

Testovanie metód cestnej siete v procese transformácie geografických dát

Michala MADAJOVÁ

Abstract: *The paper evaluates the accuracy of population estimates using road network transformation methods (the road length method, road-weighting method and node counts method) which belong to so-called intelligent transformation methods and in process of data reallocation from one set of geographical units onto another (from districts to functional urban regions) indicate better results compared with the simplest transformation method – areal weighting.*

Keywords: *areal weighting, variable estimate, length and type of the road, node, transformation*

Úvod

Metódy cestnej siete zaraďujeme v rámci procesu transformácie geografických dát medzi tzv. inteligentné transformačné metódy, ktoré umožňujú odhadnúť dáta z jedného systému územných jednotiek do druhého, nekompatibilného zonálneho systému za účasti pomocných informácií o cestnej sieti dostupných za sadu tzv. zdrojových geografických jednotiek. Termín zdrojové a cieľové geografické jednotky zavádzajú Markoff a Shapiro (1973) a v prvom prípade sa vzťahujú na skupinu zón, za ktoré máme k dispozícii určité údaje, v druhom prípade na územné jednotky, za ktoré tieto údaje potrebujeme. V zmysle delenia transformačných metód podľa formy a obsahu (Madaiová, 2010a) ich zaraďujeme do skupiny tzv. priestorovo-interpoláčnych a obsah zachovávajúcich metód odhadu dát. V nasledujúcich častiach príspevku sa pokúsime porovnať výsledky a vyhodnotiť presnosť odhadov vybranej premennej (počtu obyvateľov) získaných aplikáciou troch typov metód cestnej siete a základnej transformačnej metódy – priestorového váhania.

Základná charakteristika testovaných metód

Za štandardnú a zároveň najjednoduchšiu metódu transformácie geografických dát z navzájom odlišných zonálnych systémov sa považuje metóda prekryvu alebo priestorového váhania. Túto techniku ako prví uviedli Markoff a Shapiro (1973), detailnejšie ju však rozpracovávajú Goodchild a Lam (1980). Metóda je založená na predpoklade, že údaje sú v zdrojových zónach rozmiestnené rovnomerne, takže dáta platné pre sadu zdrojových geografických jednotiek sa úmerne pridelia sade cieľových geografických jednotiek. Hodnoty premennej pre cieľovú zónu sa odhadnú na základe známych hodnôt premennej v zdrojovej zóne a z tzv. váhy, ktorá udáva časť zdrojovej zóny, ktorá leží v cieľovej zóne. Premenné sa teda najprv pridelia korešpondujúcim zónam prieniku, a to ako funkcia časti územia zdrojovej zóny, ktorá je obsiahnutá v zóne prieniku. Zóny prieniku sa následne reagregujú do cieľových geografických jednotiek.

Pre dáta humánno-geografickej povahy je však predpoklad homogenity vnútri zdrojových zón nerealistický. Niektorí geografi sa preto pokúsili zmierniť predpoklad homogenity použitím iných typov metód, a to takých, ktoré sa snažia narušiť predpoklad rovnomerného rozmiestnenia premenných v zdrojových zónach tým, že berú do úvahy akési pomocné informácie o rozmiestnení premenných vnútri zdrojových zón. Zapojenie takýchto doplnkových informácií do procesu transformácie geografických dát zabezpečuje lepšie výsledky a celkovo napomáha k skvalitneniu transformačného procesu. Metódy, ktoré využívajú takéto pomocné informácie sa potom zvyknú označovať aj ako „inteligentné“ interpolačné metódy (Flowerdew, 1988; Flowerdew a Green, 1989, 1991, 1992 a 1994; Flowerdew et al., 1991; Langford et al., 1991; Fisher a Langford, 1995; Mennis a Hultgreen, 2005; Mennis a Hultgreen, 2006a a 2006b). V rámci delenia transformačných metód podľa Madaiová (2010a) ich označujeme aj ako zložené metódy transformácie geografických dát a zaraďujeme k nim aj metódy cestnej siete.

Metódy cestnej siete sa v rámci procesu areálovej transformácie po prvý krát aplikovali v 90-tych rokoch 20. st. v období rozvíjajúcich sa techník geografického informačného systému a stále sa zvyšujúceho a dostupnejšieho množstva digitálnych databáz. Xie (1995) rozvinul 3 tzv. „prekrývajúce sa sieťové algoritmy“ („overlaid network algorithms“). Ide o metódu dĺžky cestnej siete, hierarchicky váženej cestnej siete a bytovej únosnosti cestnej siete (priradenie premennej jednotlivým cestným úsekom na zá-

klade dĺžky, tried a bytovej únosnosti cestných úsekov). Všetky sú založené na využití GIS techník a relevantných, k odhadovanej premennej priamo sa vzťahujúcich pomocných informácií. V GIS terminológii je sieť charakterizovaná ako konkrétne špecifikovaná údajová štruktúra čiarových prvkov na báze teórie grafov tvorená hranami (arcs) a uzlami (nods). Xie (1995) hovorí o cestnej sieti ako o sade jednodimenziálnych objektov a o zdrojových a cieľových geografických jednotkách ako o dvojdimenziálnych objektoch – polygónov. Termín „sieť“ („network“) v tomto prípade znamená, že prerozdelenie atribútových hodnôt premennej zdrojovej zóny je založené na cestných segmentoch. Prenos dát zo zdrojových do cieľových geografických jednotiek sa uskutoční na základe:

- prekryvu zdrojových a cieľových zón, čím vznikne tzv. prieniková zóna, ktorá obsahuje prepojenia – identifikátory na obe sady geografických jednotiek.
- prekryvom prienikových zón a vrstvy cestnej siete (prostredníctvom váh vypočítaných na základe dĺžky a počtu uzlov v cestnej sieti) nastáva transformácia údajov do cieľových zón, a to nasledovne:

Metóda dĺžky cestnej siete je najjednoduchšou metódou cestnej siete, ktorá predpokladá homogénne rozmiestnenie premennej pozdĺž cestnej siete. Premenná cieľovej zóny sa odhadne na základe sčítania dĺžky ciest v zónach prieniku, ktoré jej prislúchajú a podielu hodnoty danej premennej vzťahnutej na jednotku dĺžky ciest korešpondujúcich zdrojových zón.

Metóda hierarchicky vázenej cestnej siete (alebo jednoduchšie metóda typov ciest) berie do úvahy jednotlivé triedy cestných úsekov. Pri tejto metóde sa predpokladá, že obyvateľstvo (alebo k nemu sa vzťahujúca premenná) je pri rôznych typoch ciest rozmiestnené odlišne. Napríklad pri diaľniciach sa dá očakávať menšia hustota zaľudnenia, ako pri mestských uliciach. Metóda je založená na rovnakom princípe ako metóda dĺžky ciest, len s tým rozdielom, že v tomto prípade sa pre zdrojové zóny vytvorí matica váh, ktorá každému typu ciest prideli váhu, ktorej hodnota je tým väčšia, čím viac sa predpokladá, že daný typ ciest je hustejšie zaľudnený. Suma váh v každej zdrojovej zóne sa musí rovnať 1, váhy sú určené subjektívne (na základe vedomostí o danom území, terénnym výskumom, ale podľa priemernej hodnoty za nejaké väčšie referenčné územie ...). O tejto metóde sa teda predpokladá, že zohľadnenie jednotlivých druhov cestných úsekov vylepší výsledky predchádzajúcej metódy.

Množstvo dostupných digitálnych informácií neskôr prispelo k rozvoju ďalšieho prístupu – **metódy bytovej únosnosti cestnej siete**, ktorej sa však v rámci tohto príspevku nebudeme bližšie venovať, len spomenieme, že táto metóda využíva informácie o počte domácností v každej zdrojovej zóne a digitálne dáta, ktoré zahŕňajú rozsahy jednotlivých adries, z ktorých je možné určiť, koľko domov je lokalizovaných pozdĺž jednotlivých cestných úsekov. Metóda je potom založená na princípe, že počty obyvateľov sú vzťahnuté k domom a domy môžu byť pripojené k častiam ulice podľa rozsahu adresy v každej zóne.

Voss et al. (1999) doplnili tieto tri algoritmy o techniku, ktorá uvažuje o počte uzlov v cestnej sieti. Túto budeme označovať ako **metódu počtu uzlov**. Uzly sa chápu ako geometricko-topologické priestorové objekty, ktoré existujú v cestnej sieti na začiatku a konci každého cestného úseku, rovnako aj v každom prieniku s ostatnými cestnými úsekmi. Premenná je cieľovým zónam pridelená na základe časti celkového počtu uzlov zdrojovej zóny, ktoré sú lokalizované v každej zóne prieniku.

Dáta a spôsob testovania jednotlivých transformačných metód

Vybraná premenná – počet obyvateľov – sa odhadovala z okresov vzťahujúcich sa k roku 2001 (zdrojové geografické jednotky) do systému funkčných mestských regiónov – FMR 91-B (cieľové geografické jednotky) navrhnutých Bezákom (2000). Keďže systém FMR 91-B vznikol agregáciou z obcí a dáta za obce, bolo možné predikované hodnoty počtu obyvateľov za FMR porovnať so skutočnými hodnotami (teda takými, ktoré sa dajú vyskladať z obcí), a tým vyhodnotiť presnosť jednotlivých transformačných metód.

Rozloha zdrojových ako aj prienikových území potrebná pri metóde priestorového váženia sa vypočítala v prostredí ArcGIS 9.2. z vrstiev *Hranice obce* podľa rokov 1991 a 2001 (GKÚ Bratislava). Čo sa týka pomocných informácií v rámci troch typov testovaných metód cestnej siete, pracovalo sa s vrstvou *Cesty* vzťahujúcou sa k obdobiu rokov 1991 až 1997, ktorá rozlišuje 2 triedy prvkov – línie a uzly (Digitálne mapové dielo SVM 50, GKÚ Bratislava). Vrstva línie obsahuje 9 typov ciest, z ktorých sme v rámci metód cestnej siete použili triedy *diaľnica*, *cesta 1. triedy*, *cesta 2. triedy*, *cesta 3. triedy*, *hlavná cesta v obci* a *vedľajšia v obci*. V rámci metód typov ciest sa vytvorila jednoduchá hierarchia, na základe ktorej sa každému typu cesty empirickým odhadom prisúdila váha, ktorá odrážala predpoklad hustoty

zaľudnenia jednotlivých cestných úsekov. Diaľniciam bolo pridelených 5 %, cestám 1., 2. a 3. triedy 10 % a hlavným a vedľajším cestám v obci 85 % hodnoty premennej zdrojovej zóny. Aplikované metódy a ich výsledky sa hodnotili a porovnávali prostredníctvom relatívnych chýb vyjadrených v percentách vypočítaných podľa vzťahu

$$S = \frac{\hat{y}_i - y_i}{y_i} * 100 ,$$

ktorých hodnoty sa pre odhadovanú $\hat{y}_i$ premennú - (y_i = skutočná hodnota premennej) a každú aplikovanú metódu znázornili ako tematické mapy, ktoré zobrazujú priestorové rozmiestnenie chýb v každej cieľovej geografickej jednotke, čo umožňuje ľahšie určiť, ako sú jednotlivé metódy hodnoverné.

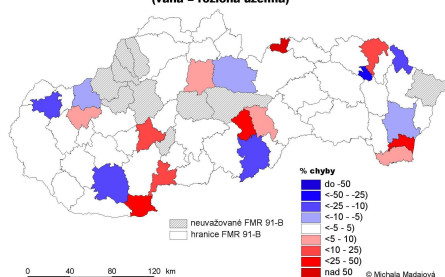
Výsledky a vyhodnotenie

Z máp relatívnych chýb na Obr. 1 vyjadrených v percentách vyplynulo niekoľko zaujímavých skutočností. V priemere za všetky metódy patrili medzi regióny s najmenej presnými odhadmi počtu obyvateľov FMR Spišská Stará Ves, Štúrovo, Šahy, Hnúšťa, Veľké Kapušany, Nová Baňa a Žarnovica, Giraltove, Medzilaborce, Senica, Nové Zámky či Rimavská Sobota. Možno teda skonštatovať, že ak sú hranice zdrojových a cieľových geografických jednotiek geometricky (rozlohou) podobné, premenná sa odhadne s minimálnymi chybami. Pri metóde priestorového váženía platí, že pokiaľ sú zdrojové jednotky oveľa väčšie ako cieľové, záujmová premenná sa odhadne s horším výsledkom, ako by tomu bolo v opačnom prípade. V prípade metód cestnej siete sa dá tento fakt výrazným spôsobom zmierniť v závislosti od toho, ako pomocná premenná odzrkadľuje rozmiestnenie odhadovanej premennej, inými slovami, či a do akej miery koreluje pomocná informácia s odhadovanou premennou.

Došlo teda k potvrdeniu počiatočného predpokladu, a síce, že metódy, ktoré využívajú pomocné informácie, sú schopné vylepšiť výsledky transformačného procesu a poskytnúť lepšie výsledky. Pri odhadoch počtu obyvateľov je obzvlášť viditeľná redukcia chýb pri metóde typov ciest a počtu uzlov (pozri Obr. 1). Pri posúdení jednotlivých individuálnych prípadov (odhadov za každú cieľovú geografickú jednotku zvlášť) však máme možnosť vidieť, že metóda priestorového váženía nie je vždy tou najhoršou metódou. Existujú dokonca prípady, kedy sa z aplikovaných metód zaraďuje medzi tie najpresnejšie metódy (napríklad vo FMR Lučenec, Medzilaborce, Poprad, Žiar nad Hronom, Myjava či Námestovo). Je to buď preto, lebo v týchto územiach je obyvateľstvo relatívne homogénne rozmiestnené, alebo preto, že rozdiel vo veľkosti zdrojových a cieľových geografických jednotiek nie je až taký veľký, prípadne cieľová zóna takmer úplne obsiahne dve alebo viac zdrojových geografických jednotiek. Napriek uvedeným skutočnostiam však stále existujú regióny, ktoré produkovali pomerne veľké chyby. Ide predovšetkým o také cieľové geografické jednotky, ktoré sú oveľa menšie ako zdrojové geografické jednotky (FMR Giraltove, Hnúšťa, Šahy, Spišská Stará Ves, Štúrovo).

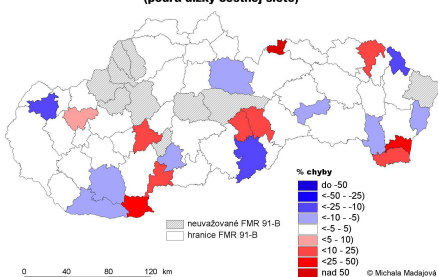
Treba však podotknúť, že hoci sú chyby odhadu v týchto zónach vysoké (až na FMR Giraltove pri metóde dĺžky ciest a FMR Hnúšťa pri metóde typov ciest, kedy sa pohybujú v rámci tolerovateľných chýb, a teda chýb v rozmedzí -5 – 5 %), v princípe sú stále nižšie ako v prípade metódy priestorového váženía. Výnimku však predstavuje FMR Spišská Stará Ves, za ktorý sa odhady počtu obyvateľov pri všetkých metódach pohybovali nad 100 %, pričom najlepšie boli pri metóde priestorového váženía

Odhad počtu obyvateľov FMR 91-B k roku 2001 - percentuálne chyby
(váha = rozloha územia)



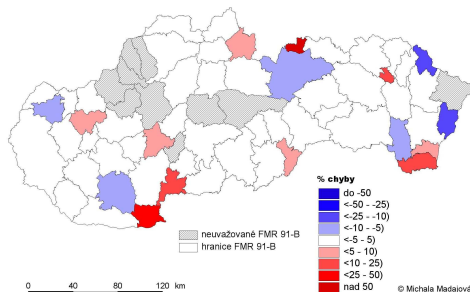
a)

Odhad počtu obyvateľov FMR 91-B k roku 2001 - percentuálne chyby
(podľa dĺžky cestnej siete)



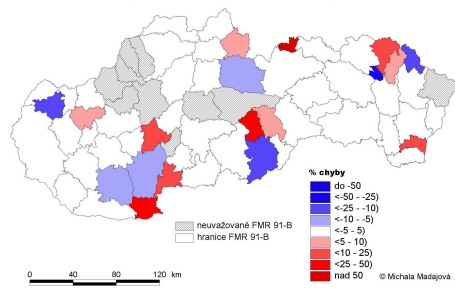
b)

Odhad počtu obyvateľov FMR 91-B k roku 2001 - percentuálne chyby
(podľa typu cestnej siete)



c)

Odhad počtu obyvateľov FMR 91-B k roku 2001 - percentuálne chyby
(podľa počtu uzlov v cestnej sieti)



d)

Obr. 1 a, b, c, d: Odhad počtu obyvateľov za FMR 91-B k roku 2001

a najhoršie pri metóde typov ciest. FMR Spišská Stará Ves dosiahol celkovo najväčšiu nadhodnotenú hodnotu počtu obyvateľov oproti skutočnosti (v prípade metódy typov ciest) a vo FMR Giraltove a Medzilaborce sa naopak počet obyvateľov zo všetkých FMR najviac pododhadol (v prípade FMR Giraltove pri metóde priestorového váženía a počtu ulov, pri FMR Medzilaborce pri metóde typov a dĺžky ciest). Doposiaľ sa nám však nepodarilo s určitosťou porozumieť nadodhadom a pododhadom jednotlivých premenných. Zatiaľ však nie je jasné, čo všetko vplyva na nadodhadovanie a pododhadovanie hodnôt príslušných premenných, ale zdá sa, že rozloha cieľovej zóny v rámci zdrojovej zóny na to nepôsobí (Madajová, 2010b). Neplatí totiž, že ak sa hodnota počtu obyvateľov v istom FMR pri jednej metóde nadodhadne, bude počet obyvateľov nadhodnocovať aj pri zvyšných metódach.

Záver

Metódy cestnej siete založené na využití dĺžky, typov a počtu uzlov v cestnej sieti ako pomocných informácií sú schopné v procese transformácie geografických dát z navzájom odlišných zónálnych systémov redukovať chyby plynúce z metódy priestorového váženía, a teda vylepšiť výsledky procesu odhadu dát. Možno skonštatovať, že niektoré cieľové zóny dosahujú konzistentné výsledky pri všetkých metódach, v iných sa však menia od metódy k metóde. Ako vyplynulo z analýzy, na proces transformácie geografických dát jednoznačne vplyva tvar zdrojových a cieľových zón, presnosť odhadov metód využívajúcich pomocné informácie však závisí aj od vzťahu medzi zdrojovými a pomocnými dátami. Čo vplyva na samotné nadodhadovanie, prípadne pododhadovanie hodnôt premennej oproti jej skutočnej hodnote zatiaľ nie je jasné a do budúcnosti predstavuje veľkú výzvu. Na formuláciu komplexnejších zistení, na základe ktorých by bolo možné formulovať akési zovšeobecnenia týkajúce sa kvality odhadov v procese transformácie geografických dát, je preto potrebná hlbšia analýza so zapojením viacerých premenných do procesu transformácie geografických a ich otestovanie na rôznych súboroch geografických jednotiek, čo, ako veríme, povedie k odhaleniu ďalších poznatkov.

Literatúra

- BEZÁK, A., 2000: Funkčné mestské regióny na Slovensku. *Geographia Slovaca* 15, GÚ SAV, Bratislava.
- FISHER, P., LANGFORD, M., 1995: Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. In *Environment and Planning A*, 1995, vol. 27, p. 211-24.
- FLOWERDEW, R., 1988: Statistical methods for areal interpolation: predicting count data from a binary variable. RR-15 Northern Regional Research Laboratory. University of Lancaster and University of Newcastle upon Tyne
- FLOWERDEW, R., GREEN, M., 1989: Statistical methods for inference between incompatible zonal systems. In *Accuracy of spatial databases*. Goodchild, M. - Gopal. S. eds., London: Taylor and Francis, 1989. p. 239–248. ISBN: 978-0-85066-847-6.
- FLOWERDEW, R., GREEN, M., 1991: Data integration: statistical methods for transferring data between zonal systems. In *Handling geographic information: Methodology and potential applications*. Masser E., - Blakemore, M. eds, 1. vyd. New York: Longman, 1991. p. 38-54. ISBN: 0-470-21792-8.
- FLOWERDEW, R., GREEN, M., 1992: Developments in areal interpolation methods and GIS. *The Annals of Regional Science*, vol. 26, 1992, p. 67–78.
- FLOWERDEW, R., GREEN, M., 1994: Areal interpolation and types of data. In *Spatial analysis a GIS*. Fotheringham, s. – Rogerson, P. eds., London: Taylor and Francis, 1994. p. 121-45.

- FLOWERDEW, R., GREEN, M. - KEHRIS, E. 1991: Using areal interpolation methods in Geographic Information Systems. *Papers in Regional Science: The Journal of the RSAI*, 1991, vol. 70, no. 3, p. 303-15.
- GKÚ, 1991: Vrstva Hranice obce. Bratislava (Geodetický a kartografický ústav).
- GKÚ, 2001: Vrstva Hranice obce. Bratislava (Geodetický a kartografický ústav).
- GKÚ, 1991-1997: Spojitá 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.
- LANGFORD, M., MAGUIRE, D. J. - UNWIN, D. J., 1991: The areal interpolation problem: Estimating population using remote sensing in a GIS framework. In *Handling geographic information*, Masser, I. - Blakemore M. B.(eds), Essex: Longman Scientific & Technical, p. 55-77. ISBN: 0-582-06730-8.
- MADAJOVÁ, M., 2010a (v tlači): Metódy transformácie priestorových dát: prehľad, klasifikácia a hodnotenie. *Acta geographica Universitatis Comenianae* 54, Bratislava.
- MADAJOVÁ, M., 2010b (v tlači). Využitie pomocných informácií pri transformácii geografických dát – vyhodnotenie vybraných postupov, *Geografický časopis* 62. No 3. GÚ SAV, Bratislava.
- MARKOFF, J., SHAPIRO, G., 1973: The linkage of data describing overlapping geographical units, *Historical Methods Newsletter*, 7: 34–46.
- MENNIS, J., HULTGREN, T., 2005: Dasymetric mapping for disaggregating coarse resolution population data. In *Proceedings of the 22nd Annual International Cartographic Conference*, 2005, July 9-16, A Coruna, Spain.
- MENNIS, J., HULTGREN, T., 2006a: 'Intelligent' dasymetric mapping and its comparison with other areal interpolation techniques. In *Proceedings of AutoCarto*, 2006, June 26-28. Vancouver, WA.
- MENNIS, J., HULTGREN, T., 2006b: Intelligent dasymetric mapping and its application to areal interpolation. In *Cartography and Geographic Information Science*, 2006, vol. 33, no. 3, p. 179-194.
- 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.
- XIE, Y., 1995: The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems* 19: 287–306.

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

Testing of Road Network Methods in the Process of Geographical Data Transformation

Michala MADAJOVÁ

Summary: *This paper evaluates the population estimates of some geographical data transformation's methods which allow data allocating from spatial inconsistent zonal systems. The basic and simplest transformation method - areal weighting - is compared with three so called network methods which use auxiliary information about the node counts, length and type of the road segments to improve the results of transformation process. One variable -population count- was estimated from 2001 census districts to the system of functional urban regions (FUR 91-B). The spatial visualisation through maps of percent errors was used to evaluate the results of estimation. The results are as follows: seems to the network methods, especially network hierarchical method, provide better results than the areal weighting method which is in general the most inaccurate method but it is not always true. There are more areas in which areal weighting perform better than the methods which use auxiliary information. We discover the accuracy of areal weighting method affect the form of source and target geographical units and the network methods yet another aspect, namely the correlation between auxiliary information and estimated variable. The areas which show out relatively bad results still exist. To evaluate and point out more general patterns require deeper and complex analysis in the future is required.*

Adresa autora:

Mgr. Michala Madajová
Katedra regionálnej geografie, ochrany a plánovania krajiny
Prírodovedecká fakulta, Univerzita Komenského
Mlynská dolina, 84215 Bratislava
madajova@fns.uniba.sk