentálními, demografickými a socioekonomickými faktory; zhodnocení vlivu automatů na čerstvé mléko jako zdroje nákazy a převedení vybraných výsledků do podoby vhodné k interaktivní exploraci v prostoru a čase.

Zvolená tematika dizertační práce je vysoce aktuální. Kampylobakterií je jedno z nejčastějších infekčních onemocnění v rozvinutých zemích. Rozsáhlá mezioborová případová studie tohoto onemocnění na území České republiky má jednoznačný přínos v oblasti vizualizace časoprostorové epidemiologie. Předložená práce má též význam i z praktického hlediska při analýze faktorů způsobujících onemocnění. Svým širokým mezioborovým charakterem má práce vysoký potenciál stát se základem pro další rozvoj vědy nejen v oblasti geoinformatiky a kartografie, ale i v oblasti epidemiologie nebo matematické statistiky.

Autor si hlavní cíl práce, kterým je komplexní časoprostorová analýza epidemiologických, velmi vhodně rozdělil na pět dílčích cílů. Jejich vyřešením vznikla představa o prostorovém šíření kampylobakterií v České republice a byly odhaleny její možné podmiňující faktory. K vlastní analýze byla použita data pocházející z databáze EPI-DAT (Epidemiologická databáze) obsahující téměř 100 tisíc záznamů. Při řešení jednotlivých dílčích cílů byly využívány nejmodernější geoinformatické a kartografické metody a metody prostorové statistiky. Při zpracování byly využívány programy ArcGIS, QGIS, GoogleEarth Pro, R, GeoDA a SaTScan.

V dizertační práci byla získána řada zajímavých výsledků. Z pohledu geoinformatiky je přínosná zejména vizualizace výskytu onemocnění na úrovni obcí, identifikace regionů s vysokým a nízkým relativním rizikem onemocnění, identifikace možných faktorů, které podmiňují šíření onemocnění a v neposlední řadě aplikace široké škály statistických metod.

Po formální i odborné stránce mám k předložené práci několik připomínek, které však nemají zásadní vliv na kvalitu předložené práce:

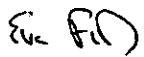
- Práce je doplněná řadou grafických výstupů, jejich velikost je však často nedostačující a obrázky jsou tak velmi špatně čitelné.
- Příprava vstupních dat by mohla být podrobněji popsána. Především zde postrádám informace o způsobu doplnění chybějících údajů, počtu neúplných záznamů, počtu zásahů do datového souboru, zda a kolik záznamů bylo zcela vynecháno, apod.

- Z práce není zřejmé, proč byla data zvolena právě v období 2008-2012. Z hlediska časového vývoje onemocnění by jistě bylo vhodnější analyzovat delší časovou řadu.
- V práci byla použita široká škála různých statistických metod. Při aplikaci jednotlivých metod často postrádám, jak byla data konkrétně zpracovávána, a tudíž, jak mají být výsledky interpretovány. Datový soubor je totiž ve formě časové řady a z uvedených výsledků a jejich interpretace často není zřejmé, zda se jedná průměrné roční hodnoty nebo kumulované hodnoty za 5 let. Z praktického hlediska je žádoucí, aby výsledky bylo možno interpretovat jako průměrné roční hodnoty.
- Proč byly zvoleny apriorní pravděpodobnosti pro Bayesovské odhady rizika onemocnění zrovna z negativně binomického rozdělení?
- Popisnou statistiku v Tabulkách 3,5,7 (str. 44-50) by bylo vhodné doplnit o boxploty nebo histogramy, aby čtenář získal lepší představu o skutečném rozložení dat.
- V korelační matici (str. 70-71) bylo identifikováno 11 významných charakteristik prostředí a 3 charakteristiky prostředí. Jaká je přesná interpretace jednotlivých faktorů? Z popisu jednotlivých charakteristik není zřejmé, co vyjadřují, jak byly spočteny, a zda je tedy jejich vzájemné srovnávání smysluplné. Pro analýzu byl používán Pearsonův a Spearmanův korelační koeficient. Popis při interpretaci výsledků je trochu nešťastný a označení koeficientů vede k dojmu, že všech 126 charakteristik bylo analyzováno pomocí Pearsonova koeficientu a až výsledných 14 charakteristik pomocí Spearmanova koeficientu. Dále je třeba upozornit, že proměnná pohlaví je nominální, a tudíž je pro ni irelevantní počítat korelaci. Pro tuto proměnnou lze užít pouze vhodné míry asociace.
- V logistické regresi ve vztazích [6.10] a [6.11] má být před sumou znaménko +.
- Odhady relativního rizika u proměnné Hustota zalidnění v multinomickém modelu v Tabulce 12 (str. 85) signalizují chybný výpočet.

Závěr

Lze konstatovat, že cíle práce byly splněny. Autor ve své práci prokázal, že je schopen zpracovat poznatky z mnoha vědních oborů (geoinformatika, statistika, epidemiologie) a podařilo se mu vytvořit ucelenou studii v oblasti časoprostorové analýzy epidemiologických dat. Práce je psána jasně a přehledně, na vysoké odborné úrovni. Typograficky je velmi pěkně upravena. **Předložená práce splňuje požadavky kladené na dizertační práci.** Vzhledem k výše uvedenému doporučuji dizertační práci Mgr. Lukáše Marka k obhajobě.

V Olomouci 2. června 2015


 doc. RNDr. Eva Fišerová, Ph.D.
 Katedra matematické analýzy a apl. mat.
 Přírodovědecká fakulta
 Univerzita Palackého v Olomouci