

VYUŽITIE POMOCNÝCH INFORMÁCIÍ PRI TRANSFORMÁCII GEOGRAFICKÝCH DÁT – VYHODNOTENIE VYBRANÝCH POSTUPOV

Michala Madajová*

* Geografický ústav SAV, Štefánikova 49, 814 73 Bratislava, geogmada@savba.sk

Use of auxiliary information in transformation of geographical data: assessment of selected procedures.

This paper introduces and evaluates some geographical data transformation methods which allow data allocation from spatially inconsistent zonal systems. The basic and simplest transformation method – areal weighting – is compared with complex techniques which use auxiliary information to improve the results of the interpolation process. Such auxiliary information is the length and type of road segments, the node counts in street networks and the class of urban fabric pursuing the CORINE land cover nomenclature. For comparison, different data types were used. Apart from the usually estimated variable – population counts – it was the area of agricultural land. These variables were available for the 1991 census districts and were estimated for the system of functional urban regions – FUR 91-B. The measures of global fit and spatial visualization of percent errors associated with all of the variables and methods were used to evaluate the results of estimation.

Key words: accuracy, area weighting, dasymetric method, interpolation, land cover, road network, transformation

ÚVOD

Transformácia geografických dát je súbor procesov, prostredníctvom ktorých je možné kombinovať a analyzovať údaje dostupné z rôznych zdrojov a za rozličné geografické jednotky. Rieši teda problém tzv. priestorového nesúladu zonálnych systémov, pod ktorým rozumieme situáciu, keď máme k dispozícii agregované údaje za taký systém geografických jednotiek, ktorý nekorešponduje s nami zvolenými (želanými) geografickými jednotkami. Problém priestorovej inkonzistencie výrazne vzrastá v oblasti interdisciplinárneho výskumu, napríklad pri snahe o integráciu demografických, sociálnych, ekonomických či fyzicko-geografických dát. Aj samotní geografi sa pri svojich analýzach častokrát stretávajú s problémom porovnania údajov dostupných z rôznych zdrojov a za odlišné geografické jednotky. Na Slovensku sa údaje o obyvateľstve publikujú spravidla v desaťročných periódach a za istý stupeň priestorovej agregácie (obce, okresy, ...). Pre geografické analýzy sú však v mnohých prípadoch vhodnejšie iné zóny ako jednotlivé administratívne jednotky. Ak teda potrebujeme použiť a hlavne zjednotiť údaje z rôznych zdrojov alebo časových období, je nevyhnutné pristúpiť k harmonizácii týchto dát, a to prostredníctvom rozličných metód.

Prevod údajov z jedného systému geografických jednotiek do nejakého iného zonálneho systému je proces, ktorý sa zvykne označovať ako areálová interpolácia, resp. ako problém areálovej interpolácie (Goodchild a Lam 1980, Lam 1983, Flowerdew a Openshaw 1987, Flowerdew et al. 1991 a Langford et al.

1991). Areálová interpolácia je v tomto zmysle potom považovaná za typ priestorovej interpolácie. My však budeme proces odhadu dát z navzájom nekompatibilných zonálnych systémov označovať všeobecnejšie – ako transformáciu geografických dát z jedného systému geografických jednotiek do druhého (pozmeneného alebo úplne odlišného) zonálneho systému. Ide predovšetkým o to, aby sme sa vyhli nejasnostiam pri používaní pojmov „areálová a priestorová interpolácia“ v súvislosti s metódami odhadu dát za navzájom nekompatibilné územné systémy. K riešeniu tejto problematiky možno pristupovať nielen cez využitie areálovo-interpoláčnych, ale aj iných priestorovo-interpoláčnych prístupov, prípadne metód pracujúcich na princípoch štatistického modelovania.

Predkladaný príspevok má preto za cieľ predstaviť základné postupy a metódy, ktoré sa využívajú v procese transformácie geografických dát získaných za nekompatibilné zonálne systémy, a na príklade odhadov hodnôt dvoch premenných odlišného charakteru uplatniť a navzájom porovnať presnosť výsledkov vybraných transformačných metód v slovenských podmienkach.

METÓDY TRANSFORMÁCIE GEOGRAFICKÝCH DÁT

V dostupnej zahraničnej literatúre máme možnosť stretnúť sa s viacerými metódami, ktoré umožňujú odhadnúť dáta za odlišné zonálne systémy. Vzhľadom na uvedené skutočnosti ich budeme označovať ako metódy transformácie geografických dát. Tieto metódy možno rozdeliť podľa rôznych kritérií do niekoľkých skupín. V princípe sa od seba odlišujú jednak na základe odlišných predpokladov o rozmiestnení premennej v danom území, použitia pomocných informácií či využitia princípov štatistického modelovania v transformačnom procese, ale aj na základe stupňa generalizácie a tzv. pyknofylactickej vlastnosti (Tobler 1979), ktorá zaručuje zachovanie rozsahu premennej v procese odhadu jej hodnôt. Výber konkrétnej metódy preto závisí od rôznych faktorov, ku ktorým mimo iného pristupujú aj také vlastnosti, akými sú napríklad časová náročnosť, dostupnosť dát, ľahká implementácia a presnosť či softvérové možnosti (podrobnosti pozri v Madajová, v tlači). V predkladanom príspevku budeme uvažovať o tých metódach transformácie geografických dát, ktoré v podstate vychádzajú z klasifikácie založenej na využívaní pomocných informácií v procese odhadu dát. Vo všeobecnosti budeme hovoriť o tzv. jednoduchých a zložitých (inteligentných) transformačných metódach (Madajová, v tlači).

V tejto súvislosti sa zaužíval pojem „zdrojové geografické jednotky“ pre zóny, kde máme k dispozícii údaje a „cieľové geografické jednotky“ pre územia, kde tieto dáta potrebujeme (Goodchild a Lam 1980).

Najčastejšie používanou metódou odhadu dát za nekompatibilné územné jednotky je metóda prekryvu alebo priestorového váženia (Goodchild a Lam 1980, Lam 1983, Flowerdew a Green 1989). Ide o najjednoduchšiu transformačnú metódu, ktorá prerozdeľuje údaje zo zdrojových zón na základe váhy odvodennej z územia prieniku zdrojových a cieľových zón. Táto metóda predpokladá homogénne rozmiestnenie záujmovej premennej v zdrojových zónach, čo sa považuje za jej najväčšiu nevýhodu. Odporúča sa použiť v prípade, keď nie sú dostupné žiadne doplnkové informácie, ktoré by pomohli skvalitniť transformačný proces a vylepšiť odhady jednotlivých premenných. Naopak, tzv. pyknofylaktická metóda, navrhnutá Toblerom (1979), predpokladá heterogénne roz-

miestnenie premennej (hustoty zaľudnenia) v každej zdrojovej zóne. Je to metóda, ktorá sa na základe informácie o hustote zaľudnenia zdrojových zón pokúša odhadnúť spojitý povrch tejto premennej v uzloch jemnej mriežky. Táto metóda berie do úvahy efekt priľahlých zdrojových zón. Vyhladený povrch hustoty zaľudnenia sa nakoniec použije na výpočet hodnôt tejto premennej v cieľových jednotkách.

Metódy transformácie geografických dát, ktoré využívajú pomocné informácie, nazývame zložené (inteligentné) transformačné metódy. Ako pomocné informácie v procese odhadu dát by sa mali použiť také údaje, o ktorých sa predpokladá, že odrážajú priestorové rozmiestnenie zobrazovaných premenných, prípadne s nimi nejakým spôsobom súvisia. V dostupnej literatúre zaoberajúcej sa danou problematikou sa pracuje zvyčajne len s počtom obyvateľov ako s odhadovanou premennou, pričom ako pomocné informácie sa v ostatnom období zaužívali údaje z diaľkového prieskumu Zeme. Existuje niekoľko spôsobov realizácií transformačných metód s využitím dát z DPZ ako pomocných informácií. Najvýznamnejšou z tejto kategórie je tzv. dazymetrická metóda. Dazymetrické mapovanie bolo v areálovej interpolácii dát navrhnuté ako alternatíva k metóde priestorového váženía ešte v 30. rokoch 20. st. americkým geografom Wrightom (1936), ktorý sa snažil využiť vedomosti o skúmanej lokalite a identifikovať územia v zónach s rozdielnou hustotou zaľudnenia. Rozvinul tak metódu mapovania hustoty zaľudnenia s využitím topografických listov ako pomocných informácií. V súčasnosti sa pri tejto technike, ako už bolo spomenuté, používajú skôr údaje z diaľkového prieskumu Zeme, pričom sa predpokladá, že rozmiestnenie obyvateľstva sa vzťahuje na jednotlivé triedy krajinskej pokrývky (land cover triedy). Najjednoduchším prístupom v rámci dazymetrického mapovania s využitím land cover dát je prístup binárneho rozdelenia, keď sa jednotlivé triedy krajinskej pokrývky klasifikujú buď ako zaľudnené alebo nezaľudnené (Fisher a Langford 1995, Holt et al. 2004). V triede urbanizovanej zástavby sa teda očakáva vyššia hustota zaľudnenia ako napríklad v poľnohospodárskych alebo lesných typoch tried krajinskej pokrývky. Princípom ostatných spôsobov implementácie tejto metódy je určiť vzťah medzi hustotou zaľudnenia a jednotlivými kategóriami krajinskej pokrývky. Hustotu tried krajinskej pokrývky možno určiť buď empiricky (Eicher a Brewer 2001, Mennis 2003) alebo regresnou analýzou (Langford et al. 1991 a Yuan et al. 1997).

Pomerne nový prístup, ktorý využíva rozvíjajúce sa techniky geografického informačného systému a stále sa zvyšujúce a dostupnejšie množstvo digitálnych databáz, navrhol Xie (1995). Ak totiž uvažujeme o probléme transformácie geografických dát, kde ako premenné vystupujú charakteristiky vzťahujúce sa k obyvateľstvu alebo k bývaniu, je možné ako pomocnú informáciu využiť cestnú sieť. Tento prístup je totiž založený na predpoklade, že hustota zaľudnenia súvisí s hustotou cestnej siete. Xie rozvinul tri sieťové algoritmy, na základe ktorých sa premenná zdrojových zón pridelí jednotlivým cestným úsekom. Ide o metódu dĺžky cestnej siete, hierarchicky váženej cestnej siete a bytovej únosnosti cestnej siete. S metódami odhadu dát, ktoré ako pomocné informácie využívajú cestnú sieť, experimentovali aj Voss et al. (1999) a Reibel a Buffalino (2005). Voss et al. (1999) uvažujú okrem dĺžky cestnej siete aj o počte uzlov v cestnej sieti v samotných zdrojových jednotkách. Ako uzol sa rozumie začiatkový a koncový bod cestného úseku, rovnako aj styčný bod jednotlivých cestných úsekov.

Ani jedna transformačná metóda však neposkytuje presné výsledky, vždy ide len o odhady premenných. Presnosti vybraných transformačných metód sa vo svojej práci venujú napríklad Flowerdew a Green (1989 a 1991), Flowerdew et al. (1991), Goodchild et al. (1993) či Hawley a Moellering (2005), ktorí odhadované hodnoty porovnávajú so známymi hodnotami alebo napr. Langford et al. (1991), ktorí pracujú s ďalšími odhadmi. Okrem týchto spomenieme ešte ďalších autorov, napr. Fisher a Langford (1995), Xie (1995), Cockings et al. (1997), Mrozinski a Cromley (1999), Sadahiro (2000) a Gregory (2002). Vo väčšine prípadov porovnávali výsledky priestorového váženého a pyknofylaktickej alebo dazymetrickej metódy, pričom sa preukázalo, že odhady, ktoré poskytla dazymetrická metóda, sú presnejšie v porovnaní so zvyšnými spomenutými metódami. Jediní, ktorí porovnali a zhodnotili výsledky rôznych metód odhadu dát (priestorové váženie, pyknofylaktická metóda, dazymetrická metóda a sieťové metódy), boli Hawley a Moellering (2005). Najlepšie výsledky spomedzi všetkých testovaných metód vykazovala sieťová metóda a ako sľubná sa javila tiež dazymetrická metóda.

DÁTA

Na testovanie a vzájomné porovnanie výsledkov sme pre potreby nášho výskumu vybrali metódu priestorového váženého ako základnú metódu transformácie geografických dát, dazymetrickú metódu s využitím kategórie 1.1. urbanizovaná zástavba z databázy CORINE landcover 1990 (Slovenská agentúra životného prostredia 2004) a sieťovú metódu s pomocnými informáciami o dĺžke a type cestných úsekov, ako aj o počte uzlov v cestnej sieti (GKÚ 1991-1997).

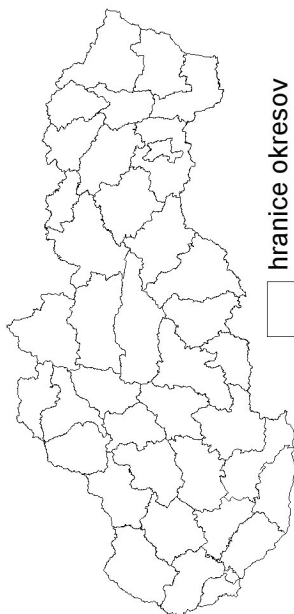
Zdrojové geografické jednotky (obr. 1) predstavovali okresy s hranicami viažucimi sa k roku 1991, z ktorých boli premenné odhadnuté do systému funkčných mestských regiónov FMR 91-B navrhnutých Bezákom (2000). Pre vzájomné porovnanie spoľahlivosti každej metódy sa okrem počtu obyvateľov odhadovala aj ďalšia premenná – rozloha poľnohospodárskej pôdy. Pre systém funkčných mestských regiónov sme sa rozhodli preto, lebo tieto jednotky na jednej strane považujeme za vhodnejší nástroj geografickej analýzy ako samotné okresy, ale predovšetkým preto, lebo sú to jednotky, za ktoré nemáme k dispozícii žiadne údaje, preto ich musíme odhadnúť. Keďže však ide o systém, ktorý vznikol agregáciou z obcí a dáta za obce poznáme, bude možné predikované hodnoty premenných za FMR porovnať so skutočnými hodnotami (teda takými, ktoré si vyskladáme z obcí) a tým vyhodnotiť presnosť jednotlivých transformačných metód.

IMPLEMENTÁCIA A TESTOVANIE JEDNOTLIVÝCH METÓD

Prekryvom zdrojových (38 okresov) a cieľových (66 FMR 91-B) geografických jednotiek v prostredí ArcGIS 9.2 sa získalo 93 zón prieniku. Pre každú zónu prieniku sa vypočítala váha, a to ako podiel rozlohy zóny prieniku a jej korešpondujúcej zdrojovej zóny. Následne sa váha aplikovala na premenné zdrojovej zóny a reagregáciou takto získaných dát sa vypočítal odhad daných premenných pre cieľovú zónu (tab. 1). Pokiaľ ide o dazymetrickú a sieťovú metódu, v prvom prípade sa prekryla vrstva okresov s triedou urbanizovanej zástavby z databázy CORINE landcover, v druhom prípade vrstva okresov a vrstva

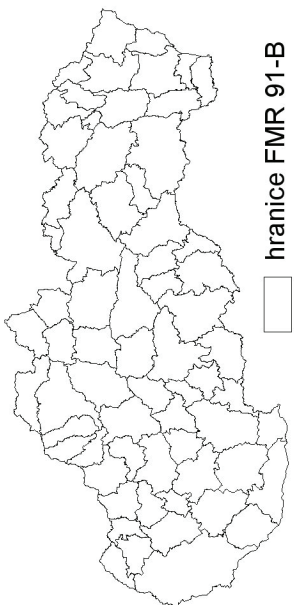
Zdrojové geografické jednotky

Okresy 1991

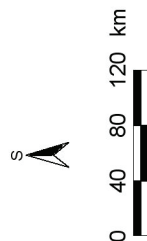
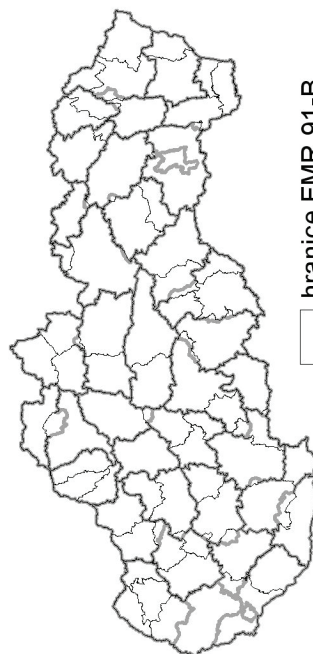


Cieľové geografické jednotky

FMR 91-B



Zóny prieniku



Obr. 1. Zdrojové a cieľové geografické jednotky, zóny prieniku

cestnej siete. Namiesto celkovej rozlohy prienikového územia sa v dazymetrickej metóde pracovalo len s rozlohou urbanizovanej zástavby. V rámci sieťovej metódy sa do úvahy zobralo šesť typov ciest z celkových deväť – diaľnice, cesty 1., 2. a 3. triedy, hlavná a vedľajšia cesta v obci. Poľné cesty, lesné cesty a chodníky neboli zahrnuté. Každému typu cesty bola expertným odhadom priradená váha, ktorá odráža predpoklad hustoty zaľudnenia. Diaľniciam bolo priradených 5 %, cestám 1., 2. a 3. triedy 10 % a hlavným a vedľajším cestám v obci 85 % hodnoty premennej zdrojovej zóny.

Tab. 1. Výpočet odhadu premennej v jednotlivých metódach transformácie geografických dát

Metóda		Váha	Odhad premennej
Priestorové váženie		$w_{zc} = \frac{A_{zc}}{A_z}$	$\hat{y}_c = \sum_z w_{zc} y_z$
Dazymetrická metóda		$w_{zc} = \frac{Z_{zc}}{Z_z}$	$\hat{y}_c = \sum_z w_{zc} y_z$
Sieťová metóda	Dĺžka cestnej siete	$w_{zc} = \frac{D_{zc}}{D_z}$	$\hat{y}_c = \sum_z w_{zc} y_z$
	Typ cestnej siete	$w_{zc} = \frac{\sum_n w_t D t_{zc}}{\sum_n D t_z}$	$\hat{y}_c = \sum_z w_{zc} y_z$
	Počet uzlov	$w_{zc} = \frac{U_{zc}}{U_z}$	$\hat{y}_c = \sum_z w_{zc} y_z$

w – váha, A – rozloha územia, Z – rozloha urbanizovanej zástavby, D – dĺžka ciest, U – počet uzlov, y – skutočná hodnota premennej, – odhadnutá hodnota premennej, t – typ cestnej siete, n – počet typov cestnej siete, indexy z a c sa vzťahujú k zdrojovej a cieľovej zóne

Presnosť a spoľahlivosť každej metódy sa hodnotila dvojakým spôsobom – na základe tzv. mier globálnej vhodnosti a vizuálnou interpretáciou. Mierami globálnej vhodnosti boli, podľa autorov Fisher a Langford (1995), stredná kvadratická chyba (S_1) a tzv. prispôbená stredná kvadratická chyba (S_2), vypočítané podľa vzťahov:

$$S_2 = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2 \right]^{1/2}$$

$$S_1 = \left[\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right]^{1/2},$$

kde n je počet cieľových zón, y je skutočný počet obyvateľov zóny i , $\hat{y}$ je odhadnutý počet obyvateľov zóny i .

S_1 pracuje s absolútnymi rozdielmi medzi skutočným a odhadnutým počtom obyvateľov cieľovej zóny. S_2 je absolútnou chybou prepočítanou vzhľadom na skutočný počet obyvateľov zóny i . K nim boli pridané ešte priemerná absolútna, priemerná relatívna a priemerná algebrická chyba (S_3 , S_4 a S_5):

$$S_3 = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$S_4 = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} (\cdot 100 \%)$$

$$S_5 = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i} (\cdot 100 \%).$$

Zatiaľ čo S_3 je chybou odhadu v jednotkách merania, S_4 je bezrozmerná miera, ktorá sa uvažuje vo vzťahu k skutočnej hodnote premennej príslušnej cieľovej geografickej jednotky. Udáva relatívny počet nesprávne pridelenej hodnoty premennej. S_5 je mierou vychýlenia odhadu. Kladné hodnoty premennej naznačujú, že daná metóda poskytuje v priemere nadhodnotenú odhadu premennej, záporné naopak signalizujú tendenciu k podhodnoteným odhadom (Bezák a Holická 1995). S_1 , S_2 , S_3 , S_4 a S_5 sú teda štatistické miery, ktoré posudzujú odhady premennej súhrnne v rámci celého súboru cieľových geografických jednotiek. Priemerné hodnoty za danú metódu však zastierajú rozdiely medzi chybami individuálnych predpovedí v rámci celého súboru cieľových geografických jednotiek. Preto sa spomínané miery zhody rozšírili o ďalší indikátor hodnotenia účinnosti testovanej metódy – o relatívnu chybu vyjadrenú v percentách, ktorá sa následne opäť použila pre celkové porovnanie jednotlivých transformačných metód prostredníctvom percentuálneho podielu presných odhadov s relatívnou chybou od -5 do +5 % a percentuálneho podielu extrémnych odhadov s relatívnou chybou väčšou ako ± 50 %.

Hodnoty relatívnej chyby vyjadrenej v percentách sa pre každú metódu znázornili ako tematické mapy, ktoré zobrazujú priestorové rozmiestnenie chýb a je z nich možné určiť, ako jednotlivá metóda pracuje. Relatívna percentuálna chyba sa ráta pre každú zónu nasledovným spôsobom:

$$PE_i = \frac{\hat{y}_i - y_i}{y_i} \times 100.$$

VÝSLEDKY A DISKUSIA

Celkové kvantitatívne charakteristiky každej premennej v jednotlivých transformačných metódach zobrazuje tab. 2. Porovnaním súčtov a stredných hodnôt pôvodného súboru dát a výsledných odhadov sa pri všetkých metódach potvrdilo zachovanie tzv. pyknofylaktickej vlastnosti, čo znamená, že sa v procese

transformácie žiadne dáta nestratili, ale naopak, zachovali si svoj rozsah. Ostatné štatistiky napovedajú, že metóda priestorového váženía poskytuje v priemere horšie výsledky ako zvyšné metódy a že vo všeobecnosti možno pozorovať lepšie výsledky pri sieťových metódach ako pri dazymetrickej metóde. Uvedené popisné charakteristiky však poskytujú len akési hrubé zhodnotenie výsledkov jednotlivých transformačných metód. Komplexnejšie tvrdenia sa preto opierajú o ďalšie indikátory hodnotenia účinnosti – miery globálnej vhodnosti (S_1 , S_2 , S_3 , S_4 a S_5) a relatívne chyby vyjadrené v percentách.

Tab. 2. Deskriptívne štatistiky za jednotlivé premenné a metódy

Štatistiky	Počet obyvateľov					
	Skutočné hodnoty	Priestorové váženie	Dazymetrická metóda	Dĺžka ciest	Typ ciest	Počet uzlov
Stredná hodnota	79 914	79 914	79 914	79 914	79 914	79 914
Medián	59 336	54 008	56 393	54 544	54 955	57 939
Smer. odchýlka	91 272	92 534	91 557	91 545	91 048	91 537
Minimum	6 013	9 041	11 959	11 555	12 807	8 720
Maximum	652 590	686 402	667 471	672 214	664 032	674 625
Rozpätie	646 577	677 361	655 513	660 659	651 225	665 906
Súčet	5 274 335	5 274 335	5 274 335	5 274 335	5 274 335	5 274 335

Štatistiky	Rozloha poľnohospodárskej pôdy					
	Skutočné hodnoty	Priestorové váženie	Dazymetrická metóda	Dĺžka ciest	Typ ciest	Počet uzlov
Stredná hodnota	37 302	37 302	37 302	37 302	37 302	37 302
Medián	29 590	31 609	30 668	31 201	30 397	31 291
Smer. odchýlka	24 749	24 478	25 036	24 486	24 706	24 446
Minimum	6 632	5 221	6 240	5 705	6 498	5 631
Maximum	136 687	144 286	133 535	135 795	130 710	136 653
Rozpätie	130 055	139 065	127 295	130 091	124 212	131 022
Súčet	2 461 937	2 461 937	2 461 937	2 461 937	2 461 937	2 461 937

Miery globálnej vhodnosti

Pri komplexnom hodnotení presnosti odhadov jednotlivých transformačných metód pre každú premennú zvlášť sa potvrdilo, že metóda priestorového váženía bola podľa očakávaní najmenej presnou metódou. Pokiaľ ide o počet obyvateľov, najmenej presnou sa ukázala byť metóda priestorového váženía, ktorá nesprávne pridelila v priemere 8 993 obyvateľov, čo predstavovalo 19 % počtu obyvateľov. Za ňou nasledovala metóda počtu uzlov, dĺžky cestnej siete a dazymetrická metóda (tab. 3). Porovnaním jednotlivých metód sa preukázalo, že najmenšie chyby dosiahla metóda typov ciest – nesprávne pridelila 4 565 osôb, t. j. 11 % celkového počtu obyvateľov, čo predstavuje oproti priestorovému váženiu až 49 % zlepšenie v prípade absolútneho počtu a 39 % zlepšenie, ak ide o podiel

nesprávne prideleného počtu obyvateľov. Nezhoda v percentuálnych podieloch zlepšenia odhadov jednotlivých metód v porovnaní s metódou priestorového váženia pri použití chýb v absolútnych a relatívnych počtoch je spôsobená variabilitou počtu obyvateľov v jednotlivých geografických jednotkách, čím sú spôsobené aj odchýlky v presnosti jednotlivých transformačných metód.

Tab. 3. Miery globálnej vhodnosti

	Počet obyvateľov				
	PV	DM	DC	TC	PU
S ₁	13 145,41	6 956,48	8 150,94	6 259,06	10 571,27
S ₂	0,31	0,25	0,25	0,24	0,28
S ₃	8 992,57	5 397,60	6 131,71	4 564,65	7 900,87
S ₄	18,98	13,43	14,21	11,48	17,33
S ₅	-8,62	-5,68	-6,91	-6,02	-7,45
Zlepšenie oproti PV (abs.)	–	29,25	25,12	39,52	8,74
Zlepšenie oproti PV (rel.)	–	39,98	31,81	49,24	12,14

	Rozloha poľnohospodárskej pôdy				
	PV	DM	DC	TC	PU
S ₁	3 609,85	2 839,15	2 570,47	3 083,49	3 518,07
S ₂	0,20	0,13	0,12	0,13	0,19
S ₃	2 584,52	2 187,86	1 965,28	2 354,77	2 739,38
S ₄	10,87	8,76	8,26	9,45	12,08
S ₅	-2,55	-0,21	-1,30	-0,78	-1,88
Zlepšenie oproti PV (abs.)	–	19,46	24,05	13,11	11,07
Zlepšenie oproti PV (rel.)	–	15,35	23,96	8,89	5,99

PV – priestorové váženie, DM – dazymetrická metóda, DC – metóda dĺžky ciest, TC – metóda typov ciest, PU – metóda počtu uzlov

Mierne odlišná situácia sa vyskytla pri odhadoch druhej premennej, rozlohy poľnohospodárskej pôdy, ktorej odhady nevykazujú až také veľké rozdiely medzi jednotlivými metódami. V tomto prípade sa ako najmenej presná javí metóda počtu uzlov (2 739 nesprávne priradených hektárov, čo predstavuje 12 % celkovej rozlohy poľnohospodárskej pôdy), nasleduje metóda priestorového váženia (chybne odhadnutých 2 584 ha, resp. 11 %). Najpresnejšou bola metóda dĺžky ciest, ktorá nesprávne prideliла 1 965 ha (8 %), tesne nasledovaná dazymetrickou metódou a metódou typov ciest.

Máme teda možnosť vidieť, že odhady jednej premennej môžu pri danej metóde prinášať väčšie nepresnosti ako odhady druhej premennej. Pre obe premenné sú chyby v ostatných metódach omnoho konzistentnejšie ako v prípade priestorového váženia. Napríklad počet obyvateľov sa v metóde priestorového váženia odhadne v porovnaní s ostatnými metódami s omnoho nepresnejším výsledkom ako rozloha poľnohospodárskej pôdy. Táto variabilita odhadov pri metóde

priestorového váženía vyjadruje extrémnu citlivosť na vnútorné zmeny v hustote zaľudnenia a narúša tak predpoklad homogénneho rozmiestnenia premennej, na ktorú sa táto metóda spolieha (Eicher a Brewer 2001). Práve táto nestálosť môže spôsobovať takéto odchýlky v chybách plynúcich z metódy priestorového váženía, tak ako to naznačujú rozdiely medzi počtom obyvateľov a rozlohou poľnohospodárskej pôdy.

Pokiaľ ide o tzv. globálne miery vhodnosti, S_1 a S_2 , výsledky sú zhodné s predchádzajúcimi tvrdeniami, čo znamená, že ako najvhodnejšia sa javí dazymetrická metóda, ďalej metóda typov a dĺžky ciest a ako najmenej presná sa ukazuje byť metóda počtu uzlov a priestorového váženía. Poradie jednotlivých metód sa líši v závislosti od odhadovanej premennej, v podstate však za najstabilnejšiu metódu, ktorá poskytuje najpresnejšie odhady oboch sledovaných premenných, možno považovať dazymetrickú metódu.

Ako už bolo naznačené a potvrdené aj pri chybe S_1 , pomerne značné rozdiely medzi chybami jednotlivých metód vidieť obzvlášť pri prvej premennej. V prípade odhadov rozlohy poľnohospodárskej pôdy nie sú až také markantné. Pokiaľ ide o S_2 , tam sa rozdiely medzi jednotlivými metódami mierne stierajú, v podstate však jej hodnoty zodpovedajú už spomenutým hodnoteniam.

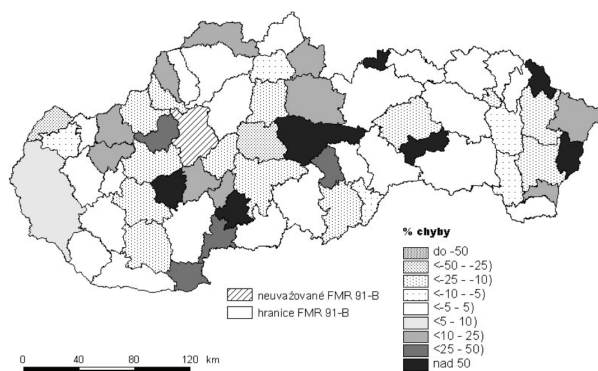
Priestorová vizualizácia chýb

Ešte k zaujímavejším zisteniam dospejeme na základe porovnania máp relatívnych chýb vyjadrených v percentách zostrojených pre obe premenné a pre každú z metód (obr. 2 až 11), keďže v tomto prípade posudzujeme jednotlivé individuálne odhady. Vo všetkých mapách je na porovnanie zvolená rovnaká klasifikácia chýb. Negatívne percentuálne chyby (podhodnotené premenné oproti skutočným hodnotám) sú znázornené rastrom bodov rôznej intenzity, pozitívne percentuálne chyby (nadhodnotené premenné oproti skutočným hodnotám) sú naopak vyjadrené rôznymi odtieňmi sivej farby. Stredná (biela) trieda rozmiestnenia reprezentuje cieľové zóny s relatívnymi percentuálnymi chybami odhadu menšími ako ± 5 % oproti skutočným hodnotám.

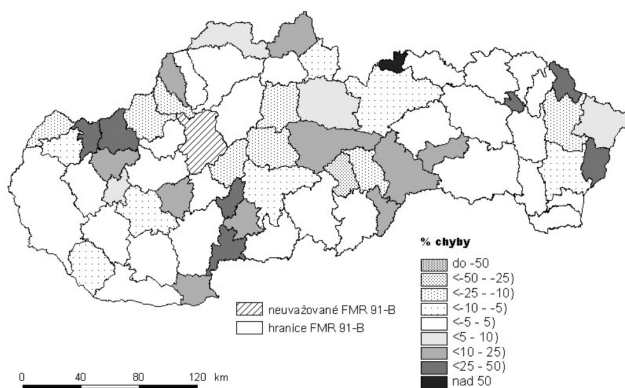
Z priestorovej vizualizácie chýb vyplynuli viaceré poznatky. S určitosťou možno skonštatovať, že ak sú hranice zdrojových a cieľových geografických jednotiek geometricky (rozlohou) podobné, interpolovaná premenná sa odhadne s minimálnymi chybami.

Ak sú zdrojové jednotky oveľa väčšie ako cieľové, záujmová premenná sa odhadne s horším výsledkom, ako by tomu bolo v opačnom prípade. To však platí len pre metódu priestorového váženía. V prípade využitia pomocných informácií v procese transformácie dát sa dá tento fakt výrazným spôsobom zmierniť v závislosti od toho, či a do akej miery koreluje pomocná informácia s odhadovanou premennou.

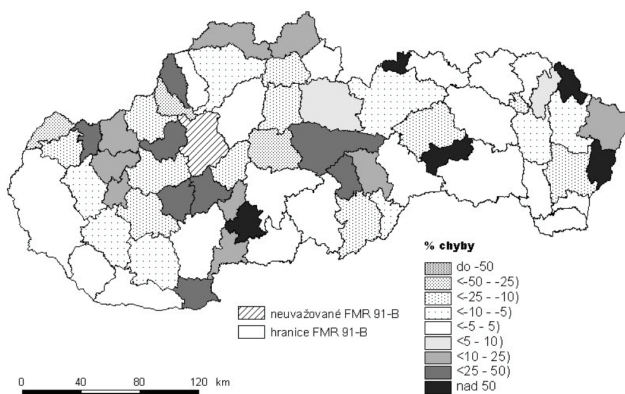
Potvrdilo sa, že metódy, ktoré využívajú pomocné informácie, sú v priemere schopné redukovať chyby odhadov a poskytnúť tak lepšie výsledky. Pri odhadoch počtu obyvateľov je obzvlášť viditeľná redukcia chýb pri metóde typov ciest (obr. 2e), ale aj pri dazymetrickej metóde (30,3 % presných a 31,8 % extrémnych odhadov pri priestorovom vážení; 45,45 % a 27,27 % pri dazymetrickej metóde; 63,64 % a 24,24 % pri metóde typov ciest). Pri odhade rozlohy poľ-



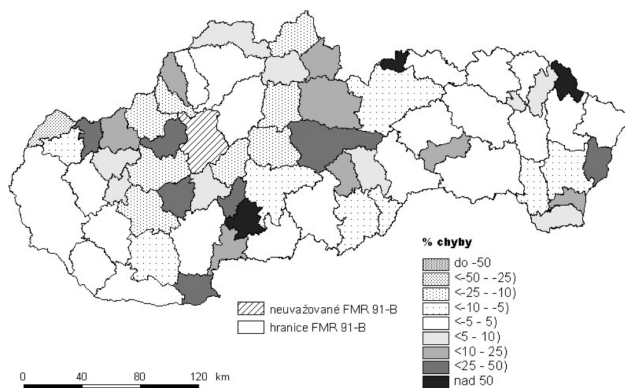
Obr. 2. Odhad počtu obyvateľov FMR 91-B k roku 1991 – percentuálne chyby (váha = rozloha územia)



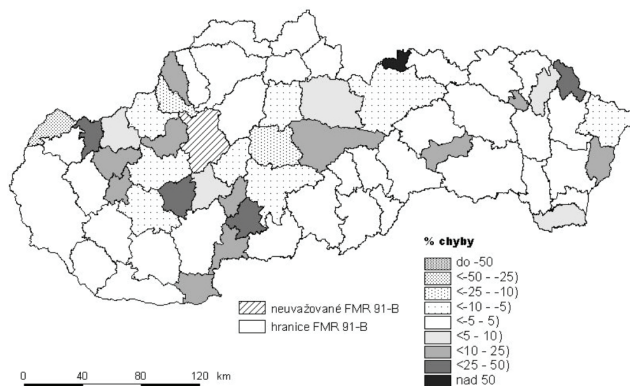
Obr. 3. Odhad počtu obyvateľov FMR 91-B k roku 1991 – percentuálne chyby (podľa kategórie 1.1. z CORINE land cover)



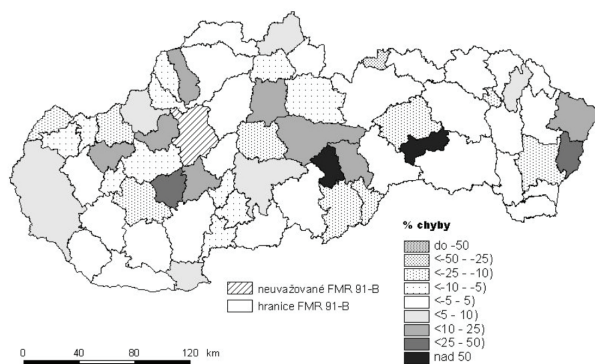
Obr. 4. Odhad počtu obyvateľov FMR 91-B k roku 1991 – percentuálne chyby (podľa počtu uzlov v cestnej sieti)



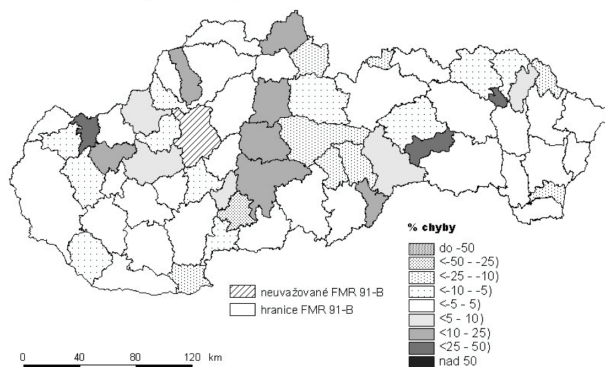
Obr. 5. Odhad počtu obyvateľov FMR 91-B k roku 1991 – percentuálne chyby (podľa dĺžky cestnej siete)



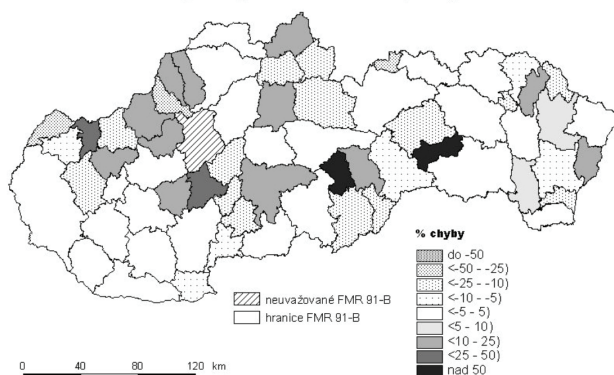
Obr. 6. Odhad počtu obyvateľov FMR 91-B k roku 1991 – percentuálne chyby (podľa typu cestnej siete)



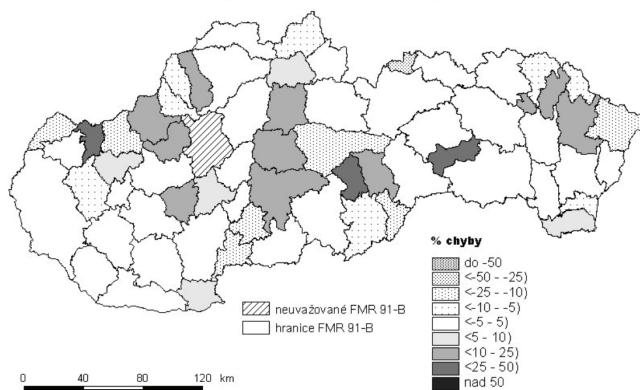
Obr. 7. Odhad rozlohy poľnohospodárskej pôdy FMR 91-B k roku 1991 – percentuálne chyby (váha = rozloha územia)



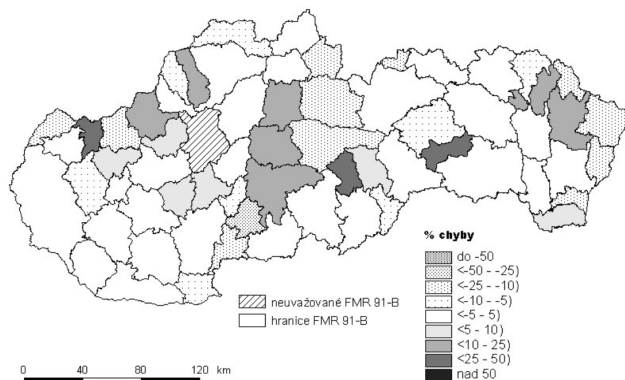
Obr. 8. Odhad rozlohy poľnohospodárskej pôdy FMR 91-B k roku 1991 – percentuálne chyby (podľa kategórie I.1. z CORINE land cover)



Obr. 9. Odhad rozlohy poľnohospodárskej pôdy FMR 91-B k roku 1991 – percentuálne chyby (podľa počtu uzlov v cestnej sieti)



Obr. 10. Odhad rozlohy poľnohospodárskej pôdy FMR 91-B k roku 1991 – percentuálne chyby (podľa dĺžky cestnej siete)



Obr. 11. Odhad rozlohy poľnohospodárskej pôdy FMR 91-B k roku 1991 – percentuálne chyby (podľa typu cestnej siete)

nohospodárskej pôdy je situácia obdobná, ako najpresnejšia sa v rámci celého súboru cieľových geografických jednotiek javí dazymetrická metóda (obr. 3b), tesne nasledovaná metódou typov ciest a dĺžky ciest (63,64 % presných a 15,15 % extrémnych odhadov pri dazymetrickej metóde; 59,09 % a 21,21 % pri metóde dĺžky ciest; 56,06 a 16,67 % pri metóde typov ciest).

Pri zohľadnení jednotlivých individuálnych prípadov však musíme konštatovať, že metóda priestorového váženía neposkytuje vždy horšie výsledky ako ostatné metódy. Existujú prípady, keď práve priestorové váženía poskytuje lepšie výsledky ako zvyšné metódy. Táto situácia je obzvlášť zreteľná pri odhade poľnohospodárskej pôdy.

Aj napriek uvedeným skutočnostiam ostáva niekoľko cieľových zón, ktoré stále produkujú pomerne veľké chyby. Ide predovšetkým o také cieľové geografické jednotky, ktoré sú oveľa menšie ako zdrojové geografické jednotky (FMR Giraltovce, FMR Spišská Stará Ves). Treba však podotknúť, že hoci sú chyby odhadu v týchto zónach vysoké, v princípe sú stále nižšie ako v prípade metódy priestorového váženía (až na jednu výnimku – pri odhade počtu obyvateľov FMR Spišská Stará Ves).

Okrem týchto skutočností je dôležité porozumieť aj nadhodnoteniam a podhodnoteniam jednotlivých premenných. Nadhodnotené a podhodnotené územia si tento fakt nie vždy zachovávajú pri všetkých použitých metódach. V miestach, kde zdrojová zóna úplne obsiahne dve alebo viac cieľových zón, máme možnosť sledovať na všetkých mapách nasledovné: ak jeden okres (napr. Banská Bystrica) obsahuje dva alebo viac FMR 91-B (FMR Banská Bystrica a Brezno), tak keď je premenná jednej z cieľových zón nadhodnotená (FMR Banská Bystrica), druhej je podhodnotená (FMR Brezno). Tento fakt je obzvlášť zreteľný pri metóde priestorového váženía, ale zachováva sa pri všetkých použitých metódach a mohol by naznačovať negatívnu priestorovú autokoreláciu v odhadnutých hodnotách. Žiaľ však nie je jasné, čo všetko vplýva na nadhodnotenie a podhodnotenie príslušných premenných, ale zdá sa, že rozloha cieľovej zóny v rámci zdrojovej zóny to nespôsobuje.

ZÁVER

Predkladaný príspevok hodnotí presnosť a spoľahlivosť vybraných transformačných metód použitých na odhad dvoch premenných odlišného charakteru (počet obyvateľov a rozloha poľnohospodárskej pôdy). Výsledky najjednoduchšej transformačnej metódy – priestorového váženía, boli porovnané so zložitejšími metódami, založenými na využití pomocných informácií, o ktorých sa predpokladá, že dokážu skvalitniť samotný proces odhadu dát a poskytnúť tak lepšie výsledky. Zhodnotenie presnosti jednotlivých metód prebiehalo v dvoch rovinách – všeobecnej a individuálnej. Na jednej strane sa posudzovala spoľahlivosť metód v rámci celého súboru cieľových geografických jednotiek a na druhej strane sa brali do úvahy aj jednotlivé individuálne odhady (odhady za každú jednu cieľovú geografickú jednotku), zvlášť za každú premennú. Z testovania vyplynulo, že hoci sa metóda priestorového váženía v podstate javí pri oboch premenných ako najmenej presná, pri posúdení jednotlivých individuálnych prípadov to nie vždy platí. Metóda priestorového váženía vykazuje lepšie výsledky v územiach, kde je premenná rovnomerne rozmiestnená. Výsledky transformačného procesu dokážu výrazným spôsobom vylepšiť metódy, ktoré využívajú doplnkové informácie, ale len do takej miery, ako koreluje odhadovaná premenná s pomocnou premennou použitou na zlepšenie výsledkov odhadov. Ďalej možno skonštatovať, že niektoré cieľové zóny dosahujú konzistentné výsledky pri všetkých metódach, kým v iných sa menia od metódy k metóde. Rovnako možno povedať, že určitá metóda sa môže pri odhadoch jednej premennej javiť ako menej presná než pri odhadoch druhej premennej. Tento fakt sa dá opäť vysvetliť koreláciou odhadovanej a pomocnej premennej. Celkovo sa však zdá, že všetky metódy vykazujú lepšie výsledky v regiónoch s väčšími hodnotami odhadovanej premennej.

Na formuláciu komplexnejších zistení, na základe ktorých by bolo možné formulovať akési zovšeobecnenia týkajúce sa kvality odhadov v procese transformácie geografických dát, je potrebná ešte podrobnejšia analýza so zapojením viacerých premenných do tohto procesu. V blízkej budúcnosti je plánované testovanie spoľahlivosti jednotlivých transformačných metód aj na inom súbore geografických jednotiek, čo, ako veríme, povedie k získaniu ďalších poznatkov.

Tento príspevok vznikol v rámci riešenia vedeckého projektu č. 1/0181/09 Súčasný procesy redistribúcie obyvateľstva na Slovensku, ktorý bol financovaný grantovou agentúrou VEGA.

LITERATÚRA

- BEZÁK, A., HOLICKÁ, A. (1995). Komparatívna analýza extrapoláčnych modelov pre regionálne populačné projekcie. *Geografický časopis*, 47, 233-246.
- BEZÁK, A. (2000). *Funkčné mestské regióny na Slovensku*. Geographia Slovaca, 15. Bratislava (Geografický ústav SAV).
- COCKINGS, S., FISHER, P., LANGFORD, M. (1997). Parameterization and visualization of the errors in areal interpolation. *Geographical Analysis*, 29, 314-328.
- EICHER, C. L., BREWER, C. A. (2001). Dasymetric mapping and areal interpolation: implementation a evaluation. *Cartography and Geographic Information Science*, 28, 125-138.

- FISHER, P., LANGFORD, M. (1995). Modelling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, 27, 211-224.
- FLOWERDEW, R., GREEN, M. (1989). Statistical methods for inference between incompatible zonal systems. In Goodchild, M., Gopal, S., eds. *Accuracy of spatial databases*. London (Taylor & Francis), pp. 239-247.
- FLOWERDEW, R., GREEN, M. (1991). Data integration: statistical methods for transferring data between zonal systems. In Masser, I., Blakemore, M. B., eds. *Haling geographical information*. Essex (Longman Scientific & Technical), pp. 38-53.
- FLOWERDEW, R., GREEN, M., KEHRIS, E. (1991). Using areal interpolation methods in GIS. Papers in Regional Science. *The Journal of the RSAI*, 70, 303-315.
- 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 no. 4. Newcastle Upon Tyne (University of Newcastle Upon Tyne).
- GKÚ (1991-1997). *Spojité vektorová mapa Slovenska, 1:50 000, vrstva cesty*. Bratislava (Geodetický a kartografický ústav).
- 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, 383-397.
- GREGORY, I. 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, 293-314.
- HAWLEY, K., MOELLERING, H. (2005). A comparative analysis of areal interpolation methods. *Cartography and Geographic Information Science*, 32, 411-423.
- 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*, 31, 103-121.
- LAM, N. (1983). Spatial interpolation methods: a review. *American Cartographer*, 10, 129-149.
- LANGFORD, M., MAGUIRE, D. J., UNWIN, D. J. (1991). The areal interpolation problem: estimating population using remote sensing in a GIS framework. In Masser, I., Blakemore, M. B., eds. *Haling geographical information*. Essex (Longman Scientific & Technical), pp. 55-77.
- MADAJOVÁ, M. (2010). Metódy transformácie priestorových dát: prehľad, klasifikácia a hodnotenie. *Acta Geographica Universitatis Comenianae*, 54, v tlači.
- MENNIS, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55, 31-42.
- 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.
- REIBEL, M., BUFALINO, M. E. (2005). Street weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A*, 37, 127-129.
- SADAHIRO, Y. (2000). Accuracy of count data estimated by the point-in-poly method. *Geographical Analysis*, 32, 64-89.
- SAŽP (1996). *GIS vrstva CORINE landcover*. Banská Bystrica (Slovenská agentúra životného prostredia).
- SAŽP (2004). *GIS vrstva CORINE landcover*. Banská Bystrica (Slovenská agentúra životného prostredia).

- TOBLER, W. (1979). Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association*, 74, 519-530.
- VOSS, P., LONG, D. D., HAMMER, R. B. (1999). *When census geography doesn't work: using auxiliary information to improve the spatial interpolation of demographic data*. Center for Demography and Ecology Working Paper. Madison (University of Wisconsin).
- WRIGHT, J. K. (1936). A method of mapping densities of population with Cape Code as an example. *The Geographical Review*, 26, 103-110.
- 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.

Michala M a d a j o v á

USE OF AUXILIARY INFORMATION IN TRANSFORMATION OF GEOGRAPHICAL DATA: ASSESSMENT OF SELECTED PROCEDURES

This paper is reported as a comparative analysis of various geographical data transformation methods which allow data allocation from spatially inconsistent zonal systems. The basic and simplest transformation method – areal weighting – is compared with difficult techniques which use auxiliary information to improve the results of the interpolation process. These methods were the dasymetric method which uses urban fabric class from the CORINE landcover classification and three network methods – network length method, network hierarchic method and node counts method. For comparison, different data types were used. Apart from the usually estimated variable – population counts – it was the area of agricultural land. These variables have been available for 1991 districts and estimated for the system of functional urban regions (FUR 91-B). The measures of global fit and spatial visualization of percentage errors associated with all variables and methods were used to evaluate the results of the estimation. The results are as follows: No method is truly satisfactory in the sense that it is impossible to derive perfect results. Concerning RMSE and adj-RMSE the dasymetric, network length and hierarchial methods provided the most accurate results for each variable. The spatial errors show similar results. The areal weighting method seems to be the most inaccurate methods but this is not always true. The maps of percentage errors shows that there are more areas in which areal weighting performs better than the methods which use auxiliary information. We suppose that the correlation between auxiliary information and estimated variables is the reason. In general, the methods which use auxiliary information are able to make more accurate estimations, but there are still areas which show out relatively bad results. Moreover, one method can perform worse by estimation of one variable than by another. An important matter is also to understand the over and underestimation of variables. What we can definitely say, when the source geographical unit obtains two target units, the negative spatial autocorrelation is observed. To evaluate and point out more general patterns requires deeper and more complex analysis in the future.

