

METÓDY TRANSFORMÁCIE PRIESTOROVÝCH DÁT: PREHĽAD, KLASIFIKÁCIA A HODNOTENIE

Michala Madajová

Univerzita Komenského v Bratislave, Prírodovedecká fakulta, Katedra regionálnej geografie, ochrany a plánovania krajiny, e-mail: madajova@fns.uniba.sk

Abstract: This paper reviews and analyzes areal data transformation's methods which can estimate data available for a one set of geographical units but needed for another set of zones. Following the available literature the classification of different data transformation's methods is offered. The paper will be conducted as a comparative analysis of these methods using population as the variable of the interest. Of these, the methods most appropriate for our research are chosen.

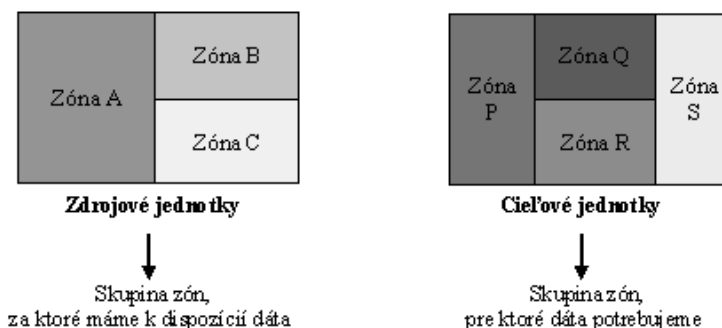
Key words: areal weighting, dasymmetric, pycnofylactic, interpolation methods, regression

1 ÚVOD

Transformácia priestorových dát umožňuje kombináciu a analýzu údajov dostupných z rôznych zdrojov a za rôzne geografické jednotky. Dáta humánno-geografickej povahy sa v snahe o zredukovanie rozsahu publikovaných údajov, ale predovšetkým v dôsledku zachovania anonymity jednotlivých subjektov publikujú za istý stupeň priestorovej agregácie. Na Slovensku sa údaje zbierajú za cenzové domácnosti a publikujú spravidla za administratívne jednotky (obce, okresy,...). Pre geografické analýzy sú však v mnohých prípadoch vhodnejšie iné zóny než tie, ktoré sú založené na politických hraniciach – napríklad fyzickogeografické regióny, funkčné mestské regióny, oblužné regióny, regióny odvodené z poštových smerových čísiel a i. Taktiež sa môže stať, že geografické hranice cenzových jednotiek sú nekompatibilné s tými, za ktoré údaje potrebujeme, prípadne s tými, za ktoré máme dáta k dispozícii. Preto, ak potrebujeme údaje za iné ako administratívne jednotky, prípadne ak chceme použiť a hlavne zjednotiť údaje z rôznych zdrojov prípadne časových období (napr. za účelom sofistikovanejších štatistických a demografických analýz), je nevyhnutné pristúpiť k harmonizácii týchto dát, a to prostredníctvom rozličných transformačných metód. Cieľom tohto príspevku je na základe dostupnej literatúry predstaviť a navzájom porovnať rôzne metódy transformácie priestorových

dát, podať ich doteraz absentujúcu ucelenú klasifikáciu a vybrať z nich tie, ktoré sú vhodné pre náš ďalší výskum.

Spracovanie demografických, ekonomických, sociálnych, či environmentálnych dát si teda môže vyžadovať ich transformáciu z jedného systému geografických jednotiek do nejakého iného (pozmeneného alebo úplne odlišného) zonálneho systému, čo je proces, ktorý sa označuje ako priestorová interpolácia. Zo známych hodnôt vybranej premennej dostupných za jeden zonálny systém je teda možné odhadnúť hodnoty premennej za inú skupinu geografických jednotiek, ktoré ale opisujú ten istý región. V tejto súvislosti má zmysel hovoriť o tých zónach, pre ktoré máme k dispozícii údaje ako o zdrojových geografických jednotkách a naopak, o tej sade zón, pre ktoré tieto údaje potrebujeme ako o cieľových geografických jednotkách (Goodchild a Lam, 1980) Pre názornosť uvádzame Obr.1.



Obrázok 1 Zdrojové a cieľové geografické jednotky

2 METÓDY TRANSFORMÁCIE PRIESTOROVÝCH DÁT V DOSTUPNEJ LITERATÚRE

Základné problémy priestorovej interpolácie načrtávajú Goodchild a Lam (1980) a Lam (1983). Podľa Goodchild a Lam (1980) sa termín priestorová interpolácia všeobecne vzťahuje na problém odhadu hodnoty premennej v nejakom bode (x,y). Je to proces, ktorý podlieha chybám v závislosti od predpokladov obsiahnutých v použitých metódach. Podľa Goodchild a Lam (1980) existuje množstvo prístupov, ktoré opisujú rôzne problémy priestorovej interpolácie dát. Na základe týchto prístupov delia interpolačné metódy do dvoch hlavných kategórií. Sú nimi tzv. obsah zachovávajúce a obsah nezachovávajúce metódy. V prípade priestorovej interpolácie sa termín „obsah zachovávajúci“ vzťahuje na zachovanie celkovej hod-

noty premennej každej zdrojovej zóny. Ak je napríklad interpolovanou premennou počet obyvateľov, potom sa musí na základe tohto kritéria v procese interpolácie zachovať suma počtu obyvateľov zdrojových zón. Obsah nezachovávajúce metódy používajú bodové údaje na odhad integrálov funkcie v určitom území. Jedná sa o tzv. bodové (alebo na bode založené) interpolačné metódy, kde sú bodmi spravidla centroidy zdrojových zón, zástupné body alebo niektoré typy vážených stredov ako napríklad centrum obyvateľstva. Metódu „bod v polygóne“ použili vo svojej práci napr. Okabe a Sadahiro (1997). Táto metóda presúva hodnoty premennej zo zdrojovej zóny do cieľovej, ak jej „reprezentatívny bod“ leží v cieľovej zóne (tamtiež). Aj keď na bodové priestorovo-interpolačné metódy aplikované na problém priestorovej interpolácie nezachovávajú obsah, existuje práca Kyriakidisa (2004), ktorý naopak uvažuje o interpolačnej technike „územie do bodu“, ktorá využíva geoštatistickú metódu krigingu, avšak na rozdiel od tradičných bodovo-interpolačných techník je táto schopná zachovať obsah interpolovanej premennej v zdrojovej zóne. Je to do budúcnosti pomerne sľubná technika, nakoľko kriging počíta aj s doplnením o pomocné informácie (Hawley, 2005).

Ako obsah zachovávajúce metódy uvádzajú Goodchild a Lam (1980) a Lam (1983) metódu prekryvu a tzv. pyknofylaktickú metódu. Metóda prekryvu sa zvyčajne označuje ako metóda priestorového váženia. Detailne sa ňou zaoberá práca Goodchild a Lam (1980). Jedná sa o jednoduchú kartografickú metódu založenú na území prekryvu zdrojových a cieľových geografických jednotiek. Hoci táto metóda uchováva obsah, predpokladá, že premenná je homogénne rozmiestnená v zdrojových jednotkách. Tento predpoklad je však pre väčšinu dát humánno geografickej povahy neprijateľný. Naopak, tzv. pyknofylaktická metóda predpokladá heterogénne rozmiestnenie premennej, čo znamená, že premenná nie je rovnomerne rozmiestnená v každej zdrojovej zóne. Jedná sa o metódu navrhnutú Toblerom (1979), ktorá sa na základe nejakej informácie o hustote zaľudnenia zdrojových zón (vo forme agregovaných populačných štatistik) pokúša odhadnúť spojitý povrch hustoty zaľudnenia v uzloch jemnej mriežky, ktorý sa potom použije na výpočet hustoty zaľudnenia cieľových jednotiek. Táto metóda berie do úvahy efekt prilahlých zdrojových zón. Pre porovnanie, metóda priestorového váženia poskytuje lepšie výsledky vtedy, ak možno skúmaný povrch považovať za nespojitý. Naopak, ak je plynulosť skutočnou vlastnosťou povrchu, vtedy je lepšie použiť pyknofylaktickú metódu. V prípadoch, keď nemožno charakter povrchu identifikovať ani ako spojitý, a ani ako nespojitý, musia sa pre spoľahlivé výsledky zaviesť rôzne vedľajšie podmienky. Takéto metódy sa ešte len rozvíjajú (Lam, 1983).

Toblerovu metódu vylepšil Rase (Rase, 2001), keď namiesto tradičnej pravouhlej mriežky použil trojuholníkovú nepravidelnú sieť (TIN). Hlavným dôvodom bola možnosť zachovania pôvodných hraníc polygónov. Pravouhlá mriežka totiž nezachováva pôvodné hranice polygónov, zatiaľ čo flexibilnejší model TIN áno, pretože je schopný použiť body geografických hraníc ako geometrické vrcholy (Hawley, 2005). Okrem toho sa tak táto metóda vyvaruje aj chýb z konvertovania zdrojových polygónov do pravidelnej mriežky.

Novšie výskumy týkajúce sa problematiky jednotlivých priestorovo-interpoláčnych metód podávajú Goodchild et al. (1993). V práci identifikujú tri hlavné typy interpolácie: kartografické, regresné a povrchové metódy. Medzi povrchové metódy zaradili už spomínané bodové priestorovo-interpoláčne metódy a pyknofylaktickú metódu. Kartografickými metódami sú naopak metóda priestorového váženía a tzv. dazymetrická metóda. Dazymetrické mapovanie navrhol ako alternatívu k metóde priestorového váženía Wright (1936). Rozvinul metódu mapovania hustoty zaľudnenia s využitím topografických listov ako pomocných informácií. Hlavným účelom tejto metódy bolo využiť vedomosti o danej lokalite a identifikovať územia v zónach, ktoré majú rozdielnu hustotu zaľudnenia. V rámci miest Wright identifikoval neobývané územia a vypočítal hustotu zaľudnenia vo zvyšných častiach mesta. Metóda teda využíva vedomosti o miestnom území na to, aby z nich mohla určiť danú premennú, v tomto prípade obyvateľstvo. Týmto spôsobom sa tak predpokladá rovnomerného rozmiestnenia premennej v zdrojovej zóne vylepší. Dazymetrickú metódu už teda možno zaradiť do skupiny metód (podľa Lam, 1983), ktoré používajú pomocné informácie o zdrojových zónach na to, aby vylepšili výsledky interpoláčného procesu. Ako pomocné informácie využívajú tieto priestorovo-interpoláčne metódy také údaje, o ktorých sa predpokladá, že odrážajú priestorové rozmiestnenie mapovaných premenných, prípadne s nimi nejakým spôsobom súvisia. V literatúre zaoberajúcej sa priestorovou interpoláciou autori pracujú predovšetkým s obyvateľstvom a ako pomocné informácie vzťahujúce sa k tejto premennej používajú hlavne údaje z land use. Je to preto, lebo v posledných rokoch sa stali pomerne široko dostupnými satelitné snímky a dáta z diaľkového prieskumu Zeme. Objavilo sa preto niekoľko spôsobov realizácií priestorovo-interpoláčnych metód s využitím dát z DPZ ako pomocných informácií, a práve najvýznamnejšou metódou z tejto kategórie je tzv. dazymetrická metóda. Ako identifikovať obývané územia zo satelitných snímok Landsat ukázali vo svojej práci Monmonier a Schnell (1984). Flowerdew a Openshaw (1987) taktiež načrtli možnosť použitia dazymetrickej metódy, no ako uvádza Fisher a Langford (1995), nikdy ju nepoužili v žiadnej zo svojich prác v rámci priestorovej interpolácie. Ako prví publikovali výsledky priestorovej interpolácie použitím dazymetrickej metódy Fisher a Langford (1995). Autori použili Openshawovu (Openshaw, 1977) simuláciu Monte Carlo pre priestorové jednotky. Tento algoritmus použili Fisher a Langford (1995) na vytvorenie cieľových zón, ktoré majú známy počet obyvateľov (je súčtom počtu obyvateľov m zón v každej cieľovej zóne). Ich metóda je variantom Wrightovej metódy v tom, že používa binárne delenie typov land use, ktoré sa vzťahuje na prítomnosť a absenciu populácie v území. Najjednoduchším prístupom v rámci dazymetrickeho mapovania s využitím dát z land use je teda prístup binárneho rozdelenia, kedy sa jednotlivé land use územia klasifikujú buď ako zaľudnené, alebo nezaľudnené. Obyvateľstvo sa tak jednoducho redistribuuje do zaľudnených land use území. Tento prístup použili aj Holt et al. (2004). Existujú však aj iné spôsoby implementácie - na základe dostupnej literatúry možno okrem spomínaných rozlíšiť aj iné varianty metód, ktoré vytvárajú dazymetrické mapovanie. Príkladom takýchto štúdií sú práce Fisher a Langford (1991), Mennis (2003), Yuan et al. (1997) a i. V týchto štúdiách je viacero možností, ako ur-

čiť vzťah medzi hustotou zaľudnenia a land use triedami. Niektoré štúdie používajú prístup empirického/pokusného výberu (Mennis, 2003), niektoré používajú preddefinované hustoty zaľudnenia (Eicher a Brewer, 2001) a niektorí regresné analýzy na určenie odhadov hustoty zaľudnenia (Yuan et al., 1997 a Fisher a Langford, 1991). Eicher a Brewer (2001) použili tri techniky dazymetrickeho mapovania pre priestorovú interpoláciu - binárnu metódu, metódu troch tried a metódu limitujúcej premennej. Metóda troch tried používa váženú schému na pridelenie populácie do mestských, poľnohospodárskych a lesných typov land use. Váhy sú subjektívne určené výskumníkom na základe jeho vedomostí o študovanom území. Metóda limitujúcej premennej pridelí populáciu podľa obmedzenia hustoty zaľudnenia, ktoré sa nastaví pre každý typ land use. Eicher a Brewer (2001) zistili, že najpresnejšie výsledky poskytuje práve metóda limitujúcej premennej. Preferovaným prístupom je však ten, ktorý používa regresné analýzy na odhad interpolovanej premennej, pretože je objektívny v dôsledku štatistického testovania presnosti modelov.

Mimo tradičných metód priestorovej interpolácie je teda potrebné rozlíšiť aj tzv. metódy štatistického modelovania. Už teórie z geografie mesta demonštrovali, že rozmiestnenie obyvateľstva v mestách je ovplyvnené rôznymi faktormi, ako sú napríklad vzdialenosť od centra alebo vzdialenosť k významným cestám. Množstvo takýchto morfológických faktorov sa postupom času dalo získať z dát z diaľkového prieskumu Zeme. Princípy štatistického modelovania na odhad počtu alebo rozmiestnenia obyvateľstva spolu s údajmi z DPZ, ktoré sa ukázali ako mimoriadne vhodné pre štúdium rozmiestnenia obyvateľstva, sa začali používať od polovice 50-tych rokov 20. st. obzvlášť s cieľom hľadania akejkoľvek alternatívy k populačnému cenzu. Motiváciou bola predovšetkým snaha o odstránenie nedostatku dát vyplývajúcich z desaťročných populačných cenzov, ale aj ich vysoká cena, či pracnosť spracovania (Wu et al., 2005). Okrem toho sa rôzne štatistické prístupy aplikovali aj na okontrolovanie spoľahlivosti cenzových výpočtov. Výskumníci zaviedli viacero metód štatistického modelovania na odhad populácie v rôznych mierkach s rôznymi typmi snímok z DPZ. Podľa Wu et al. (2005) možno vo všeobecnosti rozlíšiť 5 kategórií prístupov. Sú to prístupy založené na vzťahu medzi obyvateľstvom a mestským územím, využitím zeme, bytovými jednotkami, pixlovými charakteristikami obrázku a ostatnými fyzickými a socioekonomickými charakteristikami. Štatistickú regresnú analýzu na odhad populácie subzón v rámci zdrojových zón navrhli Flowerdew (1988) a Flowerdew a Green (1989). Ako prví však aplikovali mnohonásobné regresné techniky na odhad obyvateľstva subzón prostredníctvom dazymetrickej metódy (Langford et al., 1991). Jedná sa teda o tzv. regresnú dazymetrickú priestorovo-interpoláčnú metódu s využitím dát z land use. Nazývajú ju aj ako „inteligentnú interpoláciu“. Flowerdew (1988) rozvinul metódu priestorovej interpolácie dát, ktorá operuje ako Poissonov proces so sadou binárnych premenných. Na priestorovú interpoláciu obyvateľstva nie je podľa neho vhodný lineárny regresný model, ale naopak, má sa použiť tzv. Poissonova distribúcia chýb. V podstate tu rozvinutá metóda pracuje na základe vytvorenia regresného vzťahu medzi záujmovou premennou a jednou alebo viacerými pomocnými premennými, ktorý sa potom použije na odhad hodnôt záujmovej premennej pre cieľové zóny. Tento vzťah sa ale musí odhadnúť

vhodným spôsobom vzhľadom na predpokladané rozdelenie záujmovej premennej (Poissonovo, binomické alebo normálne rozdelenie). Parametre sa teda stanovujú regresiou na zdrojové zóny a potom sa aplikujú na cieľové zóny na základe hodnoty binárnej premennej. V prípade prítomnosti viacerých premenných v jednej cieľovej zóne sa parametre vážia na základe určitých mier, ako je napríklad priestorový rozsah danej premennej (v prípade land cover) alebo na základe nejakej inej významnej racionálnej miery. V prípade land use typov ako binárnych premenných, je táto metóda podobná tej, ktorú použili Langford et al. (1991). Flowerdew a Green (1991) túto metódu rozšírili a zahrnuli do nej časť štatistickej teórie známej ako EM algoritmus (pravdepodobnosť - Expectation - maximalizácia - Maximization). EM algoritmus je technika, ktorá pomáha vyrovnávať sa s chýbajúcimi dátami. Pozostáva z dvoch reťazových krokov: z kroku E a M. V kroku E sa vypočítajú podmienené očakávania chýbajúcich dát, výsledkom ktorých sú modelové a pozorované údaje. V kroku M sa hodnoty vypočítané v kroku E (modelové údaje) prispôbia s maximálnou vierohodnosťou štatistickému modelu („kompletnej“ sady údajov vrátane odhadov z kroku E). Tieto kroky sa opakujú, až pokiaľ algoritmus neskonverguje. Práce Flowerdewa a Greena (1989 a 1991) a Flowerdewa et al. (1991) sa týkajú dát s diskretným rozdelením, a to hlavne dát, ktorých modelom je Poissonovo rozdelenie pravdepodobnosti (s takýmto typom dát sa pravidelne stretávame v jednotlivých výskumoch, ale hlavne v sčítaniach obyvateľstva) alebo dát s binomickým rozdelením. Neskôr Flowerdew a Green (1992) túto metódu aplikovali na spojité premenné (priestorová interpolácia dát s normálnym rozdelením). Autori teda zaviedli niekoľko interpolácií, ich algoritmus však používa premenné cieľovej zóny na predikovanie záujmovej premennej. Ako indikátory za cieľové zóny sa použili viaceré premenné ako napr. volebné výsledky, počet voličov, počet áut na domácnosť, počet ľudí pochádzajúcich z Pakistanu a i. Výsledky ukázali významné zlepšenie v porovnaní s metódou priestorového váženia, avšak na druhej strane prekvapivo slabšie prispôbenie ako pri binárnych premenných. Z ich analýzy vyplýva, že jednoduchšie modely sú lepšie ako komplexnejšie (Fisher a Langford, 1995). Zároveň treba podotknúť, že ich metóda je obsah zachovávajúcou metódou.

Aj Goodchild et al. (1993) vo svojej práci uviedli niekoľko priestorovo-interpoláčnych metód založených na regresných modeloch. Použili 5 rôznych metód ako aj metódu priestorového váženia a pyknofylaktickú metódu. Rozvinuli alternatívnu metódu, ktorá využíva pomocné informácie z inej sady priestorových jednotiek, známej ako tzv. kontrolné zóny. Tieto by mal byť výskumník schopný vytvoriť na základe svojich vedomostí o študovanom území. O kontrolných zónach sa predpokladá, že ich obyvateľstvo je homogénne rozmiestnené. Po prekryve zdrojových a kontrolných zón na jednej strane, a kontrolných a cieľových zón na strane druhej, sa odhadne hustota kontrolných zón a na základe homogénneho charakteru kontrolných zón sa môže odhadnúť populácia cieľových zón. Výsledok ich práce spočíval v zistení, že kontrolné zóny výrazne vylepšujú interpolačný proces.

Pomerne novým prístupom, ktorý využíva stále sa rozvíjajúce a zlepšujúce techniky geografického informačného systému (nielen doteraz používané jednoduché mapové prekrytie, ale aj rôzne iné techniky rozdelenia a agregácie priestorových

dát) a stále sa zvyšujúce a dostupnejšie množstvo digitálnych databáz, rozvinul vo svojej práci Xie (1995). Ak totiž uvažujeme o probléme priestorovej interpolácie dát, kde ako premenné vystupujú charakteristiky vzťahujúce sa k obyvateľstvu alebo k bývaniu, je možné ako pomocnú informáciu využiť aj cestnú sieť. Xie rozvinul tri algoritmy pre priestorovú interpoláciu s využitím cestnej siete ako pomocných informácií. Sú to algoritmy, na základe ktorých sa premenná zdrojových zón (obyvateľstvo) prideliť cestným segmentom. Jedná sa o metódu dĺžky cestnej siete, hierarchicky váženej cestnej siete a bytovej únosnosti cestnej siete. Xie zistil, že všetky jeho metódy poskytujú lepšie výsledky ako metóda priestorového váženia a že metóda hierarchicky váženej cestnej siete vykazuje najlepšie výsledky. S priestorovo-interpolačnými metódami, ktoré ako pomocné informácie využívajú cestnú sieť experimentovali aj Voss et al. (1999) a Reibel a Buffalino (2005).

O tzv. modernej inteligentnej priestorovo-interpolačnej metóde hovorí Deichmann (1996). Používa množstvo rôznych pomocných údajov ako umiestnenie a veľkosť mestských území, riek, dopravných štruktúr, parkov a iných údajov, ktoré sa vzťahujú na rozmiestnenie obyvateľstva. Je to na rastru založená interpolačná metóda, ktorá vytvára sériu váh pre všetky bunky v zdrojovej zóne. Tieto váhy sa určujú heuristicky na základe poskytnutých pomocných informácií. Tento typ priestorovo-interpolačnej metódy rozšírili Turner a Openshaw (2001), keď do odhadu parametrov modelu začlenili tzv. neurálne siete. Neurálne siete zahŕňajú premenné ako vzdialenosti k parkom, k hlavným cestám, k riekam, mestským územiám, či dokonca odhad populácie z pyknofylaktického modelu (Hawley, 2005). Predmetom tejto štúdie bola najmä tzv. dezagregovaná priestorová interpolácia – t.j. prípad, keď majú cieľové geografické jednotky vyšší stupeň priestorového rozlíšenia ako zdrojové geografické jednotky (cieľové jednotky sú oveľa menšie ako zdrojové jednotky). Problém interpolácie dát je tak výrazne ťažší a založený na moderných priestorovo-interpolačných metódach, ktoré zahŕňajú používanie rôznych pomocných informácií, ktoré sprevádzajú proces interpolácie.

Na záver je potrebné zdôrazniť, že metódy bez a s využitím pomocných informácií sa môžu navzájom prepojiť a aplikovať za určitým účelom. Napríklad Langford a Unwin (1994) aplikovali povrchovú metódu (kernel vyhladenie) na výsledky dazymetrickej metódy za účelom vytvorenia kartograficky prijateľnejšej a informatívnej mapy. Množstvo metód bez využitia pomocných informácií môže tieto informácie do seba inkorporovať, ak sa tieto stanú dostupnými. Napríklad v rámci pyknofylactickej metódy – ak sú dostupné informácie o obývaných územiach, môže sa obyvateľstvo zo zdrojových jednotiek prideliť najprv do týchto obývaných polygónov, ktoré v nich ležia, pričom sa predpokladá, že neobývané polygóny majú nulovú populáciu. Následne sa uskutoční interpolácia prostredníctvom vyhladenia. Taktiež existuje množstvo prác, v ktorých sa môžu princípy štatistického modelovania inkorporovať napríklad do dazymetrickej metódy (priemerná hustota zaľudnenia land use území je odvodená z regresnej analýzy - vid' vyššie spomínané práce Langford et al., 1991 alebo Yuan, 1997).

3 PRESNOSŤ JEDNOTLIVÝCH METÓD TRANSFORMÁCIE PRIESTOROVÝCH DÁT

Ako vidieť z prehľadu literatúry zaoberajúcej sa problematikou interpolácie priestorových dát, ani jedna metóda neposkytuje úplne presné výsledky - vždy sa jedná o odhady. Spomínaní autori vo svojich štúdiách navrhli rôzne metódy, nie všetci ich však aj aplikovali v reálnej situácii a venovali pozornosť zhodnoteniu ich spoľahlivosti. Presnosti vybraných metód sa vo svojej práci venujú napríklad Flowerdew (1988), Flowerdew a Green (1989, 1991), Flowerdew et al. (1991), Goodchild et al. (1993) či Hawley (2005), ktorí interpolované hodnoty porovnávajú zo známymi hodnotami alebo napr. Fisher a Langford (1991), ktorí pracujú s ďalšími odhadmi. Okrem týchto spomenieme ešte Cockings et al. (1997), Fisher a Langford (1995), Gregory (2002) a Sadahiro (2000). Sadahiro (2000) predstavuje niekoľko detailných stochastických modelov chýb priestorového váženia. Podľa neho predstavuje interpolácia z menších zdrojových do väčších cieľových jednotiek relatívne menšie množstvo chýb ako v opačnom prípade. Podľa Fisher a Langford (1995), v minulosti mali vedci obmedzené možnosti ako porovnať výsledky interpolačného procesu a ako zhodnotiť chyby. Vo svojej práci použili techniku Monte Carlo, ktorá pracuje s problémom modifikovateľných územných jednotiek. Táto umožňuje násobnú interpoláciu vybranej premennej (obyvateľstvo) z jednej sady zdrojových jednotiek do viacerých sád cieľových jednotiek. Vytvorí sa tak viacero testovacích situácií pre metódy priestorovej interpolácie tých istých dát pre to isté geografické územie. Takto sú preskúvané vlastnosti chýb spojených s určitým interpolačným modelom. Fisher a Langford (1995) testovali metódu priestorového váženia, regresné metódy a dazymetrickú metódu. Podľa tejto dvojice autorov je priestorové váženie najmenej presnou interpolačnou metódou. Naopak, najpresnejšie výsledky v porovnaní s ostatnými metódami poskytla binárna dazymetrická metóda. Ukázala sa odolná voči zmenám hustoty zaľudnenia, ktoré sú spojené s určitým typom land use ako aj voči anomáliam vysokourbanizovaných, ale riedko osídlených území (Fisher a Langford, 1995). Je to predovšetkým v dôsledku jej vlastnosti zachovania obsahu interpolovanej premennej, kedy sa v procese transformácie do rastrovej reprezentácie uchová počet obyvateľov zdrojových zón. Všetky chyby sú tak neodmysliteľne limitované zmenou v každej individuálnej zdrojovej jednotke. Dazymetrická metóda s využitím dát land use je podľa tejto dvojice autorov odolnejšia v porovnaní s regresnými alebo povrchovými metódami aj voči chybám z klasifikácie snímky. Z výskumu ďalej vyplynulo, že presnosť všetkých interpolačných metód sa zvyšuje, ak množstvo cieľových jednotiek klesá. Presnosť výsledkov podľa nich čiastočne zvyšuje aj komplexnosť (resp. obtiažnosť) regresných modelov. To kontrastuje s názormi Flowerdew a Green (1994), ktorí sú presvedčení, že jednoduchšie modely sú úspešnejšie. Práca Cockings et al. (1997) nadväzujúc na predchádzajúcu prácu Fishera a Langforda (1995) navrhuje opatrenia na parametrizáciu chýb priestorovej interpolácie. Použila sa podobná technika Monte Carlo, testovali sa ale len dve metódy, a to dazymetrická metóda a priestorové váženie. Zistili, že geometrické charakteristiky cieľových zón sú významným faktorom chýb, čo sa týka metódy

priestorového váženia, ale nie chýb, ktoré sú spojené s dazymetrickou metódou. Chyby v dazymetrickej metóde korelujú s hustotou obyvateľstva.

Podobne, Mrozinski a Cromley (1999) poukázali na to, že odhady, ktoré poskytnú dazymetrická metóda sú presnejšie v porovnaní s tými, ktoré boli výsledkom priestorového váženia a pyknofylactickej metódy. Taktiež v práci Gregory (2002) sa preukázali presnejšie výsledky binárnej dazymetrickej metódy kombinovanej s EM algoritmom v porovnaní s ostatnými priestorovo-interpoláčnymi metódami. Z novších štúdií, Hawley (2005) porovnal a zhodnotil výsledky rôznych metód priestorovej interpolácie (priestorové váženie, pyknofylactická metóda, dazymetrická metóda a metóda hierarchicky váženej cestnej siete), ktoré ešte neboli doposiaľ navzájom otestované. Ako zdrojové a cieľové geografické jednotky použil cenové hierarchické zóny, čo mu umožnilo porovnať hodnoty odhadnutej a aktuálnej populácie. Najlepšie výsledky spomedzi všetkých testovaných metód vykazovala sieťová metóda. Ako sľubná sa javila aj dazymetrická metóda. V prípade nedostatku pomocných informácií sa ako lepšia voľba javí pyknofylactická metóda v porovnaní s metódou priestorového váženia. Celkovo sa však ukázalo, že všetky metódy vykazujú lepšie výsledky v oblastiach s vysokou hustotou zaľudnenia.

Ako však konštatuje Gregory (2002), väčšina prác, ktoré sa zaoberajú presnosťou interpoláčnych techník, je zameraná na údaje typu celkového počtu obyvateľov a len málo autorov sa venuje komplikovanejším premenným. Doteraz používané metódy sú obmedzené aj v dôsledku približného odhadu, ktorý je založený na predpokladanom matematickom vzťahu medzi záujmovou premennou a pomocnými premennými. Podstatné však je, že ani jedna metóda neposkytuje úplne presné odhady cieľových dát. Zásadne by mala byť preto každá metóda sprevádzaná kvantitatívnou mierou spoľahlivosti. To by sa malo brať do úvahy aj pri inkorporácii interpolovaných hodnôt do prostredia GIS. Témou chýb v GIS a ich reprezentáciou sa z rôznych hľadísk zaoberajú Goodchild a Gopal (1989).

4 KLASIFIKÁCIA METÓD TRANSFORMÁCIE PRIESTOROVÝCH DÁT

Ako vyplýva z vyššie uvedeného prehľadu literatúry, existuje pomerne veľké množstvo rôznych transformačných metód, ktoré sa od seba líšia z viacerých hľadísk. Vo všeobecnosti sa v dostupnej literatúre nemáme možnosť stretnúť s akýmsi uceleným pohľadom na jednotlivé metódy transformácie priestorových dát. Všetci autori vo svojich prácach podávajú návrh jednej (či už priestorovo-interpoláčnej alebo štatistickej) metódy, prípadne uvažujú o viacerých priestorovo-interpoláčnych metódach, z ktorých jedna je takmer vždy priestorové váženie. Delenie jednotlivých metód podľa určitých kritérií implicitne vidieť vo väčšine spomenutých prác, nikde však nie je uvádzaná ich komplexná klasifikácia. Akési triedenie jednotlivých metód odhadu chýbajúcich dát možno vidieť v prácach Goodchild a Lam (1980), Goodchild a Kemp (1990), Goodchild et al. (1993), Hawley (2005), Lam (1983) a Wu et al. (2005).

Goodchild a Lam (1980) a Lam (1983) rozdeľujú priestorovo-interpoláčn  metody na z klade toho,  i s  jednotliv  metody schopn  uchovať rozsah premennej interpolovanej zo zdrojov ch do cieľov ch geografick ch jednotiek. Rozlišuj  teda tzv. **obsah nezachovávajúce** a **obsah zachovávajúce metody**. K prv m menovan m zaraďuj  bodovo založen  (point-based) interpoláčn  metody, k obsah zachovávajúcim zase  zemne založen  (areal-based) interpoláčn  metody: metodu priestorov ho v zenia a pyknofylaktick  metodu.

Novšie v skumy t kajúce sa tejto problematiky podávajú Goodchild et al. (1993), ktorí identifikuj  tri hlavn  typy interpolácie: **kartografick , regresn  a povrchov  metody**. Tak to rozdelenie jednotliv ch interpoláčn ch met d pouzivaj  aj Fisher a Langford (1995). Ku kartografick m met dam zaraďuj  metodu priestorov ho v zenia a dazymetrick  metodu. K regresn m met dam patria tie priestorovo-interpoláčn  premenn , ktor  v sebe zahrňaj  ist  variantu regresn ch modelov. S  to tie metody, ktor  vo svojich pr cach pouzili napríklad Flowerdew (1988), Flowerdew a Green (1989 a 1991), Goodchild et al. (1993) a Langford et al. 1991. K povrchov m met dam patria bodovo založen  priestorovo-interpoláčn  metody a Toblerova pyknofylaktick  metoda.

Podľa Goodchild a Kemp (1990) existuje veľk  množstvo algoritmov navrhnut ch pre probl m priestorovej interpolácie. Tieto klasifikovali na z klade dichotomie do  iestich p rov: bodov  vs. plošn , glob ln  vs. lok ln , exaktn  vs. aproximuj ce, stochastick  vs. deterministick , obsah zachovávajúce vs. obsah nezachovávajúce a s visl  vs. nes visl . Keďže sa jednotliv  algoritmy od seba navz jom fundament lne odlišuj , a to v met dach prerozdelenia hodn t zdrojov ch jednotiek, rozhodla sa t to dvojica autorov identifikovať tri v seobecn  typy priestorovo-interpoláčn ch techniek. S  nimi **priestorovo vyhladzujúce** (založen  na bodoch alebo rastoch) **metody**, ktor  delia na **glob lny** a **lok lny vyhladzovac  algoritmus**. Glob lny algoritmus je skupina met d, ktor  produkuje priblizn  interpoláčn  v sledky. Mono ich v podstate stotoniť s v ššie spom nan mi bodovo založen mi priestorovo-interpoláčn mi met dami. Interpolácia hodn t nast va prostredn ctvom jednotliv ch techniek ako kriking, anal za trendov ho povrchu, Furierova anal za, anal za klzav ho priemeru a i. Lok lny vyhladzovac  algoritmus je akousi vylepšenou verziou predchádzajúceho typu, ktor  zvyčajne produkuje presn  interpoláciu. Patrí sem Toblerov pyknofylaktick  pr stup, ale aj tzv. B spline,  i proxim lna anal za. Ďalšou skupinou s  **priestorovo v zen ** (založen  na ploche alebo vektore), kde autori zaraďuj  metodu priestorov ho v zenia. Poslednou skupinou met d s  **modelovacie** (štatisticky založen ) priestorovo-interpoláčn  metody. S  to metody, ktor  mono prirovnať k regresn m met dam z predchádzajúceho delenia.

Hawley (2005) vo svojej pr ci pod va doteraz najutriedenejši pohľad na jednotliv  priestorovo-interpoláčn  metody vhodné pre harmonizáciu d t hum nnogeo-grafickej povahy. Podľa Lam (1983) ich rozdeľuje na tie, ktor  nevyuzivaj  pomocn  informácie a tie, ktor  pouzivaj  pomocn  informácie. Tieto deli na metody priestorovej interpolácie s vyutím d t z DPZ, metody priestorovej interpolácie s vyutím d t cestnej siete a ostatn  metody priestorovej interpolácie s vyutím pomocn ch d t. Medzi **metody priestorovej interpolácie bez vyutia pomocn ch**

informácií zaraďuje tento autor *metódu prekryvu, pyknofylaktickú metódu a bodovo založené interpolačné metódy*. K **metódam priestorovej interpolácie s využitím dát z DPZ** zaraďuje *regresné metódy a dazymetrické metódy*. Medzi **metódy priestorovej interpolácie s využitím dát cestnej siete** radí *tri algoritmy pre priestorovú interpoláciu s využitím cestnej siete* od Xie (1995) - metódu dĺžky siete, metódu hierarchicky váženej siete a metódu bytovej únosnosti cestnej siete. K tzv. **ostatným metódam priestorovej interpolácie s využitím pomocných dát** zaraďuje *EM algoritmus Flowerdewa a Greena* (1991, 1992 a 1994), *metódu kontrolných zón* (Godchild et al., 1993) a tzv. inteligentnú priestorovo-interpolačnú metódu (Deichman, 1996 a Turner a Openshaw, 2001).

Na rozdiel od vyššie uvedených triedení, práca Wu et al. (2005) rozlišuje zvlášť **metódy priestorovej interpolácie dát** a zvlášť **metódy štatistického modelovania**. Podľa Wu et al. (2005), sú metódy priestorovej interpolácie dát v prvom rade určené na problém transformácie zón, ktorý zahŕňa prerozdelenie dát z jednej sady geografických jednotiek do druhej pomocou určitých interpolačných operácií. Na druhej strane, štatistické modelovanie sa viac zaujíma o odvodenie vzťahu medzi odhadovanou premennou (obyvateľstvom) a inými, zväčša socioekonomickými premennými prostredníctvom teórií geografie mesta. Tento prístup bol pôvodne navrhnutý na odhad populácie v medzicenzovom období alebo populácie územia odlišného od toho, za ktoré sa údaje vypublikovali z cenzu. Tieto prístupy môžu a nemusia byť inkorporované aj do procesu transformácie. Metódy priestorovej interpolácie ďalej delí do dvoch kategórií podľa využitia pomocných informácií. V podstate sa jeho klasifikácia zhoduje s triedením, ktoré použil Lam (1983). Metódy štatistického modelovania delí na základe vzťahu medzi obyvateľstvom a 1) mestskými územiami, 2) land use, 3) obytnými jednotkami, 4) charakteristikami pixlov obrázka a 5) inými fyzickými a socioekonomickými charakteristikami.

Metódy transformácie priestorových dát možno na základe dostupnej literatúry a najdôležitejších charakteristík jednotlivých metód používaných na odhad vybranej premennej rozlíšiť do troch rovín, v rámci ktorých delíme jednotlivé metódy:

A. Podľa formy:

1. **Klasické priestorovo-interpolačné metódy**

1.1. Bodovo založené (bodové) PIM

1.2. Územne založené PIM

1.2.1. Rastrové PIM

1.2.2. Vektorové PIM

2. **Priestorovo-interpolačné metódy s využitím princípov štatistického modelovania**

K rastrovým PIM zaraďujeme pyknofylaktickú metódu, k vektorovým metódu priestorového váženia, dazymetrickú metódu a sieťovú metódu.

K PIM s využitím princípov štatistického modelovania zaradujeme také metódy odhadu neznámej premennej, ktoré inkorporujú matematické a štatistické modely do priestorovej interpolácie.

B. Podľa obsahu:

1. **Obsah nezachovávajúce PIM**
2. **Obsah zachovávajúce PIM**

PIM, ktoré nezachovávajú obsah interpolovanej premennej sú bodové PIM a PIM s využitím princípov štatistického modelovania, k obsah zachovávajúcim patria územne založené PIM z predchádzajúceho delenia.

C. Podľa využitia pomocných informácií:

1. **Jednoduché PIM (PIM bez využitia pomocných informácií)**
2. **Zložené/inteligentné PIM (PIM s využitím pomocných informácií)**
 - 2.1. Špecifické – s využitím pomocných informácií za zdrojové jednotky
 - 2.2. Flexibilné – s využitím pomocných informácií za cieľové jednotky alebo za inú sadu zón, prípadne s využitím viacerých pomocných sád dát

K jednoduchým priestorovo-interpoláčnym metódam patria bodové PIM, priestorové váženie a pyknofylaktická metóda, k špecifickým zahŕňame dazymetrickú metódu a sieťové metódy. Ako o flexibilných PIM uvažujeme o PIM s využitím štatistického modelovania, kontrolných zón a o neurálnych sieťach.

5 VŠEOBECNÉ POROVNANIE A VYUŽITEL'NOSŤ METÓD TRANSFORMÁCIE PRIESTOROVÝCH DÁT

V predchádzajúcej časti sme čitateľovi poskytli prehľad množstva metód používaných na odhad neznámej hodnoty vybranej premennej za zvolený zonálny systém. Výber konkrétnej metódy však závisí od rôznych faktorov, akými môžu byť napríklad časová náročnosť, dostupnosť dát, ľahká implementácia, presnosť či softvérové možnosti. Ak napríklad uvažujeme o obyvateľstve ako o záujmovej premennej, potom metóda priestorového váženia predpokladá jej homogénne rozmiestnenie. Hoci je to metóda, ktorá sa vyznačuje pomerne jednoduchou aplikáciou, záujmová premenná je v skutočnosti len zriedkavo rovnomerne rozmiestnená. Pyknofylaktická metóda naopak uvažuje o heterogénnom rozmiestnení údajov. Bodové PIM vytvárajú pomerne dramatické vizualizácie rozmiestnenia interpolovanej premennej (Langford et al., 1991). Spolu s rastrovými PIM sú okrem toho pomerne technicky náročné metódy. Spravidla obsahujú interaktívne a iteratívne zisťovania priestorového správania študovaného fenoménu. Implementácia týchto prístupov si vyžaduje presné a jasné vedomosti o dátach, technickú zručnosť a dokonalé porozumenie priestorovej štatistike. Navyše, existuje len zopár softvérových balíčkov, ktoré obsahujú

funkcie schopné vykonať takúto priestorovú interpoláciu. Čo sa týka zložitejších metód, metódy ktoré využívajú údaje z DPZ ako pomocné informácie, predpokladajú, že rozmiestnenie obyvateľstva sa vzťahuje na jednotlivé land use územné typy. Tieto metódy očakávajú vyššiu hustotu zaľudnenia v mestských a prímestských land use typoch ako napríklad v poľnohospodárskych alebo zalesnených typoch land use. Sieťové metódy zase predpokladajú, že obyvateľstvo súvisí s hustotou cestnej siete. Okrem toho, hierarchická sieťová metóda berie do úvahy aj triedy jednotlivých ciest. PIM s využitím princípov štatistického modelovania aplikujú uznávané formuly pre odhad hodnôt záujmovej premennej. Používajú náhodný spôsob na vyriešenie neurčitosti problému priestorovej interpolácie. Predpokladajú, že rozmiestnenie záujmovej premennej možno úplne presne opísať matematickým jazykom alebo definovať vzorcom. Tento predpoklad síce zvyšuje komplexnosť problematiky priestorovej interpolácie, ale na druhej strane limituje jej aplikáciu, pretože s takýmito predpokladmi o štatistických vlastnostiach dát sa len zriedka stretávame v praxi. Tieto metódy okrem toho prijali aproximujúci prístup, ktorý stanovuje jednoduchú funkciu rozvrhnutú naprieč celým študovaným územím. Z jednotlivých pozorovaní sa teda odvodí všeobecný vzorec na základe nejakých predpokladaných štatistických pravidiel a potom sa tento aplikuje na individuálne študované jednotky. Počas tohto procesu sa ale zvyčajne stratí pôvodný rozsah odhadovanej premennej, čím sa zredukuje aj presnosť odhadu pre cieľové jednotky v porovnaní so zdrojovými. Tieto metódy berú do úvahy hodnoty tzv. predikovaných premenných, ktoré sú v nejakom vzťahu k záujmovej premennej, ale tieto sú dostupné za skupinu cieľových geografických jednotiek. Metóda kontrolných zón naopak predpokladá využitie tretej sady zón, o ktorej predpokladá, že má homogénnu hustotu zaľudnenia. Metódy neurálnych sietí zase využívajú množstvo rozmanitých údajov na vylepšenie presnosti interpolácie.

Tabuľka 1 porovnáva vybrané doteraz spomenuté metódy transformácie dát z hľadiska niektorých ich dôležitých charakteristík, ako sú predpoklady o rozmiestnení premenných, využitie pomocných informácií, priestorovej dimenzie a funkcií. Charakteristika rozmiestnenia premennej udáva, či niektorá z metód predpokladá homogénne alebo heterogénne rozmiestnenie premennej v zdrojových zónach, charakteristika pomocných dát hovorí, či a aké doplnkové informácie sa použili na zlepšenie interpolačného procesu, dimenzia sa vzťahuje na pôvodné dáta a na rôzne pomocné dáta, ktoré môžu byť zahrnuté v určitej priestorovo-interpolačnej metóde. Charakteristika funkcie sa vzťahuje na hlavné typy priestorových operácií, ktoré sa používajú na implementáciu určitej priestorovo-interpolačnej metódy (Hawley, 2005).

Jednotlivé metódy transformácie priestorových dát možno využiť v rôznych oblastiach:

- na odhad počtu obyvateľov v určitej sade zón (napr. máme k dispozícii údaje za určitú sadu zón, ale chýbajú nám dôležité informácie, ako napr. počet obyvateľov týchto zón. Údaje zhromaždené pre inú sadu zón však takéto informácie k dispozícii majú. Obyvateľstvo nami zvolených zón možno odhadnúť zo zdrojových zón prostredníctvom rôznych transformačných nástrojov

- Okrem odpovede na otázku koľko obyvateľov žije v určitom záujmovom území sa môžu metódy priestorovej interpolácie zároveň využiť aj v iných oblastiach, napr. prostredníctvom priestorovo-interpoláčnych metód možno odhadnúť hranice dosahu územia zasiahnutého prírodnou katastrofou a v nich určiť počet postihnutých obyvateľov, domov, obchodov atď.
- Na porovnanie priestorovej štruktúry (politických, administratívnych a i. hraníc) za určité časové obdobie

Tabuľka 1 Porovnanie vybraných metód transformácie priestorových dát (Zdroj: Hawley, 2005)

METÓDA	ROZMIESTNENIE PREMENNEJ	POMOCNÉ DÁTA	DIMENZIA	FUNKCIE
Priestorové vázanie	homogénne	žiadne	2-D	priestor. prekryv
Pyknofilaktická	heterogénne	žiadne	2-D, 2 ½-D, 3-D	susedstvo, vyhladenie, zonálna štatistika
Bod v polygóne	homogénne	žiadne	2-D, 0-D	Bod v polygóne
Regresná	heterogénne	land use/land cover	2-D	regresia, klasifikácia, priestor. prekryv
Dazymetrická	heterogénne	land use/land cover	2-D	priestor. prekryv
Sieťová	heterogénne	cestná sieť	2-D, 1-D	klasifikácia, priestor. prekryv
EM algoritmus	heterogénne	rôzne premenné	2-D	štatistický algoritmus
Kontrolné zóny	heterogénne	používateľom definované zóny	2-D	priestor. prekryv, digitalizácia
Inteligentná	heterogénne	rôzne premenné	0-D, 1-D, 2-D, 3-D	heuristický algoritmus

6 ZÁVER

Metódam odhadu geografických dát sa venuje veľké množstvo zahraničných autorov predovšetkým z oblasti GIS a DPZ. Všetky metódy transformácie priestorových dát sa vyznačujú pomerne veľkou zložitosťou, preto je nevyhnutná ich podrobná analýza. Prístupy k riešeniu tejto problematiky, ako aj samotný výber transformačných metód však závisí od viacerých faktorov. Ani jedna metóda však nie je úplne spoľahlivá, nakoľko všetky metódy poskytujú „iba“ odhady danej premennej.

Kvalita odhadov závisí predovšetkým od toho, ako sú definované zdrojové a cieľové jednotky, od stupňa generalizácie v interpolačnom procese, ako aj od charakteristik deleného povrchu. Všetci autori venujúci sa spomínanej problematike uvažujú vo svojich prácach len o jednej odhadovanej premenej, a to o obyvateľstve. Ostatné socioekonomické premenné využívajú len ako pomocné informácie. Všetci autori vo svojich prácach podávajú návrh jednej (či už priestorovo-interpolačnej alebo štatistickej) metódy, prípadne uvažujú o viacerých transformačných metódach, z ktorých jedna je takmer vždy priestorové váženie. Delenie jednotlivých metód podľa určitých kritérií implicitne vidieť vo väčšine spomenutých prác, nikde však nie je uvádzaná ich komplexná klasifikácia. Vo všeobecnosti možno metódy transformácie priestorových dát rozdeliť na základe stanovených cieľov a dostupných a požadovaných informácií na metódy priestorovej interpolácie a metódami štatistického modelovania. Z doteraz spomenutých metód transformácie priestorových dát vyberieme tie, ktoré sú vhodné na humánogeografické analýzy a v blízkej budúcnosti sa ich pokúsime aplikovať a otestovať v slovenských podmienkach. Bude to metóda priestorového váženia, pyknofylaktická metóda, dazymetrická metóda s využitím dát z land use a jedna zo siet'ových metód. Metódu priestorového váženia sme si vybrali preto, lebo je to základná priestorovo-interpolačná metóda so širokými možnosťami uplatnenia. Pyknofylaktickú metódu sme si vybrali preto, aby sme ju mohli porovnať s metódou priestorového váženia, nakoľko obe metódy nevyužívajú pomocné informácie, ale líšia sa v predpoklade o rozmiestnení premennej v sledovanom území. Zo skupiny jednoduchých metód, ktoré využívajú pomocné informácie vyberieme dazymetrickú metódu s využitím dát z land use. Z modernejších metód sa pre porovnanie s dazymetrickou metódou vybrala metóda dĺžky a hierarchicky váženej cestnej siete. Výsledky z metódy využívajúcej land use bude totiž zaujímavé porovnať s výsledkami metódy, ktoré ako pomocné dáta využíva cestnú sieť.

PodĎakovanie

Príspevok vznikol ako súčasť grantu VEGA 1/0181/09 „Súčasný procesy redistribúcie obyvateľstva na Slovensku“.

Literatúra

- COCKINGS, S., FISHER, P., LANGFORD, M. 1997. Parameterization and visualization of the errors in areal interpolation. *Geographical Analysis* 29(4): 314-328.
- EICHER, C. L., BREWER, C. A. 2001. Dasymeric Mapping and Areal Interpolation: Implementation a Evaluation. *Cartography and Geographic Information Science*. 28(2): 125-138.
- FISHER, P., LANGFORD, M. 1995. Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A* 27: 211-24.
- FISHER, P., LANGFORD, M. 1996. Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping. *Professional Geographer* 48(3): 299-309.

- FLOWERDEW, R., OPENSHAW, S. 1987. A review of the problems of transferring data from one set of areal units to another incompatible set. Northern Regional Research Laboratory: research report.
- FLOWERDEW, R., GREEN, M. 1989. Statistical methods for inference between incompatible zonal systems. In: M. Goodchild and S. Gopal (eds), Accuracy of spatial databases. London, U.K.: Taylor a Francis, pp. 239-47.
- FLOWERDEW, R., GREEN, M. 1991. Data integration: Statistical methods for transferring data between zonal systems. In: I. Masser and M. B. Blakemore (eds), Haling geographical information. Essex, U.K.: Longman Scientific & Technical, pp. 38-53.
- FLOWERDEW, R., GREEN, M. 1992. Developments in areal interpolation methods and GIS. *Annals of Regional Science* 26: 67-78.
- FLOWERDEW, R., GREEN, M. 1994. Areal interpolation and types of data. In: S. Fotheringham and P. Rogerson (eds), *Spatial analysis a GIS*. London, U.K.: Taylor and Francis. pp. 121-45.
- FLOWERDEW, R., GREEN, M., KEHRIS, E. 1991. Using areal interpolation methods in GIS. *Papers in Regional Science: The Journal of the RSAI* 70(3): 303-15.
- GOODCHILD, M. F., LAM, N. S. 1980. Areal interpolation: A variant of the traditional spatial problem. *Geoprocessing* 1: 297-312.
- GOODCHILD, M. F., ANSELIN, L., DEICHMANN, U. 1993. A framework for the areal interpolation of socioeconomic data. *Environment and Planning A* 25: 383-97.
- GREGORY, IAN N. 2002. The accuracy of areal interpolation techniques: Standardising 19th and 20th century census data to allow long-term comparisons. *Computers Environment and Urban Systems*, 26 (4). pp. 293-314. ISSN 0198-9715
- LAM, N. 1983. Spatial interpolation methods: a review. *American Cartographer* 10,129-149
- GREEN, M. 1989. Statistical methods for areal interpolation: the EM algorithm for count data. Lancaster, UK: North West Regional Research Laboratory.
- HOLT, J., LO, C.P., HODLER, T. W. 2004. Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science*, April 1.
- LANGFORD, M., MAGUIRE, D. J., UNWIN, D. J. 1990. Mapping the density of population: Continuous surface representations as an alternative to choropleth and dasymetric maps. Leicester, U.K.: Midlas Regional Research Laboratory, University of Leicester and Loughborough, University of Technology, U.K.
- LANGFORD, M., MAGUIRE, D. J., UNWIN, D. J. 1991. The areal interpolation problem: Estimating population using remote sensing in a GIS framework. In: I. Masser and M. B. Blakemore (eds), *Haling geographic information*. Essex, U.K.: Longman Scientific & Technical. pp. 55-77.
- LANGFORD, M., UNWIN, D. J. 1994. Generating a mapping population density surfaces within a geographical information system. *The Cartographic Journal* 31: 21-6.
- MENNIS, J., 2003. Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55: 31-42.
- MONMONIER, M., SCHNELL, G. A. 1984. Land use and land cover data and the mapping of population density. *International Yearbook of Cartography* 24: 24-9.
- MROZINSKI, R.D., JR. CROMLEY, R.G., 1999. Singly . and doubly . constrained methods of areal interpolation for vector-based GIS. *Transactions in GIS*, 3: 285-301.
- OPENSHAW S, 1977. "Optimal zoning systems for spatial interaction models" *Environment and Planning A* 9: 169 – 184
- REIBEL, M., BUFALINO, M.E. 2005. Street weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A*, 37(1): 127-9.
- SADAHIRO, Y. 2000. Accuracy of count data estimated by the point-in-poly method. *Geographical Analysis* 32 (1): 64-89.

- TOBLER, W. 1979. Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association* 74:519–30.
- VOSS, P., LONG, D.D., HAMMER, R.B. 1999. When census geography doesn't work: Using ancillary information to improve the spatial interpolation of demographic data. *Center for Demography and Ecology Working Paper*: 98-107.
- WRIGHT, J. K. 1936. A method of mapping densities of population with Cape Code as an example. *The Geographical Review* 26: 103-110.
- WU, S., QIU, L. W. 2005. Population estimation methods in GIS and remote sensing: A review. *Gis and Remote Sensing* 42: 80-96.
- XIE, Y. 1995. The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems* 19: 287–306.
- YUAN, Y., SMITH, R., LIMP, W. 1997. Remodelling census population with spatial information from Landsat TM imagery, *Computers, Environment and Urban Systems* 21: 245–258.

Spatial data transformation's methods: review, classification and evaluation

Summary

This paper reviews existing areal data transformation's methods which can estimate arbitrary socioeconomic variables available for one geographical system but needed for another zones. Many methods for data transformation have been reported in the GIS and remote sensing literature. All the authors reflect on one interpolated variable, namely population. No method is truly satisfactory in the sense that it is impossible to derive perfect results – results of these methods are always “only” estimates. One can see the implicit dividing of transformation's methods in the most of the studies about this problem but nowhere the complex classification occurs. Depending on the goal of the research and the required information, we propose classify these methods in three levels: according to content we distinguish areal interpolation's methods and statistical methods, according to form we can divide two categories depending on whether ancillary information is used or not and finally we can identify volume preserving and volume non-preserving methods.

The choice of which method to use may be dependent on various factors such as data availability, ease of implementation, accuracy and time. In this paper we compare different transformation's methods. Into the future we want to apply and compare the chosen areal interpolation's methods, test their reliability and result's accuracy. Probably it will be the areal weighting method, the pycnophylactic method, a dasymetric method using remote sensing data, and the network length or hierarchical weighted method.