

PRIESTOROVÁ DEZAGREGÁCIA POPULAČNÝCH DÁT S VYUŽITÍM MÁP KRAJINNEJ POKRÝVKY A NEPRIEPUSTNOSTI POVRCHUKonštantín ROSINA¹, Pavol HURBÁNEK², Peter ATKINSON³

¹ Geografický ústav, Slovenská akadémia vied, Štefánikova 49, 814 38 Bratislava, Slovensko
konstantin.rosina@gmail.com

² Katedra geografie, Pedagogická fakulta, Katolícka univerzita v Ružomberku, Hrabovská cesta 1,
034 01 Ružomberok, Slovensko
pavolhurbanek@gmail.com

³ School of Geography, University of Southampton, University Road, SO17 1BJ, Southampton,
United Kingdom

P.M.Atkinson@soton.ac.uk

Abstrakt

Priestorové rozmiestnenie obyvateľstva bolo v minulosti spravidla konceptualizované pomocou diskretných objektov (obvykle reprezentovaných vektorovými typmi údajov, napr. hranice obcí s atribútom počtu alebo hustoty obyvateľov). Takýto druh údajov o obyvateľstve je široko dostupný prostredníctvom cenzov či populačných registrov, avšak mnohí autori poukazovali na nevýhody spojené s týmto konceptom, mimo iných problém modifikovateľných územných jednotiek. Alternatívnou konceptualizáciou sú spojité polia (populačné povrchy) vzorkované pravidelnou mriežkou so želanou veľkosťou. Niektoré krajiny už pristúpili k poskytovaniu populačných dát agregovaných do pravidelných mriežok, avšak táto snaha je ešte ďaleko od celoeurópskeho rozsahu. Ako náhradné riešenie boli vyvinuté početné metódy na dezagregáciu populačných údajov do pravidelných mriežok. Najlepšie výsledky spravidla dosahujú metódy využívajúce pomocné údaje ako napr. krajinná pokrývka alebo sklonitosť terénu (tento prístup býva niekedy označovaný ako dazymetrické mapovanie). V súčasnosti najpresnejšou rastrovou databázou o obyvateľstve na celoeurópskej úrovni je „Population density grid of the EU“, vytvorený J. Gallegom v období 2001-2009, dostupný na web stránke Európskej environmentálnej agentúry (EEA), ďalej len PDG. Hlavné zdroje dát využité pri tvorbe tejto databázy boli údaje o počte obyvateľov z cenzov, hranice obcí EÚ a CORINE Land Cover (európska mapa krajinej pokrývky so 44 triedami). V predkladanom príspevku je prezentovaný pokus o ďalšie vylepšenie existujúceho modelu s využitím ďalšieho pomocného zdroja dát dostupného pre celú EÚ („EEA Fast Track Service Precursor on Land Monitoring - Degree of soil sealing at 100m spatial resolution” – rastrová databáza udávajúca stupeň nepriepustnosti povrchu od 0 do 100% s priestorovým rozlíšením 100m, ďalej len SSL) a novej iteratívnej metódy dezagregácie. Pri dazymetrickom mapovaní je pre celkový výsledok spravidla dôležitejšia kvalita použitých pomocných dát než výber konkrétnej metódy dezagregácie. Napriek skutočnosti, že presnosť SSL je rôzna od krajiny ku krajine s relatívne nízkou presnosťou na Slovensku (záujmové územie výskumu), validácia ukázala, že predbežný výsledok navrhovanej metódy dosiahol lepšiu presnosť oproti poslednej verzii PDG.

Abstract

Spatial distribution of population has been traditionally conceptualized by discrete objects (usually represented by vector data structures, e.g. boundaries of municipalities with the attribute of population count or density). While this type of population datasets is widely available from population registers and censuses, many researchers have pointed at limitations related to this concept – the modifiable areal unit problem among others. The alternative conceptualization suggested is a continuous field (a population surface) sampled with a regular grid of desired size. Agencies of several countries started to provide population data aggregated to a regular grid, but this effort is far from European coverage. To overcome this issue, numerous methods for disaggregation of population data into regular grids have been proposed. The best performing ones usually rely on ancillary data sources such as those on land cover or slope that serve as a proxy for population distribution (this approach is sometimes referred to as dasymmetric mapping). Currently, the most accurate gridded population density dataset covering the EU completely is Population density grid

of the EU, developed by Javier Gallego in 2001-2009, available from the European Environment Agency (EEA) website (hereinafter referred to as PDG). The main data sources for the development of the dataset include: census population by commune, commune boundaries, and CORINE Land Cover (European dataset with 44 land cover classes). Here, an attempt to further improve the fit of the model is presented, employing an additional EU-wide dataset (EEA Fast Track Service Precursor on Land Monitoring - Degree of soil sealing at 100m spatial resolution; hereinafter referred to as SSL) and a novel iterative method of disaggregation. In dasymetric mapping, the quality of employed ancillary data is often more important than the choice of a particular method of disaggregation. The accuracy of SSL has been proved to vary from country to country, with relatively low accuracy in Slovakia (the study area). Nevertheless, according to the performed validation, preliminary result of the suggested method outperformed the benchmark (the last version of the PDG).

Kľúčové slová: dezagregácia; dazymetrické mapovanie; hustota zaľudnenia; populačný raster

Keywords: downscaling; desaggregation; dasymetric mapping; population density; population grid

1. ÚVOD

Hustota zaľudnenia je jedna zo základných socioekonomických premenných. Dôkladné poznanie jej priestorovej diferenciácie je potrebné pre mnohé oblasti ľudskej činnosti a stalo sa častým predmetom skúmania v geografickej literatúre. Ako uvádzajú Lu, Weng a Li (2006), aktuálny a presný odhad rozmiestnenia obyvateľstva má značný význam pre decíziu sféru, plánovanie využitia krajiny a pre lepšie pochopenie interakcií medzi rastom populácie a sociálnymi, ekonomickými či prírodnými podmienkami. Príklady priestorových problémov vyžadujúcich podrobné populačné dáta ponúkajú aj Wu a Murray (2007). Ide o procesy priestorového plánovania ako lokalizácia parkov, čističiek odpadových vôd, knižníc, nemocníc a dopravnej infraštruktúry či vymedzovanie spádových zón obchodných prevádzok (napr. Martin and Williams 1992; Plane and Rogerson 1994).

Ďalším okruhom tém, kde sú potrebné podrobnejšie populačné dáta ako tie, ktoré sú bežne dostupné, sú hodnotenie rizík, negatívnych vplyvov na obyvateľstvo, tvorba scenárov krízového riadenia a zmierňovanie následkov prírodných či antropogénnych katastrof. Primárne za účelom hodnotenia rizík bola vytvorená aj globálna rastrová databáza hustoty zaľudnenia s rozlíšením 30 sekúnd – LandScan (Dobson et al. 2000). Európska rastrová databáza hustoty zaľudnenia s priestorovým rozlíšením 100m, vytvorená v Joint Research Centre (JRC) of European Commission (Gallego 2010) bola využitá na odhad rozsahu škôd spôsobených povodňami (Tralli et al. 2005) a na dopad hluku na obyvateľstvo bývajúce v okolí letísk (Vinx and Vise'e 2008).

2. SÚČASNÝ STAV PROBLEMATIKY

2.1 Kritika agregovaných údajov a ich reprezentácie formou kartogramu

Obvykle sa dostupné dáta o rozmiestnení obyvateľstva obmedzujú len na počty obyvateľov agregované do administratívnych, resp. štatistických územných jednotiek a najčastejšie sú reprezentované kartogramom. Tento spôsob reprezentácie je však nedostatočný pre analytické i vizualizačné účely a je často kritizovaný (Openshaw and Taylor 1981; Openshaw 1983; Bracken and Martin 1989; Goodchild, Anselin and Deichmann 1993; Langford and Unwin 1994; Martin 1996). Jedným z jeho hlavných nedostatkov je problém modifikovateľnej územnej jednotky, ktorý Openshaw (1983) popísal ako situáciu keď modifikácia hraníc a miery agregácie údajov významne ovplyvní výsledok priestorovej analýzy dát. Pozorované priestorové štruktúry tak môžu byť do veľkej miery dané konkrétnou konfiguráciou hraníc použitých zón a zistené vzťahy pozorované medzi premennými sú platné iba pre daný spôsob agregácie dát (Martin 1996).

Ďalším problémom je ekologická chyba, ktorú Goodchild (2011) definuje ako prípad „cross-scale inference“. Tapp (2010) vysvetľuje, že agregované dáta (napr. za sčítacie obvody) znázornené prostredníctvom kartogramu predstavujú ekologickú chybu – že všetky pozorovania v určitom areáli sú podobné priemeru za celý areál. Inými slovami, kartogram zvädza čitateľa k domnienke, že hustota zaľudnenia je rovnaká v každom mieste znázorneného areálu, čo spravidla nie je pravda.

Langford a Unwin (1994) považujú za problém aj samotnú nerovnakú veľkosť jednotiek – agregácia údajov je formou generalizácie (extrémne hodnoty sa spriemerujú), čím väčšia je jednotka, tým je tento efekt silnejší a rôzne veľké jednotky sú tým pádom vzájomne neporovnateľné. V prípade Slovenska sú priestorovo najpodrobnejšou bežne dostupnou úrovňou populačných dát obce. Najväčšia obec je svojou rozlohou viac ako 1000-krát väčšia v porovnaní s najmenšou obcou.

Posledným problémom, ktorý zmienime je fakt, že hranice administratívnych/štatistických územných jednotiek nie sú optimalizované pre dáta, ktoré sú nimi znázorňované, ale vznikali pod vplyvom mnohých iných faktorov, napríklad historických, politických, praktických (uľahčenie zberu dát) a pod. Ako zdôrazňuje Martin (1989), logický vzťah medzi premennou a nepravidelnými jednotkami je neznámy, na rozdiel od fyzicko-geografických premenných, ktoré možno znázorniť pomocou ich „prírodných“ hraníc, takže sa zachová paralela medzi zónami reprezentovanými v dátach a tými, ktoré existujú v realite.

2.2 Priestorová dezagregácia ako riešenie

„Niektoré z uvedených problémov, najmä ekologická chyba, sú fundamentálne späté s agregovanou povahou zonálnych údajov a bez prístupu k pôvodným, individuálnym záznamom (pozn. a poznania ich presnej polohy v priestore) nemôžu byť prekonané. Avšak existujú alternatívy k neustálemu narábaniu s údajmi v zonálnej podobe, pričom najväčšia pozornosť bola venovaná konceptu modelovania tých istých informácií v podobe hustotných povrchov reprezentovaných pravidelnou mriežkou, resp. rastrom (angl. grid), ktorý je vhodný pre spracovanie v prostredí rastrového GIS-u“ (Martin 1996). Výhodou rastrovej reprezentácie je pravidelná veľkosť, tvar a časová stálosť jej základných jednotiek – buniek rastra.

V krajinách, kde sú známe polohové súradnice individuálnych záznamov z cenzu (napr. Rakúsko), už prišli k publikovaniu výsledkov aj vo forme pravidelného rastra. Tento prístup zatiaľ na Slovensku nebolo možné aplikovať a v rámci Európy zatiaľ prevažuje publikovanie údajov iba v sústave tradičných štatistických územných jednotiek. Nevýhodou agregácie individuálnych záznamov do podrobného rastra je skutočnosť, že niektoré bunky nadobudnú príliš malé hodnoty, ktoré nemôžu byť zverejnené z dôvodu ochrany dôverných údajov, keďže by z nich následne mohli byť získané dáta o jednotlivcoch.

Preto je väčšinou jediným riešením dezagregácia údajov z dostatočne malých štatistických jednotiek (napr. obcí). Ide o aproximatívnu metódu, ktorá síce nikdy nedokáže úplne „zrekonštruovať“ detaily priestorovej štruktúry, ktoré sa stratili pri agregácii dát (Martin 1989), avšak ako potvrdzujú existujúce práce (napr. Gallego 2010; Dobson et al. 2000), možno dospieť k prijateľným výsledkom.

2.3 Jednoduché metódy dezagregácie

V existujúcej literatúre nájdeme množstvo rôznych prístupov k dezagregácii populačných dát, resp. príbuzným témam označovaným angl. termínmi ako napr. downscaling, change of support, areal interpolation, dasymmetric mapping, population surface modeling, atď. Prvou skupinou metód sú jednoduché metódy, ktoré sa spoliehajú iba na matematické a (geo)štatistické metódy, ktoré interpolujú hodnoty zo známych reprezentatívnych bodov (napr. populačne vážených centroidov v prácach D. Martina) do ľubovoľného iného bodu bez pomocných priestorových údajov, len na základe funkcie, u ktorej sa predpokladá, že bude verne popisovať priestorovú diferenciáciu skrytú pod zdrojovými zónami (underlying distribution).

Jednou z prvých a najčastejšie citovaných prác na tému reprezentácie hustoty obyvateľstva ako spojitého povrchu je Toblerov text Smooth pycnophylactic interpolation for geographical regions (1979). Tobler tu formálne popísal metódu tvorby reprezentácie hustoty obyvateľov ako hladkého povrchu, hustota je funkciou súradníc x , y a to tak aby sa počet obyvateľov zdrojových zón zachoval – zavádza pojem pyknofylaktický (objem zachovávajúci). Pyknofylaktická vlastnosť je splnená vtedy, keď výsledok získaný zintegrovaním povrchu nad územím zdrojovej zóny sa rovná pôvodnému počtu obyvateľov, čo je dôležitý predpoklad presnosti modelu.

V oblasti jednoduchých metód je významný i prínos D. Martina, ktorý interpoloval počty obyvateľov z centroidov sčítacích obvodov v UK pomocou adaptívnej kernelovej funkcie (modelujúcej klesajúcu hustotu zaľudnenia s rastúcou vzdialenosťou od centroidu). Vzhľadom na to, že v tej dobe neboli dostupné hranice obvodov (iba ich centroidy) nebolo možné splniť ani overiť pyknofylaktickú vlastnosť. (Martin 1989; Bracken and Martin 1989; Bracken and Martin 1991).

Z jednoduchých metód je treba uviesť aj areálovú interpoláciu, ktorú popísali Goodchild a Lam (1980); Goodchild, Anselin a Deichmann (1993). Ich práce sa však nezameriavajú na dezagregáciu údajov do pravidelných územných jednotiek podstatne menších ako sú zdrojové zóny (napr. do buniek rastra), ale na prevod údajov medzi nekompatibilnými systémami zón.

2.4 Dazymetrické metódy dezagregácie

Druhou skupinou sú metódy, ktoré na spresnenie odhadu využívajú pomocné priestorové údaje. Tieto metódy sa označujú ako dazymetrické (u niektorých autorov tiež ako „inteligentné“ – Sadahiro 1999; Langford 2007). Ako pomocné údaje sa veľmi často využívajú mapy krajiny pokrývky a využitia krajiny, niekedy iba v binárnej podobe (rezidenčné – nerezidenčné plochy). Langford a Unwin (1994) testovali

využitie vrstvy rezidenčných plôch ad-hoc klasifikovanej z družicových snímok programu Landsat na dezagregáciu hustoty obyvateľov v grófstve Leicestershire. Mennis (2003) podobne dezagregoval hustotu obyvateľov v časti Pennsylvánie, avšak satelitné snímky klasifikoval do troch tried. Lu, Weng a Li (2006) využili kombináciu rezidenčných plôch a stupňa nepriepustnosti povrchu na odhad populácie v jednom okrese štátu Indiana.

Iným typom pomocných údajov sú cestné siete, resp. ich uzly (Voss, Long and Hammer 1999) alebo adresné body, ktoré využila vo svojej práci Tapp (2010).

Osobitne však treba spomenúť dve dazymetrické mapy, ktoré pokrývajú podstatne rozsiahlejšie územie. V americkom Oak Ridge National Laboratory je vyvíjaná databáza LandScan, ktorá má globálne pokrytie s priestorovým rozlíšením 30 sekúnd (pre USA existuje podrobnejšia verzia s rozlíšením 3 sekundy). Ako pomocné údaje boli využité viaceré druhy údajov - krajinná pokrývka, intenzita nočného osvetlenia, sklonitosť reliéfu, cestné siete a i. (Dobson et al. 2000; Bhaduri et al. 2002).

JRC vyvíja európsku dazymetrickú mapu hustoty obyvateľstva s priestorovým rozlíšením 100m od roku 2000 (the population density grid of the EU – ďalej PDGEU). Prvá verzia bola zverejnená v roku 2004 a vychádzala ešte z populačných údajov k roku 1991, ostatné verzie už boli založené na údajoch okolo roku 2001. Zatiaľ ostatná, piata verzia (v. 5) bola zverejnená v roku 2009. Jednotlivé verzie sa odlišujú predovšetkým metódou dezagregácie a čiastočne aj pomocnými údajmi, kľúčovým zdrojom pomocných údajov však vždy bola európska mapa krajiny pokrývky CORINE Land Cover 2000 (CLC). PDGEU je zdarma dostupná na web stránke Európskej environmentálnej agentúry (EEA 2009) a jej hlavným autorom je Javier Gallego (Gallego and Peedell 2001; Gallego 2010). PDGEU vo verzii 5 (v. 5) predstavuje doposiaľ najpodrobnejšiu a najpresnejšiu rastrovú reprezentáciu hustoty zaľudnenia, ktorá pokrýva územie Slovenska.

3 DÁTA A METÓDY

Cieľom nášho experimentu bolo vytvoriť dezagregovanú mapu hustoty zaľudnenia pre celé územie Slovenska s priestorovým rozlíšením 100m a iba s využitím voľne dostupných zdrojov dát. Výsledný produkt by mal byť presnejší ako PDGEU, verzia 5. Ako východiskové dáta boli použité počty obyvateľov za obce, teda základné územné jednotky (ZUJ). Z cenzu v roku 2001 je síce možné získať údaje za menšie jednotky – základne sídelné jednotky (ZSJ), avšak tieto je potrebné si ponechať pre validáciu výsledkov a porovnanie s existujúcimi produktmi.

Hlavným nedostatkom PDGEU je skutočnosť, že minimálna mapovacia jednotka (MMJ) pomocných údajov – CLC – je 25ha. Dôsledkom toho sa približne v 15% ZUJ na Slovensku nenachádza CLC trieda umelých povrchy. Ide predovšetkým o najmenšie obce a obce z rozptýleným osídlením. V týchto obciach potom PDGEU nedokázal presnejšie lokalizovať jadro s vyššou hustotou zaľudnenia a výsledok sa približoval homogénnemu kartogramovému vyjadreniu.

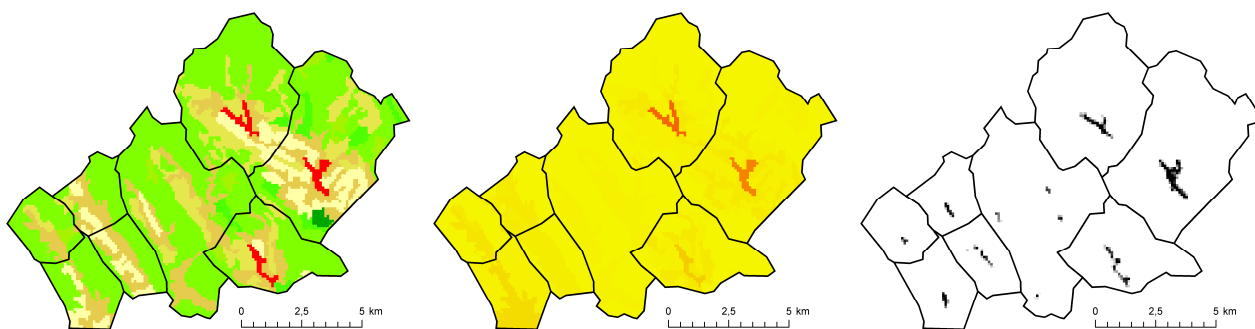
3.1 Európska mapa zastavaných plôch – vhodný zdroj pomocných dát

Ako najlepšie riešenie tejto situácie sa javilo využitie viacerých zdrojov pomocných údajov. V rámci európskej iniciatívy Global Monitoring for Environment and Security (GMES) bola vytvorená rastrová databáza s veľkosťou bunky 100m znázorňujúca v každej bunke stupeň zastavanosti, resp. umelej nepriepustnosti povrchu (angl. degree of soil sealing) ako percentuálnu hodnotu od 0 do 100. Celý názov produktu je Fast Track Service Precursor on Land Monitoring - Degree of soil sealing 100m (uprednostňujeme skrátený názov soil sealing layer, ďalej len SSL) a je zdarma dostupný na web stránke EEA (EEA 2010). Podobne ako CLC 2006, SSL bol odvodený z európskej satelitnej mozaiky s priestorovým rozlíšením cca 20m, IMAGE 2006.

Zastavané plochy do veľkej miery reprezentujú prítomnosť a aktivitu ľudí. Za predpokladu, že SSL úplne presne reprezentuje realitu, už 1m² zastavanej plochy by mal byť teoreticky zaznamenaný ako hodnota 1% príslušnej 100x100m bunky. Takáto presnosť v praxi nie je dosiahnuteľná u dát odvodených z diaľkového prieskumu Zeme (bližšie o presnosti SSL pojednávajú napr. Hurbánek et al. 2010), napriek tomu, ako vidieť na obr. 1, SSL umožňuje identifikovať jadrá tých obcí, ktoré neboli zaznamenané v CLC. Obcí, ktoré majú

iba nulové bunky SSL je na Slovensku 0,7% (výrazne menej ako 15% obcí s chýbajúcou triedou CLC umelé povrchy). Istou nevýhodou SSL je, že vysoký stupeň zastavanosti majú všetky plochy pokryté umelým materiálom - aj parkoviská, letiská, ale tiež budovy skladov, priemyslu a obchodu a pod., hoci majú z hľadiska rezidenčnej populácie zväčša nulovú, príp. len veľmi nízku hustotu zaľudnenia. CLC obsahuje niektoré triedy odrážajúce využitie krajiny človekom (napr. trieda 121 plochy priemyslu a obchodu alebo 124 letiská), ktoré umožňujú čiastočne odlišiť nerezidenčné zastavané plochy od rezidenčných. Hoci limitom je MMJ CLC s rozlohou 25ha, je zrejmé, že pri použití SSL spolu s CLC prichádza ku vzájomnej kompenzácii ich nedostatkov. Navyše, SSL a CLC majú takmer zhodné, celoeurópske pokrytie, takže v prípade vyvinutia uspokojivej metódy na území Slovenska je možnosť testovať jej aplikáciu na podstatne väčšom území.

Myšlienka integrácie údajov o stupni zastavanosti a krajinej pokrývke za účelom dezagregácie nie je úplne nová, podobný prístup už testovali Lu, Weng a Li (2006) a Steinocher et al. (2010).



Obr. 1. Príklad obcí obsahujúcich a neobsahujúcich triedu umelých povrchov v CLC (červená farba na obrázku vľavo), PDGEU odvodené z CLC (v strede) a stupeň zastavanosti povrchu (napravo)

3.2 Použitá metóda dezagregácie

Pri dezagregácii sme sa rozhodli vychádzať z existujúcej metódy popísanej v literatúre, ktorú sme následne modifikovali tak, aby zodpovedala našim požiadavkám. Z piatich rôznych metód využívajúcich CLC, ktorými bolo vytvorených 5 verzií PDGEU sme sa rozhodli pre prvú metódu (Gallego and Peedell 2001) pre jej jednoduchosť a preto, že ako jediná využíva iba voľne dostupné dáta. Na začiatok boli prijaté základné zjednodušujúce predpoklady:

1. Údaje využité pre dezagregáciu odrážajú objektívnu realitu s dostatočnou presnosťou;
2. obyvateľstvo môže byť lokalizované výlučne v zastavaných plochách; za rozlohu zastavanej (t.j. potenciálne obývateľnej) plochy v 100m bunke rastra v hektároch sa považuje jej hodnota SSL (0-100) vydelená 100; napr. rozloha zastavanej plochy v bunke s hodnotou SSL 20 bude 0,2 ha;
3. hustota zaľudnenia zastavaných plôch je rovnaká v každej unikátnej kombinácii obce a triedy CLC;
4. pomer hustôt zaľudnenia zastavaných plôch v každej unikátnej dvojici tried CLC je konštantný pre všetky obce v určitom zoskupení obcí (regióne).

Celé územie Slovenska bolo rozdelené do 100x100m buniek identických s európskym rastrom založeným na súradnicovom systéme ETRS-LAEA (Lambertova azimutálna rovnako-plošná projekcia z geodetického dátumu ETRS 1989). Legenda CLC, ktorá má pôvodne 44 tried bola zoskupená do 17 tried podľa Gallega a Peedella (2001). V každej z pôvodných 44 tried, ktoré boli zoskupené do jednej novovytvorenej triedy sa predpokladala podobná hustota zaľudnenia. Pre každú z buniek na území Slovenska bola zaznamenaná príslušnosť k jednej zo 17 tried CLC a hodnota SSL.

Hustota zaľudnenia na zastavanú plochu pre každú unikátnu kombináciu obce (označenie indexom m) a triedy CLC (označenie indexom c) môže byť vypočítaná podľa rovnice (1) (Gallego and Peedell 2001). Y predstavuje hustotu zaľudnenia, X známy skutočný počet obyvateľov, S rozlohu a U je koeficient zabezpečujúci, že pomer hustôt zaľudnenia v každej dvojici tried CLC je konštantný pre všetky obce

v jednom zoskupení obcí. S_{cm} predstavuje iba sumu rozlôh zastavanej plochy všetkých buniek v triede c a obci m a nie celkovú rozlohu, ako tomu bolo v pôvodnej metóde Gallega a Peedella (2001).

Rovnica (1) má pyknofylaktickú vlastnosť, takže suma počtu obyvateľov vo všetkých bunkách v jednej obci sa rovná skutočnému počtu obyvateľov obce.

$$Y_{cm} = U_c \frac{X_m}{\sum_c S_{cm} U_c} \quad (1)$$

Aby sme mohli určiť hodnoty koeficientov U_c (udávajúce pomery hustoty medzi triedami CLC), ktoré by vytvorili čo najlepší odhad, potrebovali by sme údaje o počte obyvateľov za jednotky menšie ako obce, vďaka ktorým by bolo možné opakovane porovnávať odhadnutý počet obyvateľov so skutočným a následne prispôbovať hodnoty koeficientov, pokiaľ nedosiahneme najlepší výsledok. Takéto dáta však nie sú k dispozícii (s výnimkou ZSJ, ktoré ale boli vyhradené na finálnu validáciu).

Gallego a Peedell (2001) navrhujú spôsob, ktorým možno uvedenú situáciu riešiť:

1. Dezagregovať údaje za väčšie jednotky ako obce (regióny, okresy) s využitím odhadnutých iniciálnych hodnôt koeficientov U_c – rovnica (2), r je index pre región;
2. vzniknutý raster agregovať na úrovni obcí – rovnica (3);
3. porovnať výsledné hodnoty za obce so skutočnými hodnotami a vypočítať indikátory nezhody – rovnice (4) a (5)

$$Y_{cr} = U_c \frac{X_r}{\sum_c S_{cr} U_c} \quad (2) \quad X_m^* = \sum_c S_{cm} Y_{cr} \quad (3) \quad \delta_r = \sum_{m \in r} |X_m^* - X_m| \quad (4) \quad \delta = \sum_m |X_m^* - X_m| \quad (5)$$

4. zmeniť hodnoty koeficientov tak, aby sa znížila nezhoda a to nasledovne:
 - a. podľa rovníc (6) a (7) sa vypočíta Pearsonov korelačný koeficient pre každú kombináciu regiónu r a triedy CLC c medzi:
 - množinou pomerov odhadnutej a skutočnej populácie všetkých obcí m v regióne r ;
 - a množinou pomerov rozlohy zastavanej plochy v triede c v obci m ku celkovej rozlohe zastavanej plochy v obci m pre všetky obce v regióne r ;
 - b. ak je zistená pozitívna korelácia, môžeme usudzovať, že zvyčajne sú nadhodnotené počty obyvateľov v obciach, ktoré majú relatívne vysoké zastúpenie triedy c , takže koeficient pre danú triedu a región musí byť znížený (a naopak v prípade negatívnej korelácie musí byť zvýšený) podľa rovnice (8);
5. predchádzajúce kroky sú opakované pokiaľ sa celková miera nezhody (rovnica 5) neustáli;
6. v poslednom kroku sa dosadia hodnoty koeficientov U_{cr} z poslednej iterácie do rovnice (9) a odhadne sa výsledná hustota zaľudnenia zastavaných plôch v obci m a triede c . Tá istá sada koeficientov sa použije pre všetky obce v jednom regióne. Rovnica (9) zaisťuje, že suma odhadnutých počtov obyvateľov za obec sa bude rovnať skutočnému počtu obyvateľov.

$$\psi_m = \frac{X_m^*}{X_m} \quad (6) \quad \rho_{cr} = \text{corr} \left(\psi_m, \frac{S_{cm}}{S_m} \right) \quad (7) \quad U'_{cr} = U_c \left(1 - \frac{\rho_{cr} \times \delta_r}{2 \times X_r} \right) \quad (8) \quad Y_{cm} = U_{cr} \frac{X_m}{\sum_c S_{cm} U_{cr}} \quad (9)$$

4 VÝSLEDKY A VALIDÁCIA

Iteratívna metóda odhadu koeficientov popísaná v predošlej kapitole bola vykonaná pre dva rôzne systémy regiónov a s dvomi rôznymi iniciálnymi hodnotami koeficientov U_{cr} , celkovo tak boli aplikované 4 rôzne konfigurácie. Využitie systémov regiónov boli NUTS-3 (zodpovedajúce samosprávnym krajom) v počte 8 a funkčné mestské regióny (FMR-A) podľa Bezáka (2000) v počte 51.

Ako iníciaľne koeficienty boli využité:

A. uniformné hodnoty 1 pre všetky triedy CLC

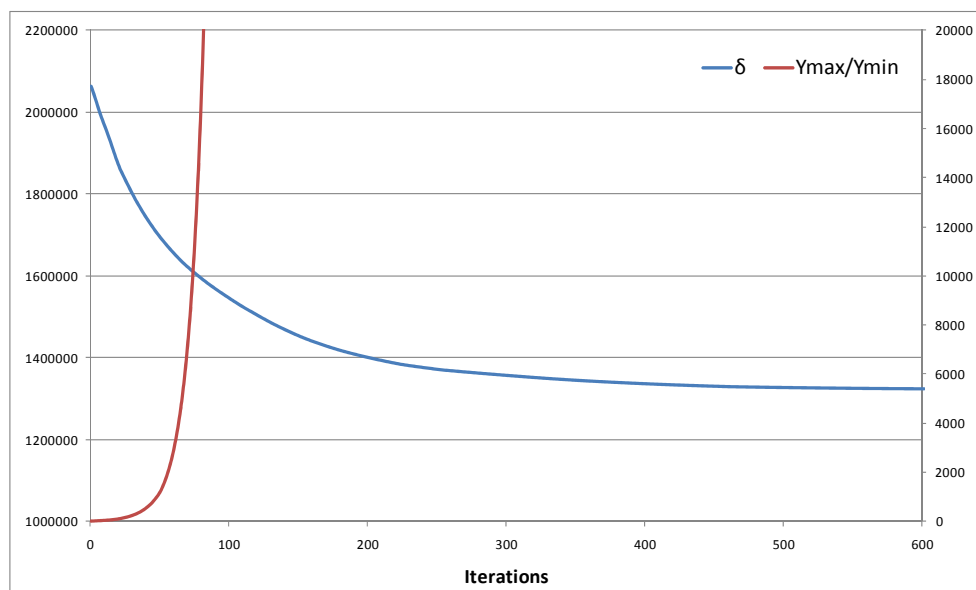
B. rôznorodé hodnoty od 32 pre triedu súvislej sídelnej zástavby až po 0 napr. pre vodné plochy.

Zmena celkového indikátora nezhody δ (rovnica 5) v priebehu iterácie je dokumentovaná v tabuľke 1.

Tab 1. Rozsah celkového indikátora nezhody δ pre rozličné parametrizácie metódy dezagregácie

		Regionálny systém			
		FMR-A		NUTS-3	
		δ_{\max}	δ_{\min}	δ_{\max}	δ_{\min}
Iniciálne koeficienty	A	1,921,096	1,286,120	1,991,017	1,787,418
	B	2,062,586	1,322,974	2,094,570	1,798,614

Hoci cieľom popísaného iteratívneho procesu bolo znížiť celkovú mieru nezhody, neznamená to, že najnižšia hodnota δ je nevyhnutne najlepšia. Po príliš veľkom počte opakovaní siete δ ešte nepatrne klesá, avšak hodnoty koeficientov majú tendenciu nadobúdať extrémne hodnoty. Gallego a Peedell (2001) navrhujú limitujúce opatrenie – pomer medzi najvyššou a najnižšou hodnotou hustoty v obci by nemal prekročiť 10 000. Na obr. 2 je znázornený priebeh indikátora celkovej nezhody δ (modrá) a maximálneho pomeru najvyššej a najnižšej hustoty v jednej obci spomedzi všetkých obcí (červená). V uvedenom prípade bol limit prekročený po 76 opakovaníach.



Obr. 2. Príklad priebehu indikátora celkovej nezhody a maximálneho pomeru hustoty dvoch tried CLC v jednej obci pre dezagregáciu s využitím systému regiónov FMR-A a rôznorodých iníciaľných koeficientov

Pre porovnanie, Bielecka (2005), ktorá tiež využívala metódu navrhnutú Gallegom a Peedellom poznamenáva, že uspokojujúce výsledky boli dosiahnuté po 30 iteráciách. V pôvodnej metóde využitej na tvorbu prvej verzie PDGEU boli použité hodnoty koeficientov z cca 40. iterácie. Z ohľadom na uvedené sme považovali za zaujímavé vykonať validáciu s viacerými sadami koeficientov vzatými z rôzneho štádia iteratívneho procesu. Pre každú zo štyroch parametrizácií boli nakoniec vybrané tri sady koeficientov – okolo 40. iterácie, z iterácie kedy bola prekročená limitujúca podmienka (priesečník červenej a modrej čiary na Obr. 2) a pre zaujímavosť aj koeficienty z ešte neskoršieho štádia čím sme získali 12 sád koeficientov, s ktorými bola uskutočnená validácia.

Následne bola vypočítaná hustota zaľudnenia na zastavanú plochu pre každú kombináciu triedy CLC a obce. Pre každú bunku výsledného rastra bol vypočítaný počet obyvateľov na základe rozlohy jej zastavanej plochy a príslušnosti k triede CLC a obci. Z výsledného rastra boli agregované hodnoty do ZSJ, ktorých počet je viac ako dvojnásobný v porovnaní so ZUJ. Jednotlivé parametrizácie boli potom porovnané navzájom a s PDGEU v. 5 prostredníctvom sumy absolútnych chýb. Výsledok je znázornený v Tab.3. Všetky parametrizácie založené na systéme regiónov NUTS-3 dosiahli lepší výsledok ako posledná verzia PDGEU. Parametrizácie využívajúce systém FMR-A dosiahli horší výsledok s výnimkou jedného prípadu.

Tab 2. Výsledky validácie na úrovni základných sídelných jednotiek

Systém regiónov	Iniciálne koeficienty	Iterácia #	Suma abs. chýb
FMR-A	Uniformné	40	2 636 465
FMR-A	Uniformné	87	2 731 697
FMR-A	Uniformné	150	2 846 496
FMR-A	Rôznorodé	40	<u>2 586 726</u>
FMR-A	Rôznorodé	73	2 701 469
FMR-A	Rôznorodé	150	2 847 268
NUTS3	Uniformné	46	<u>2 488 171</u>
NUTS3	Uniformné	120	<u>2 503 751</u>
NUTS3	Uniformné	216	<u>2 566 998</u>
NUTS3	Rôznorodé	60	<u>2 528 961</u>
NUTS3	Rôznorodé	124	<u>2 592 926</u>
PDGEU v. 5			2 598 599

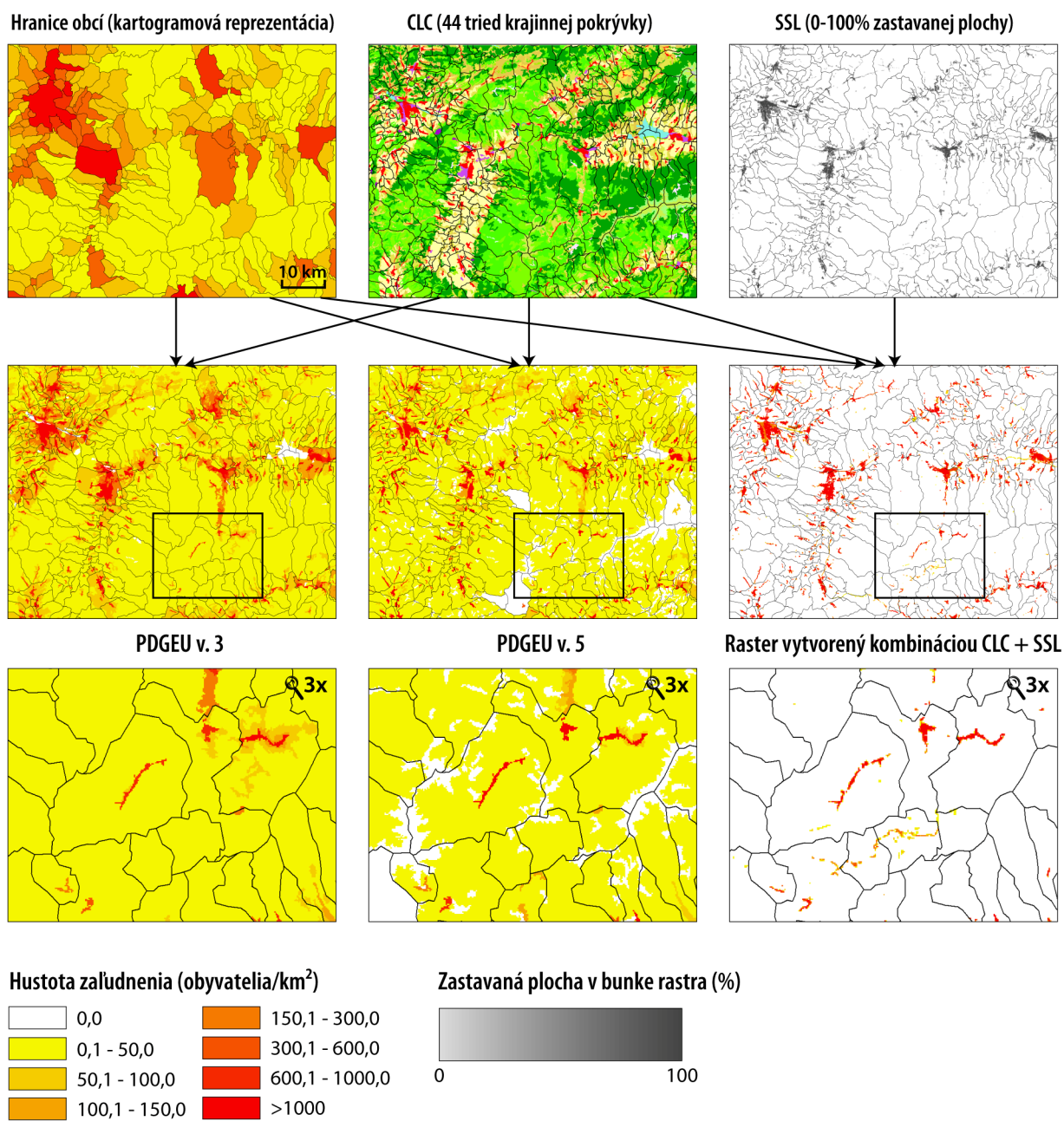
5 ZÁVER

Väčšina testovaných parametrizácií dezagregačnej metódy využívajúcej iteratívny algoritmus na odhad koeficientov dokázala z celkového pohľadu lepšie predpovedať priestorové rozmiestnenie obyvateľstva ako posledná verzia PDGEU, čo možno považovať za naplnenie cieľa práce. Ako vidieť na zväčšených výrezoch na Obr. 3, vďaka zapojeniu nového zdroja pomocných údajov, SSL, bolo možné presnejšie identifikovať rozmiestnenie obyvateľstva aj v obciach, kde chýbala reprezentácia sídelného jadra obce v CLC. Na Obr. 4 – 6 je znázornená Bratislava a jej okolie (pohľad zo severovýchodu), najskôr ako kartogram, potom ako PDGEU a následne ako raster vytvorený dezagregáciou na základe SSL a CLC. Je tu vidieť, že CLC hrá významnú úlohu v odlíšení nerezidenčných zastavaných plôch (diaľnica, letisko), ktoré tak nadobudli nízku hodnotu zaľudnenia.

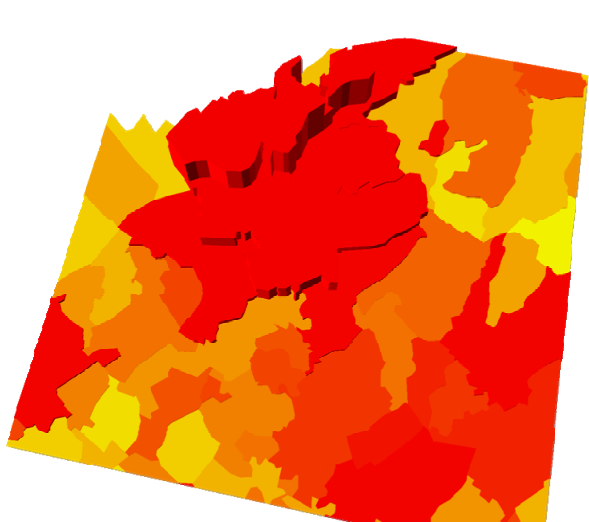
Avšak treba podotknúť, že predpoklady na zlepšenie presnosti neboli celkom naplnené. Od využitia novej, podrobnejšej databázy sa očakávalo výraznejšie zlepšenie odhadu, ako však vidieť v Tab. 2, pokrok je iba veľmi mierny. Možné vysvetlenie je nepotvrdenie prvého zjednodušujúceho predpokladu – že údaje využité pre dezagregáciu odrážajú objektívnu realitu s dostatočnou presnosťou. Ako sa ukázalo pri hodnotení SSL na Slovensku (Rosina 2010, Hurbánek et al. 2010), korelácia medzi hodnotami buniek SSL a skutočným stupňom zastavanosti v realite je nie vždy pozitívna ani lineárna. Tieto nedostatky by bolo možné kompenzovať lepším poznaním vzťahu medzi hodnotami SSL a realitou a jeho zohľadnením pri dezagregácii.

Ďalším problémom je spôsob validácie výsledkov, kde boli využité najpodrobnejšie dostupné údaje o obyvateľstve – ZSJ, ktorých je na Slovensku 6850. Je to síce výrazne viac ako 2920 ZUJ, avšak ich rozmiestnenie je veľmi nerovnomerné – až 2/3 ZUJ sú tvorené iba jednou ZSJ, takže validácia prakticky prebehla iba v tretine obcí. Je pravdepodobné, že ak by boli k dispozícii podrobnejšie dáta, výsledok validácie by bol značne odlišný.

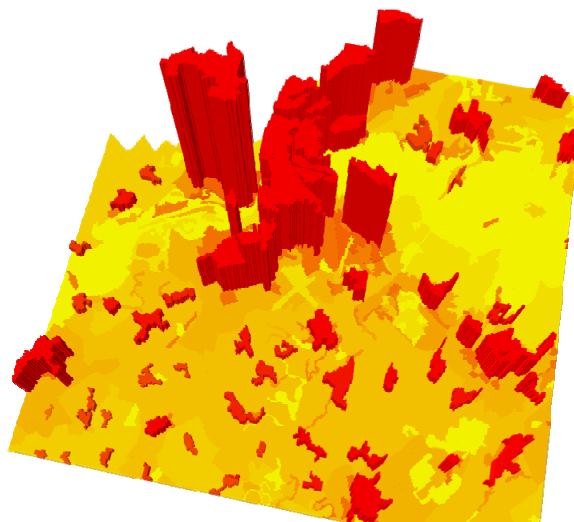
Okrem kvality použitých dát je nesporne priestor na zlepšenie aj v samotnej metóde dezagregácie, kde je potrebné ďalej skúmať citlivosť výsledku na konkrétne zoskupenie tried CLC do menšieho počtu tried, použitý regionálny systém či iníciaľne hodnoty koeficientov.



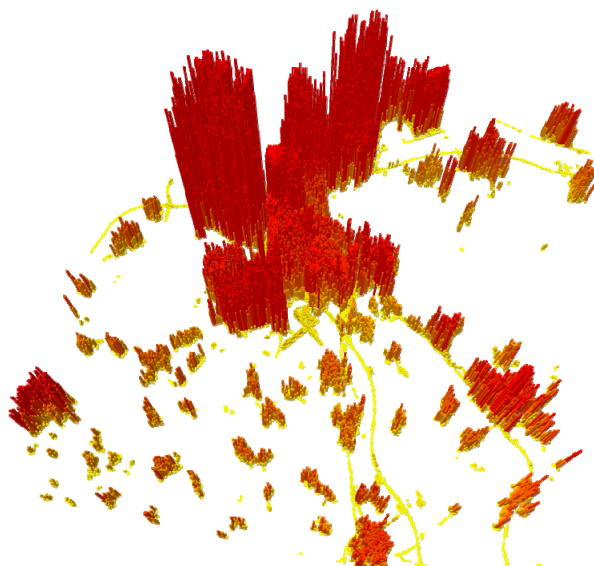
Obr. 3. Schéma vzťahov medzi vstupnými údajmi a výslednými produktmi



Obr. 4. 3D zobrazenie kartogramu



Obr. 5. 3D zobrazenie PDGEU



Obr. 6. 3D zobrazenie výsledného rastra vytvoreného kombináciou CLC a SSL

Tento príspevok je jedným z výstupov dosiahnutých riešením vedeckého projektu č. 2/0018/10 "Časovo-priestorová analýza využívania krajiny: hodnotenie dynamiky zmien, fragmentácie a stability aplikáciou dátových vrstiev CORINE land cover" na Geografickom ústave SAV za podpory grantovej agentúry VEGA.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 220832.

6. ZOZNAM LITERATÚRY

- Bezák, A. (2000): Funkčné mestské regióny na Slovensku. *Geographia Slovaca*, 15, 2000, pp. 1 – 89.
- Bhaduri, B., Bright, E., Coleman, P., & Dobson, J. (2002). LandScan: Locating people is what matters. *Geoinformatics*, 5(2), 34-37.
- Bielecka, E. (2005): A dasymetric population density map of Poland. In: Proceedings of the International Cartographic Conference, July 9-15, A Coruña, Spain.
- Bracken, I. and Martin, D. (1989). The generation of spatial population distributions from census centroid data. *Environment and Planning A* 21:537–43.
- Bracken, I. and Martin, D. (1991). Techniques for modelling population-related raster databases. *Environment and Planning A* 23:1069-1075
- Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., & Worley, B. A. (2000). LandScan: A global population database for estimating populations at risk. *Photogrammetric Engineering and Remote Sensing*, 66(7), 849–857.
- EEA (2009). Population density disaggregated with Corine land cover 2000. Dostupné online: http://www.eea.europa.eu/data-and-maps/data/ds_resolveuid/F6907877-C585-45DE-B93F-E7FC0975DE2A
- EEA (2010). EEA Fast Track Service Precursor on Land Monitoring - Degree of soil sealing 100m. Dostupné online: http://www.eea.europa.eu/data-and-maps/data/ds_resolveuid/57ecd963002327759d99a3f2d1e1dfe1
- Gallego, J. (2010). A population density grid of the European Union. *Population and Environment*, Vol. 31, No. 6., pp. 460-473.
- Gallego, J., & Peedell, S., (2001). Using CORINE Land Cover to map population density. Towards Agrienvironmental indicators, Topic report 6/2001 European Environment Agency, Copenhagen, pp. 92–103.
- Goodchild, M. (2011). Scale in GIS: An overview. *Geomorphology*, 130(1-2):5-9.
- Goodchild, M., & Lam, N. (1980). Areal interpolation: a variant of the traditional spatial problem. *Geo-Processing*, 1, 297–312.
- Goodchild, M., Anselin, L., & Deichmann, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, 25, 383–397.
- Hurbanek, P., Atkinson, P. M., Chockalingam, J., Pazur, R. and Rosina, K. (2010). Accuracy of Built-up Area Mapping in Europe at Varying Scales and Thresholds. In *Accuracy 2010 : proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Leicester : University of Leicester, 385-388.
- Langford, M. (2007). Rapid facilitation of dasymetric based population interpolation by means of raster pixel maps. *Computers, Environment and Urban Systems* 31: 19-32.
- Langford, M., & Unwin, D. J. (1994). Generating and mapping population density surfaces within ageographical information system. *Cartographic Journal*, 31(1), 21–26.
- Lu, Weng and Li (2006). Residential population estimation using a remote sensing derived impervious surface approach. *International Journal of Remote Sensing*, 27(16), 3553-3570.
- Martin, D. (1989). Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers*, NS 14, 90-97.
- Martin, D. (1996). An assessment of surface and zonal models of population. *International Journal of Geographical Information Systems*, 10(8), 973–989.
- Martin, D., and Williams, H. (1992). Market-Area Analysis and Accessibility to Primary Health-Care Centers. *Environment and Planning A* 24, 1009–1019.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer* 55(1): 31-42.

Openshaw, S. (1983). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography* vol. 38. Norwich: Geobooks.

Openshaw, S. And Taylor, P. J. (1981). The modifiable areal unit problem. In *Quantitative Geography: A British View*. Wrigley, N. - Bennett R, J. (eds). London: Routledge and Kegan Paul, 1981. p. 335-50.

Plane, D. A., and Rogerson, P.A. (1994). *The Geographical Analysis of Population with Applications to Business and Planning*. New York: Wiley.

Rosina, K. (2010). Európska mapa zastavaných plôch a jej validácia na území Slovenska. Diplomová práca. Univerzita Komenského v Bratislave, Prírodovedecká fakulta. 59s.

Sadahiro, Y. (1999). Accuracy of areal interpolation: a comparison of alternative methods. *Journal of Geographical Systems*, 1, 323–346.

Steinnocher, K., Kaminger, I., Köstl, M. Weichselbaum, J. (2010). Gridded Population – new data sets for an improved disaggregation approach. EFGS 2010 conference, Tallin, Estonia 6th October 2010.

Tapp, A. F. (2010). Areal Interpolation and Dasymetric Mapping Methods Using Local Ancillary Data Sources. *Cartography and Geographic Information Science*, Vol. 37, No. 3, July 2010, pp. 215-228.

Tobler, W. R. (1979). Smooth pycnophylatic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, 519–530.

Tralli, D. M., Blom, R. G., Zlotnicki, V., Donnellan, A., & Evans, D. L. (2005). Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(4), 185–198.

Vinkx, K., & Vise'e, T. (2008). Usefulness of population files for estimation of noise hindrance effects. ICAO Committee on Aviation Environmental Protection. CAEP/8 Modelling and Database Task Force (MODTF). 4th Meeting. Sunnyvale, USA, pp 20–22, February 2008.

Voss P. R., Long D. D., and Hammer R. B. (1999). When Census Geography Doesn't Work: Using Ancillary Information to Improve the Spatial Interpolation of Demographic Data. Madison, WI, University of Wisconsin-Madison, Center for Demography and Ecology Working Paper. No. 99-26.

Wu, C. and Murray, A.T. (2007). Population Estimation Using Landsat Enhanced Thematic Mapper Imagery. *Geographical Analysis* 39 (2007) 26–43.