

**AREÁLOVÁ TRANSFORMÁCIA
GEOGRAFICKÝCH DÁT:
PRINCÍPY, METÓDY A APLIKÁCIA**

AREAL TRANSFORMATION OF GEOGRAPHICAL DATA:
PRINCIPLES, METHODS AND APPLICATION

MICHALA
SLÁDEKOVÁ MADAJOVÁ
PAVOL HURBÁNEK

GEOGRAPHIA SLOVACA

Geographia Slovaca je recenzovaný vedecký časopis, ktorý uverejňuje monografie a monotematické súbory príspevkov z rôznych oblastí geografie, environmentalistiky a regionálnej vedy. Geographia Slovaca vychádza v Geografickom ústave SAV od roku 1992.

Geographia Slovaca is a reviewed journal bringing monographs and monothematic sets of contributions from all geographical sciences including environmental and regional sciences. Geographia Slovaca is issued at the Institute of Geography, SAS, since 1992.

Hlavný redaktor / Editor-in-Chief

Vladimír Ira – GgÚ SAV Bratislava

Redaktor / Editor

Martin Šveda – GgÚ SAV Bratislava

Redakčná rada / Editorial Board

Matej Gabrovec – GIAM ZRC SAZU Ljubljana, Mikuláš Huba – GgÚ SAV Bratislava, Karel Kirchner – Ústav geoniky AV ČR Brno, Éva Kiss – MTA CSFK FI Budapest, Tomasz Komornicki – IGI PAN Warszawa, Milan Lehotský – GgÚ SAV Bratislava, Anton Michálek – GgÚ SAV Bratislava, Ján Oťaheľ, – GgÚ SAV Bratislava, Peter Podolák – GgÚ SAV Bratislava

Redakcia / Editorial Office

Geografický ústav SAV, Štefánikova 49, 814 73 Bratislava, Slovenská republika (Slovak Republic)
tel: +421 2 5751 0210, e-mail: geogira@savba.sk

GEOGRAPHIA SLOVACA 32

Sládeková Madajová, Michala¹ – Hurbánek, Pavol^{1,2}

Areálová transformácia geografických dát: princípy, metódy a aplikácia

Areal Transformation of Geographical Data: Principles, Methods and Application

¹ Geografický ústav Slovenskej akadémie vied, Štefánikova 49, 814 73 Bratislava /
Institute of Geography, Slovak Academy of Sciences, Stefanikova 49, 814 73 Bratislava

² Katedra geografie, Pedagogická fakulta, Katolícka univerzita v Ružomberku, Hrabovská cesta 1,
034 01 Ružomberok / Department of Geography, Faculty of Education, Catholic University in Ruzomberok,
Hrabovska cesta 1, 034 01 Ruzomberok

Recenzenti / Reviewers

Doc. RNDr. Marián Halás, PhD.

Mgr. Ladislav Novotný, PhD.

Grafická úprava: Erika Mészárosová

Návrh a grafické spracovanie obálky: Martin Šveda

Tlač: Romulus Beňo JUNIOR PRESS, Častá

Jazyková úprava: Rút Facunová

Náklad: 100 výtlačkov

Geografický ústav SAV, 2016

ISSN: 1210-3519

ISBN: 978-80-89548-03-3

GEOGRAPHIA SLOVACA

32 – 2016

Sládková Madajová, Michala – Hurbánek, Pavol

**AREÁLOVÁ TRANSFORMÁCIA GEOGRAFICKÝCH DÁT:
PRINCÍPY, METÓDY A APLIKÁCIA**

**Areal Transformation of Geographical Data: Principles, Methods
and Application**

**GEOGRAFICKÝ ÚSTAV
SLOVENSKEJ AKADÉMIE VIED
BRATISLAVA**

OBSAH

POĎAKOVANIE	5
ÚVOD	7
1. VÝSKUM V OBLASTI AREÁLOVEJ TRANSFORMÁCIE GEOGRAFICKÝCH DÁT	10
1.1 Teoretická báza skúmanej problematiky, vymedzenie základných pojmov	13
1.2 Metodologická báza skúmanej problematiky	16
1.2.1 Klasifikácia metód areálovej transformácie geografických dát	31
1.2.2 Stručná charakteristika a všeobecné porovnanie základných metód areálovej transformácie geografických dát	35
2. APLIKAČNÁ ČASŤ	51
2.1 Výber metód areálovej transformácie geografických dát	52
2.2 Dáta použité na testovanie vybraných metód areálovej transformácie	54
2.3 Implementácia vybraných metód areálovej transformácie	57
2.3.1 Priestorové váženie	58
2.3.2 Binárna dazymetrická metóda	58
2.3.3 Metódy cestnej siete – metóda dĺžky a typov ciest	62
2.3.4 Geografické konverzné tabuľky	65
3. VÝSLEDKY A DISKUSIA	72
3.1 Spôsob hodnotenia výsledkov metód areálovej transformácie	72
3.2 Porovnanie výsledkov vybraných metód	74
3.2.1 Miery globálnej vhodnosti a histogramy	74
3.2.2 Priestorová vizualizácia chýb	88
3.3 Celkové zhodnotenie výsledkov vybraných metód areálovej transformácie geografických dát	92
ZÁVER	98
LITERATÚRA	102
SUMMARY	108

CONTENT

ACKNOWLEDGEMENT	5
INTRODUCTION	7
1. RESEARCH IN THE SUBJECT OF AREAL TRANSFORMATION OF GEOGRAPHICAL DATA	10
1.1 Theoretical base of the subject, definition of basic concepts	13
1.2 Methodological base of the subject	16
1.2.1 Classification of methods for areal transformation of geographical data	31
1.2.2 Brief description and general comparison of basic methods for areal transformation of geographical data	35
2. APPLICATION	51
2.1 Selection of methods for areal transformation of geographical data	52
2.2 Data used for assessment of the selected methods for areal transformation	54
2.3 Implementation of the selected methods for areal transformation	57
2.3.1 Areal weighting method	58
2.3.2 Binary dasymetric method	58
2.3.3 Road network methods – methods of road lengths and types	62
2.3.4 Geography conversion tables	65
3. RESULTS AND DISCUSSION	72
3.1 Means for assessment of results of the selected methods	72
3.2 Comparison of results of the selected methods	74
3.2.1 Measures of global fit and histograms	74
3.2.2 Spatial error visualisation	88
3.3 Overall assessment of results of the selected methods	92
CONCLUSION	98
REFERENCES	102
SUMMARY	108

POĎAKOVANIE

Základom monografie je dizertačná práca „Harmonizácia a areálová transformácia geografických dát: princípy, metódy a aplikácia na území Slovenska“, ktorú M. Madajová úspešne obhájila dňa 29. septembra 2011 na Katedre regionálnej geografie, ochrany a plánovania krajiny Prírodovedeckej fakulty Univerzity Komenského v Bratislave pod vedením prof. RNDr. Antona Bezáka, DrSc. Týmto autorka ďakuje prof. RNDr. Antonovi Bezákovi, DrSc. za vedenie dizertačnej práce a cenné rady, ktoré veľkou mierou prispeli k formovaniu jej vedeckej osobnosti. Publikovanie časti práce v podobe predloženej monografie a realizáciu ďalšieho výskumu v tejto oblasti umožnil grantový projekt VEGA č. 1/0275/13 „Tvorba, verifikácia a aplikácia priestorových modelov zaľudnenia a osídlenia na báze európskych služieb pre monitoring krajiny“.

Autori ďakujú oponentom, doc. RNDr. Mariánovi Halásovi, PhD. a Mgr. Ladislavovi Novotnému, PhD., ktorých pozorné čítanie a podnetné pripomienky nepochybne prispeli k vylepšeniu pôvodného rukopisu.

Autorka ďakuje svojmu manželovi za lásku, všestrannú podporu a trpezlivosť, ktoré jej vždy, a hlavne v hektickom období zostavovania tejto monografie, poskytol. Osobitné poďakovanie venuje mame za starostlivosť pri finalizácii práce a svojmu synovi za to, aký je.

ÚVOD

Jednou z kľúčových úloh regionálnogeografického výskumu je analýza vzťahov a vzájomných súvislostí medzi jednotlivými javmi a procesmi v priestore a čase. Pri pokusoch analyzovať zmeny v celku i v regionálnom priereze však možno naraziť na množstvo problémov, ktoré geografom znemožňujú korektnú analýzu, modelovanie, priame porovnanie či kombináciu dát z rôznych zdrojov. Jednotlivé charakteristiky, ako aj územné jednotky, za ktoré sú korešpondujúce dáta dostupné, sa môžu meniť v čase, alebo v mnohých prípadoch potrebujeme dáta analyzovať za iné zóny ako bežne dostupné administratívne/štatistické jednotky. Hoci sa na Slovensku údaje o obyvateľstve zbierajú za individuálne objekty, z rôznych dôvodov sa publikujú za istý stupeň priestorovej agregácie (obce, okresy ...). Okrem toho, že takéto územné jednotky sú pre mnohé geografické analýzy nevhodné, pre rôzne výskumné úlohy a predovšetkým pre plánovacie prax sú mnohokrát aj príliš „hrubé“. Problém priestorovej inkonzistencie výrazne vzrastá aj v oblasti interdisciplinárneho výskumu, napríklad pri snahe o integráciu demografických, ekonomických, environmentálnych či sociálnych dát dostupných za rôzne zonácie geografického priestoru.

Hoci sa možno väčšina geografov, ktorá sa s uvedenými problémami stretla, vzdala ich riešení, existuje niekoľko možností, ako ich prekonať. Preto, ak potrebujeme použiť a hlavne zjednotiť údaje z rôznych zdrojov, geografických jednotiek či časových období, je nevyhnutné pristúpiť k zosúladieniu týchto navzájom nekonzistentných entít, čo je proces, ktorý nazývame harmonizácia geografických dát. Predkladaná monografia sa sústreďí na časť tohto procesu, ktorá rieši problém priestorového nesúladienia – na tzv. areálovú transformáciu dát z jednej skupiny geografických jednotiek na dáta odhadnuté za inú skupinu geografických jednotiek. Keďže táto téma patrí na Slovensku medzi doposiaľ málo rozvinutú oblasť geografického výskumu, práca má za cieľ podrobnejšie predstaviť problematiku areálovej transformácie dát z navzájom nekompatibilných zonálnych systémov, vysvetliť jej princípy, sumarizovať, charakterizovať a klasifikovať rôzne prístupy a metódy umožňujúce prevod a odhad dát, otestovať vybrané metódy v slovenských podmienkach a navzájom porovnať presnosť ich výsledkov.

Vzhľadom na takto široko koncipovaný cieľ práca pozostáva z niekoľkých častí. Na začiatku sú predstreté teoretické a metodologické východiská skúmanej problematiky spolu s prehľadom základnej literatúry, ktoré slúžia na hrubé oboznámenie sa s témou

a s autormi a štúdiami, ktoré sa ňou aspoň okrajovo zaoberali. Osvojenie si základných pojmov, s ktorými pri procese areálovej transformácie geografických dát bežne narábame, je nevyhnutné, preto sme ich chápaniu a vysvetleniu venovali samostatnú podkapitolu. V metodologickej časti sú prezentované prístupy a metódy súvisiace s harmonizáciou údajov dostupných za rôzne geografické jednotky. Pokúsili sme sa podať ich doposiaľ chýbajúcu komplexnú klasifikáciu a základnú charakteristiku s cieľom čo najlepšie objasniť spôsob použitia jednotlivých metód a poukázať na rozdiely medzi nimi.

Z množstva metód používaných na areálovú transformáciu geografických dát z navzájom nekompatibilných územných systémov boli vybrané len niektoré, ktoré sme sa v rámci aplikačnej časti pokúsili implementovať a otestovať na území Slovenska. Výsledky základnej interpolačnej metódy, priestorového váženía, boli porovnané s jednoduchými metódami založenými na využití pomocných informácií, o ktorých sa predpokladá, že dokážu skvalitniť samotný proces odhadu dát a poskytnúť tak lepšie výsledky. Pre vzájomné porovnanie spoľahlivosti jednotlivých metód sme okrem zvyčajne zaužívanej odhadovanej premennej – počtu obyvateľov – odhadovali ďalšie dve charakteristiky – počet obyvateľov v predproduktívnom veku a počet ekonomicky aktívnych obyvateľov pracujúcich v poľnohospodárstve. Hodnoty premenných boli dostupné za jednotlivé okresy Slovenska a odhadnuté do systému funkčných mestských regiónov FMR 91-B (FMR) navrhnutého Bezákom (2000). Rozhodli sme sa použiť dve modelové situácie (okresy s hranicami viažucimi sa k rokom 1991 a 2001), v snahe čo najlepšie testovať vybrané metódy, keďže v roku 1996 nastala zmena územnosprávneho členenia Slovenskej republiky. Okrem nižšej úrovne (myslí sa analýza za každú situáciu zvlášť) prebiehalo teda testovanie aj na vyššej úrovni (okresy 1991 → FMR 91-B vs. okresy 2001 → FMR 91-B) s cieľom určiť, či a ako bude vplývať veľkosť/rozsah zdrojových a cieľových geografických jednotiek na proces areálovej transformácie dát. Pre systém funkčných mestských regiónov sme sa rozhodli z viacerých dôvodov. Na testovanie presnosti jednotlivých metód sme na jednej strane potrebovali regióny, ktorých hranice neboli kompatibilné s okresmi. Zároveň to však museli byť také územia, za ktoré sme poznali skutočné hodnoty odhadovaných dát. FMR predstavujú systém regiónov, za ktoré nemáme primárne k dispozícii žiadne údaje. Keďže však ide o systém, ktorý vznikol agregáciou z obcí a dáta za obce poznáme, bolo možné odhadnuté hodnoty premenných za FMR porovnať so skutočnými hodnotami (teda takými, ktoré získame agregáciou dát za obce a ktoré sú očistené od územných zmien, ktoré jednotlivé obce prekonali), a tým zhodnotiť presnosť jednotlivých metód.

Keďže transformácia geografických dát z navzájom nekompatibilných zonálnych systémov je proces, ktorý podlieha chybám a ani jedna metóda neposkytuje úplne presné výsledky, ale vždy ide len o odhady premennej, posledná časť monografie je venovaná pokusu o systematickejšie zhodnotenie výsledkov použitých metód. Presnosť odhadov jednotlivých metód sa hodnotila na základe tzv. globálnych mier vhodnosti, histogramov a priestorovej vizualizácie relatívnych chýb vyjadrených v percentách jednotlivo za každú premennú a použitú metódu. Okrem takéhoto hrubého porovnania, ktorého výsledkom je poradie úspešnosti jednotlivých metód odhadu dát dostupných za rôzne geografické jednotky, sme sa v rámci hodnotenia presnosti vybraných metód transformácie

geografických dát pokúsili odpovedať na otázky, čo je príčinou, prípadne čo ovplyvňuje presnosť jednotlivých metód, predovšetkým tých, ktoré využívajú pomocné údaje na skvalitnenie interpolačného procesu.

1. VÝSKUM V OBLASTI AREÁLOVEJ TRANSFORMÁCIE GEOGRAFICKÝCH DÁT

Pri skúmaní vývojových trendov určitých javov a procesov v jednotlivých regiónoch možno naraziť na viacero prekážok týkajúcich sa korektnej analýzy, modelovania či priameho porovnania údajov (v danom časovom období alebo z rôznych časových období). V priebehu sledovaného obdobia môže napríklad dôjsť k zmene priestorových jednotiek, za ktoré máme k dispozícii určité údaje (geografické jednotky prekonajú územné zmeny, alebo vzniknú úplne iné geografické jednotky), prípadne sa zmenia samotné pozorované údaje – sledovaná charakteristika zanikne, resp. vznikne nová charakteristika, môže nastať zmena jej definície, klasifikácie či aktuálnej dostupnosti. Mnohokrát však v rámci výskumných analýz jednoducho len potrebujeme analyzovať dáta za odlišné územné jednotky, ako sú bežne dostupné štatistické regióny, či kombinovať údaje z rôznych zdrojov, s čím sa neraz spája problém priestorového nesúladu. Hlavné faktory, ktoré znemožňujú analýzu, kombináciu či priame porovnanie dát možno podľa Martina et al. (2002) rozdeliť do štyroch kategórií: *územné usporiadanie, premenné, prostredie a prístupový mechanizmus*.

Územné usporiadanie sa týka priestorového rozdelenia krajiny s cieľom zhromaždenia a publikácie dát. Ako sme už spomenuli v úvode, na Slovensku sa množstvo dát humánno-geografickej povahy síce zbiera za individuálne objekty, z rôznych dôvodov (predovšetkým z hľadiska ochrany dôverných údajov) sa však publikujú za istý stupeň priestorovej agregácie: obce, okresy a pod.). Takéto bežne dostupné administratívne/štatistické jednotky sú však pre mnohé výskumné úlohy nevhodným konceptom. Spomeňme napríklad v geografickej literatúre častokrát rozoberaný problém ekologickej chyby a modifikovateľných územných jednotiek (Gehlke a Biehl 1934; Openshaw 1983), keď zmena hraníc a miera agregácie údajov významne ovplyvňuje výsledky priestorovej analýzy dát. Ďalšie problémy spojené s agregovanou povahou zonálnych údajov načrtáva Martin (1996) a v domácej literatúre sumarizuje Rosina et al. (2012).

Okrem administratívnych jednotiek, za ktoré sa spravidla zhromažďujú a publikujú údaje z jednotlivých sčítaní obyvateľstva a populačných registrov, existujú aj iné typy územných jednotiek, za ktoré sú dostupné dáta rôznej povahy (napríklad volebné a zdravotné obvody, regióny dochádzky, povodňové zóny či iné fyzickogeografické regióny a i.). Obzvlášť v prípade decíznej sféry, krízového manažmentu a vôbec pre plánovaciú prax je ča-

stokrát potrebné kombinovať údaje z rôznych zdrojov a výsledky analyzovať a publikovať za iné, častokrát podrobnejšie zóny, ako sú administratívne/štatistické jednotky.

Situácia sa komplikuje aj v prípade, keď síce narábame s tými istými geografickými jednotkami, ale tieto sa v priebehu nejakého časového obdobia (napríklad v medzicenzovom období) menia. Na Slovensku ide predovšetkým o problém porovnania územno-správneho členenia pred a po roku 1996. V iných krajinách je však táto problematika oveľa zložitejšia. Napríklad Veľká Británia prekonala po roku 1974 viac územných zmien (myslí sa zmien administratívnych hraníc) ako celý zvyšok Európy (Martin et al. 2002). Komplikovanú situáciu porovnávania dát z rôznych časových okamihov a za rôzne geografické jednotky tu sťažuje aj existencia viacerých typov územných jednotiek (historické grófstva, sčítacie dištrikty, volebné obvody a i.), z ktorých sa každý používa s iným cieľom. Preto v tejto krajine už niekoľko rokov prebieha snaha o vytvorenie jednotných geografických jednotiek, ktoré by prostredníctvom harmonizácie dát poskytovali kompatibilitu s tými, ktoré boli použité v predchádzajúcich obdobiach (napr. Martin 1989; Martin a Bracken 1991; Norman et al. 2001, 2003; Martin et al. 2002; Simpson a Yu 2003; Durham 2004; Norman 2006a,b,c). Pokusy o vytvorenie konzistentných geografických jednotiek v čase viedli k tvorbe rôznych stratégií štandardizácie priestorového systému. Tieto možno podľa Normana et al. (2003) rozdeliť do štyroch skupín:

1. Stratégia „Freeze history“ – stabilizácie histórie

Predstavuje akési „zafixovanie“ priestorového systému v nejakom časovom bode a následné systematické sledovanie jeho zmien tak, že dáta zhromaždené pre neskorší priestorový systém sa dajú spätne upraviť do pôvodných hraníc. Na takomto prístupe sú napríklad založené NUTS 5 v Eurostate.

2. Stratégia „Update“ – aktualizácia do súčasného zonálneho systému

Znamená aktualizáciu dát z predchádzajúcich priestorových systémov do súčasného systému.

3. Konštrukcia modelových zón

Táto stratégia je postavená na vytvorení tzv. modelových zón, ktoré sú založené na hraniciach, ktoré sú spoločné za rôzne časové obdobia, alebo sa vytvorí úplne nová skupina geografických jednotiek, ktoré sú užívateľovi prístupnejšie ako existujúca skupina jednotiek.

4. Geokódovanie individuálnych dát

Predstavuje prístup založený na geokódovaní dát (napr. na stupni domácností) zvyčajne prostredníctvom poštových smerových čísiel, adries alebo buniek pravidelného rastra. Táto stratégia umožňuje detailné modelovanie dát a odráža každú zmenu hraníc. Našla uplatnenie hlavne v škandinávskych krajinách, postupne sa však začína rozvíjať aj v ďalších európskych krajinách (Rakúsko, Slovinsko).

Ďalšími kategóriami, ktorým však nebudeme v predkladanej monografii venovať pozornosť, pre úplnosť ich však spomíname, sú **premenné, prostredie a prístupový mechanizmus**. *Premenné* sa týkajú samotných atribútových informácií, napr. údajov, ktoré sa zisťujú (napr. prostredníctvom cenzu), ako aj položených otázok a ich meniacemu sa významu. Jedna a tá istá premenná môže mať totiž v jednotlivých obdobiach rôzny význam (týka sa to napr. sociálnej klasifikácie alebo klasifikácie povolání). Okrem toho vedú zmeny v spoločnosti často k zmenám vo výskume a v politike záujmov. To si vyžaduje širší rámec

otázok, ktoré sú zahrnuté v sčítacom hárku. Tu sa ale naráža na napätie medzi potrebou zahrnúť nové témy, ktorých sociálny význam postupne vzrástol, a zachovaním kompatibility medzi jednotlivými zisťovaniami. Premenné úzko súvisia s ďalším, pomerne širokým rámcom faktorov, ktoré súhrnne nazývame *prostredie*. Vo všeobecnosti si pod týmto pojmom treba predstaviť sociálny a politický kontext, v ktorom sa zisťovanie uskutočňuje a ktorý má vplyv na publikované dáta (napríklad vládne vplyvy na sociálnu alebo ekonomickú štatistiku). Poslednou kategóriou, ktorá okrem iného ovplyvňuje aj kvalitu harmonizačného procesu, je *prístupový mechanizmus*. Týka sa organizačných a technických procedúr, ktoré využíva používateľ s cieľom prístupu k dátam.

Ak teda potrebujeme použiť a hlavne zjednotiť údaje z rôznych zdrojov, geografických jednotiek či časových období, je nevyhnutné pristúpiť k zosúladieniu týchto navzájom nekonzistentných entít, čo je proces, ktorý v zmysle Durham (2004), nazývame *harmonizácia geografických dát*. Možné spôsoby riešenia harmonizácie premenných (t. j. harmonizácia v zmysle atribútových informácií) sú načrtnuté v dizertačnej práci Madajovej (2011). V predkladanej monografii sa budeme venovať výlučne problému, ktorý Martin et al. (2002) identifikujú ako územné usporiadanie, a ktorý súvisí s priestorovou inkompatibilitou a s prevodom dát z jedného systému geografických jednotiek do druhého, odlišného systému geografických jednotiek. Situácie, keď územné usporiadanie vystupuje ako jedna z príčin potrieb harmonizácie geografických dát možno podľa autorov Simpson (2002) a Qui a Cromley (2013) klasifikovať takto:

- a) analýza časových radov na konzistentnej územnej báze (z angl. *temporal mismatch problem*), t. j. vytvorenie konzistentných zón v čase;
- b) analýza dát a prezentácia výsledkov za želaný zonálny systém podobnej mierky, resp. za také územia, ktoré sú pre výskum vhodnejšie a „prirodzenejšie“ (tzv. *alternative geography problem*);
- c) agregácia, resp. dezagregácia dát do jednotiek, ktoré sú dostatočne veľké/podrobné na to, aby poskytli spoľahlivé výsledky (z PŠČ záznamov do samosprávnych území alebo naopak do jednotiek s vysokým priestorovým rozlíšením – tzv. *small area problem*);
- d) spájanie rôznych zonálnych systémov, resp. kombinácia dát z rôznych zdrojov.
Vo všeobecnosti teda ide o problém analýzy údajov a ich odhadu:
 - z tých istých geografických jednotiek, ktoré sa ale zmenili v čase,
 - z rôznych zonálnych systémov do jedného spoločného systému,
 - zo štandardných do neštandardných jednotiek,
 - z neštandardných do štandardných jednotiek.

Za štandardné geografické jednotky považujeme administratívne (územnosprávne štatistické) jednotky, neštandardnými môžu byť rôzne analytické regióny, ako napríklad funkčné mestské regióny, regióny dochádzky, volebné a zdravotné obvody, regióny odvodené z poštových smerových čísiel, národné parky, povodňové zóny či iné fyzickogeografické regióny, alebo rôzne analytické zonácie vytvorené v prostredí geografických informačných systémov (bunky pravidelného rastra alebo bufferové zóny, ako napr. územie v dosahu 50 km od jadrovej elektrárne).

Skôr, ako sa budeme zaoberať samotným problémom a riešeniami, ktoré súvisia s harmonizáciou geografických dát v zmysle prevodu dát medzi rôznymi zonálnymi systémami, je potrebné osvojiť si niekoľko termínov a vysvetliť si ich význam.

1.1 Teoretická báza skúmanej problematiky, vymedzenie základných pojmov

Situáciu, keď máme k dispozícii údaje za taký systém geografických jednotiek, ktorý nekorešponduje s územnými jednotkami, za ktoré je potrebné dáta analyzovať, možno vo všeobecnosti označiť ako problém priestorového nesúladu (inkonzistentnosti). Transfer a odhad dát za takéto nekompatibilné územné jednotky je proces, ktorý býva v literatúre zvyčajne pomenovaný ako *areálová interpolácia*, resp. *problém areálovej interpolácie* (Goodchild a Lam 1980; Flowerdew a Openshaw 1989; Mrozinski a Cromley 1999; Langford et al. 2001; Lin et al. 2013). Jedine Durham (Durham 2004) používa termín *harmonizácia geografických dát* a označuje takto proces zoskupenia a transformácie dát z jednej skupiny geografických jednotiek na dáta odhadnuté pre inú skupinu geografických jednotiek. Môžu to byť tie isté geografické jednotky v rôznych časových okamihoch, ktoré prekonali zmenu hraníc, alebo to sú rôzne geografické jednotky v tom istom časovom okamihu. Prevod dát za rôzne zonálne systémy je jedným zo základných problémov aj v geografickom informačnom systéme. Niekedy sa v tomto prostredí označuje ako *problém polygónového vrstvenia* (Goodchild 1978).

Problém odhadu hodnoty premennej z v nejakom bode (x, y) , ktorý dáva premennej z známe hodnoty pri zvyčajne náhodne usporiadanom počte údajových bodov, sa podľa Goodchilda a Lama (1980) všeobecne chápe ako *priestorová interpolácia* (*spatial interpolation*). Situáciu, keď potrebujeme získať hodnoty premennej pre inú skupinu regiónov než za akú máme k dispozícii dané údaje, označujú títo autori, rovnako ako Lam (1983), Flowerdew a Openshaw (1987) a Flowerdew et al. (1991) za *problém územnej, resp. areálovej interpolácie* (*the areal interpolation problem*). Flowerdew a Green (1991) formulujú problém prevodu údajov za rôzne zonálne systémy ako problém odvodenia dát z jednej skupiny zón, ktoré poskytujú relevantné dáta, do inej skupiny zón a označujú ho ako *areálovú interpoláciu*. Podobne, Langford et al. (1991) definujú areálovú interpoláciu ako prevod dát z jednej skupiny geografických jednotiek do druhej skupiny prekrývajúcich sa a nehierarchizovaných zón. Naopak Turner a Openshaw (2001) nazývajú proces transformácie hodnôt premenných z jednej skupiny geografických jednotiek do hodnôt za inú skupinu geografických jednotiek ako jeden z *typov priestorovej interpolácie*. *Priestorovú interpoláciu* chápajú ako druh priestorového modelovania a definujú ju ako proces transformácie hodnôt priestorovej premennej dostupnej za určité body, línie alebo regióny do hodnôt tej istej premennej vzťahujúcej sa ale k odlišným bodom, líniam alebo regiónom. Pre úplnosť treba dodať, že títo autori zavádzajú ešte pojem tzv. *dezagregovanej priestorovej interpolácie*. Týmto pojmom označujú prípad, keď má jedna skupina geografických jednotiek (cieľová) vyššiu úroveň priestorového rozlíšenia ako druhá skupina geografických jednotiek (zdrojová). Myslí sa tým situácia, keď sú cieľové geografické jednotky oveľa menšie ako zdrojové jednotky. Odhad hodnôt premennej je tak výrazne zložitejší. S termínom priestorová dezagra-

gácia sa stretávame predovšetkým v súvislosti s prístupmi, ktoré transformujú populačné dáta do pravidelnej mriežky, resp. rastra (z angl. *grid*) s vysokým priestorovým rozlíšením (tvorba rastrových modelov rozmiestnenia obyvateľstva, napr. Dobson et al. 2000; Gallego 2010; Rosina et al. 2012).

Vzhľadom na istú rôznorodosť v používaní pojmov *harmonizácia geografických dát, priestorová a areálová interpolácia* v literatúre zaoberajúcej sa problematikou transferu dát z jednej skupiny geografických jednotiek do druhej, nekompatibilnej skupiny geografických jednotiek, je potrebné uviesť naše chápanie týchto procesov.

Pod pojmom **harmonizácia geografických dát** rozumieme proces zosúladenia určitým spôsobom nekonzistentných entít, pod ktorými si treba predstaviť tak atribútové, ako aj priestorové informácie. Môžu to byť teda nielen samotné atribútové informácie (štatistické dáta), ale aj geografické jednotky, za ktoré sú tieto dáta dostupné.

Priestorovú interpoláciu chápeme ako proces odhadu hodnoty premennej vzťahujúcej sa nielen k určitým regiónom, ale aj k bodom a líniám. Preto možno vyčleňovať akési typy priestorovej interpolácie, ktoré možno rozdeliť do niekoľkých rovín v závislosti od rôznych faktorov, ako napr. spracovaných údajov, použitých funkcií či predpokladov o danom území. Sú nimi: bodová vs. areálová, globálna vs. lokálna, deterministická vs. stochastická, exaktná vs. aproximujúca či súvislá vs. prerušená interpolácia (Goodchild a Kemp 1990). Väčšina interpolačných metód je zameraná skôr na spracovanie bodových údajov, a preto sa zvykne aj samotný pojem priestorovej interpolácie častokrát stotožňovať s tzv. bodovou interpoláciou. V práci budeme bližšie rozoberať problém odhadu dát za nekompatibilné geografické jednotky, t. j. za regióny, je teda vhodné podľa uvedenej klasifikácie označiť takto chápaný proces ako areálovú interpoláciu, prípadne ako problém areálovej interpolácie. **Areálovú interpoláciu** považujeme za druh priestorovej interpolácie. V súvislosti s metódami, ktoré sú schopné odhadnúť dáta z jedného systému geografických jednotiek do iného, odlišného územného systému, sme sa však rozhodli termín *areálová interpolácia* nahradiť termínom **areálová transformácia dát**, a to z viacerých dôvodov. Hoci sa v literatúre rozlišujú dva typy priestorovej interpolácie – bodová a areálová, už samotné metódy bodovej a areálovej interpolácie sú rôznorodé a k riešeniu problému areálovej interpolácie možno pristupovať nielen prostredníctvom využitia techník, ktoré sú založené na území (ploche), ale aj bodoch a líniách. Aby sme sa teda vyhli po prvé nejasnostiam s používaním pojmov *areálová interpolácia* ako taká a *metódy areálovej interpolácie*, ktoré nie sú v pravom zmysle slova založené len na území (ale môžu byť napríklad aj bodovo interpolačné a i.) a po druhé vzhľadom na to, že k riešeniu tejto problematiky možno pristupovať nielen prostredníctvom využitia rôznych priestorovo-interpolačných techník, ale aj princípov štatistického modelovania, rozhodli sme sa problém areálovej interpolácie nazvať všeobecnejšie – areálovú transformáciu a jej metódy následne nazvať ako **metódy areálovej transformácie geografických dát**. V prípadoch, keď sa budeme venovať metódam odhadu dát za nekompatibilné zonálne systémy, tak ako sú uvádzané v literatúre, alebo keď nebudeme uvažovať výlučne o štatistických metódach, možno termín *metódy areálovej transformácie* nahrádzať termínom *metódy areálovej interpolácie*.

Areálovú transformáciu geografických dát budeme teda považovať len za jeden z harmonizačných prístupov, ktorý rieši problém prevodu a odhadu údajov za navzájom odlišné územné jednotky.

Zdrojové a cieľové geografické jednotky, premenné

Spracovanie údajov z rôznych zdrojov (demografických, ekonomických, sociálnych, či environmentálnych dát) a/alebo časových období a s tým spojený problém kompatibility rôznych územných jednotiek si teda vyžaduje transformáciu dát z jedného systému geografických jednotiek do nejakého iného (pozmeneného alebo úplne odlišného) systému geografických jednotiek. V tejto súvislosti má zmysel hovoriť o jednej skupine geografických jednotiek ako o zdrojovej a o druhej skupine geografických jednotiek ako o cieľovej (Markoff a Shapiro 1973; Goodchild a Lam 1980).

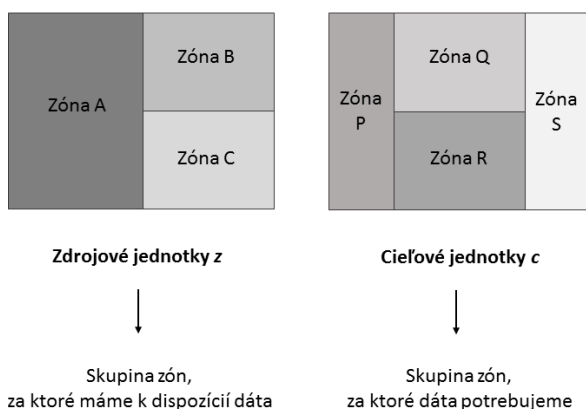
Zdrojové geografické jednotky sú územné jednotky – zóny, za ktoré máme k dispozícii údaje. Budeme ich označovať písmenom *z*.

Cieľové geografické jednotky sú územné jednotky – zóny, za ktoré tieto údaje potrebujeme. Označujeme ich písmenom *c* (obr. 1). Cieľovými geografickými jednotkami môžu byť:

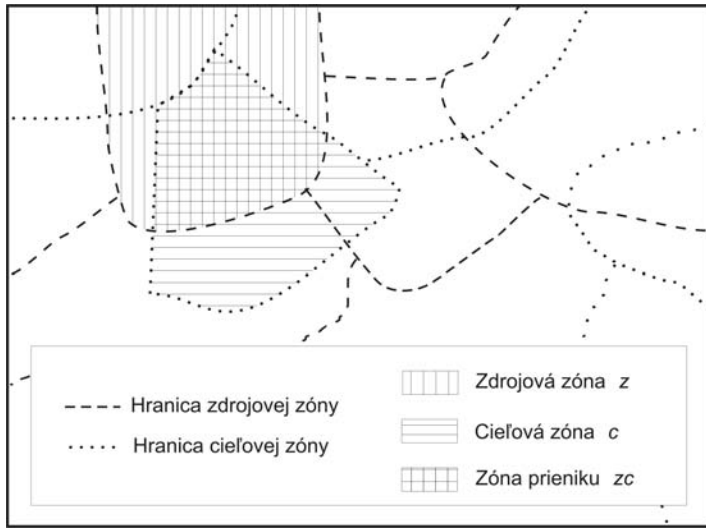
- geografické jednotky, ktoré sa zmenili v priebehu nejakého časového obdobia) alebo
- úplne iné geografické jednotky.

Je potrebné podotknúť, že pojem *geografická jednotka* častokrát nahrádzame termínom *zóna*, resp. *región*, ktorý používame na vyjadrenie bližšie nešpecifikovanej územnej jednotky (podobne termín *skupina*, resp. *systém geografických jednotiek* = *zonálny systém*).

Prekryvom zóny *z* a zóny *c* vznikne tzv. **zóna prieniku** z_c (obr. 2). Niekedy leží celá zdrojová zóna vnútri cieľovej zóny (alebo naopak), ale zvyčajne je každá zdrojová zóna rozdelená hranicami cieľovej zóny do niekoľkých zón prieniku a opačne. Územia týchto zón preto označíme ako A_z , A_c a A_{z_c} .



Obr. 1. Příklad zdrojových a cílových geografických jednotiek



Obr. 2. Príklad zón prieniku (upravené podľa Flowerdewa a Greena 1991)

Pojem **premenná** zastupuje ľubovoľný objekt (predmet, vlastnosť alebo vzťah) z nejakého súboru objektov. V našom ponímaní budeme pod premennou myslieť atribútovú informáciu geografických dát.

Ak premennú označíme Y , potom hodnota Y pre zónu z sa označuje ako y_z (táto je známa). Hodnota Y pre zónu c tak bude y_c (nepoznáme ju) a pre zónu prieniku y_{zc} .

Zvyčajne existuje medzi hodnotou y_c pre cieľovú zónu a hodnotami y_{zc} pre zóny prieniku jednoduchý vzťah. Problém odhadu hodnôt pre cieľové zóny sa potom redukuje na problém odhadu hodnôt zón prieniku (Flowerdew a Green 1994).

Ak využívame pomocné informácie o zdrojovej alebo cieľovej zóne, aby sme vylepšili proces odhadu, potom zavádzame najmenej jednu **pomocnú premennú** x_z dostupnú pre zdrojové zóny alebo x_c pre cieľové zóny.

1.2 Metodologická báza skúmanej problematiky

V súvislosti s prevodom dát za odlišné zonálne systémy možno rozlíšiť niekoľko typov území, pre ktoré potrebujeme údaje analyzovať. Môžu to byť:

- územia, ktoré sú dokonalou (úplnou) agregáciou jednej alebo viacerých zdrojových geografických jednotiek,
- územia, ktoré prekrývajú zdrojové geografické jednotky a nemôžu byť z nich priamo vystavané,
- územia, ktoré sú definované ako geometrické zóny,
- územia, ktoré majú pravidelný tvar (raster, grid).

Na základe toho možno rozlíšiť tzv.:

1. *Perfektnú (dokonalú, úplnú) agregáciu* – keď menšie zdrojové zóny presne zapadnú do väčších cieľových zón.

Na získanie údajov za cieľové zóny možno len jednoducho agregovať dáta za jednotlivé zdrojové zóny:

Zdrojové zóny	Cieľové zóny
1	A
2	B
3	A
4	B

Cieľové zóny	Zdrojové zóny
A	1,3
B	2,4

2. *Nedokonalú agregáciu* – keď zdrojové zóny presne nezapadnú do nových území, zdrojové a cieľové zóny sa prekrývajú.

Na odhad hodnôt chýbajúcich údajov sa používa tzv. váha, ktorá na základe nejakého kritéria udáva podiel zdrojovej jednotky, ktorý leží v cieľovej jednotke. Váhy sa vo všeobecnosti odvodzujú z pomocných informácií (v závislosti od použitej metódy), ktoré nejakým spôsobom súvisia s rozmiestnením danej premennej v zdrojových jednotkách. Z váh sa odvodnia údaje za cieľové zóny, zvyčajne nasledovne (viac v podkapitole 1.2.2 pri jednotlivých metódach):

Zdrojové zóny	Cieľové zóny	Váha
1	A	0,4
1	B	0,6
2	A	0,5
2	B	0,5

Cieľové zóny	Zdrojové zóny
A	0,4*zóna 1 + 0,5*zóna 2
B	0,6*zóna 1 + 0,5*zóna 2

3. *Geometrickú agregáciu* – keď je potrebné zistiť údaje pre geometricky definované územie (napr. územie v dosahu 50 km od jadrovej elektrárne).

Môže sa využiť prístup, pri ktorom sa vypočíta euklidovská vzdialenosť D_{ij} medzi ťažiskom cieľovej zóny i a ťažiskom každej zdrojovej zóny j podľa vzorca¹ :

$$D_{ij} = \sqrt{[(x_i - x_j)^2 + (y_i - y_j)^2]}$$

kde:

x_i, y_i = súradnice ťažiska cieľovej zóny i

x_j, y_j = súradnice ťažiska zdrojovej zóny j .

4. *Dezagregáciu* – keď sú cieľové zóny oveľa menšie ako zdrojové, napr. keď dáta pre rozdeľujeme do buniek rastra s podrobným priestorovým rozlíšením (100 x 100 m, 1 km x 1 km).

¹ http://www.geog.leeds.ac.uk/courses/postgrad/geog5105/unit8/Unit8_2004/GEOG5105Unit8_2004.pdf

Pravidelná mriežka môže slúžiť aj ako tzv. kontrolná zóna, ktorou sa prekryjú zdrojové geografické jednotky. Následne sa dáta zo zdrojových zón v závislosti od použitej metódy vhodne prerozdedia (dezagregujú) do jednotlivých buniek rastra. Z takýchto rastrov je následne možné jednoducho aproximovať hodnoty pre ľubovoľné polygóny (cieľové geografické jednotky).

Každý z týchto prístupov používa rôzne metódy a techniky, ktoré umožňujú odhadnúť dáta za odlišné zonálne systémy. Ako sumarizujú Martin et al. (2002), prístupy na spájanie dát dostupných za rôzne geografické jednotky sú založené buď na premodelovaní dát využitím priestorovej interpolácie na prenos dát, alebo na použití konverzných tabuliek na najvhodnejšie pridelenie dát z jednej skupiny geografických jednotiek do druhej. Prvý prístup je podľa nich vhodnejší pri práci s rozlohou menšími geografickými jednotkami, druhý sa zvykne používať pri väčších geografických jednotkách.

Geografické konverzné tabuľky (GKT) predstavujú techniku, ktorá našla svoje uplatnenie predovšetkým vo Veľkej Británii pri analýze intercenzových zmien a opakovaných problémoch porovnávania demografických údajov. Všeobecný pojmový rámec pre GKT a konverziu dát medzi rôznymi geografickými jednotkami poskytuje Simpson (2002). Opisuje rôzne možnosti konštrukcie geografických konverzných tabuliek a dopĺňa ich prostredníctvom rôznych príkladov. Detailnejšie modelovanie chýbajúcich údajov podľa neho umožňuje metóda lokalizácie poštových smerových čísiel a ich následného využitia v rámci GKT. Konverzným tabuľkám a PSČ záznamom venuje pozornosť viacero autorov. Za všetkých spomenieme napr. Martina (1992), Wilsona a Reesa (1998, 1999), Boyla a Fenga (2001), Normana et al. (2001, 2003), Normana (2006a, 2006b, 2006c) a Simpsona a Yua (2003). Martin et al. (2002) obnovili, resp. sprístupnili približne 100 kľúčových cenových premenných pre široký rámec geografických jednotiek vo Veľkej Británii prostredníctvom GKT umožňujúcich konverziu údajov o obyvateľstve z rokov 1971 a 1991 do územných jednotiek z roku 1981. Na ich základe sa Durham (2004) pokúša definovať systematickú metodológiu, ktorá by dovoľila porovnať demografické údaje zo sčítania v roku 2001 s ostatnými britskými censami. Opisuje a vysvetľuje základné pojmy a princípy harmonizačného procesu, zaoberá sa zhodnotením doteraz vo Veľkej Británii použitých metód harmonizácie geografických dát a podrobnejšie opisuje metódu geografických konverzných tabuliek, ktoré sa použili pri vytvorení On-line demografického atlasu. Durham použila 88 premenných z cenov 1971, 1981 a 1991 a pripojila k nim údaje z roku 2001. Všetky tieto dáta potom prepojila so sériou generalizovaných hraníc a kartogramov a vsunula ich do on-line atlasu, ktorý na základe softvéru CommonGIS poskytuje cez webové rozhranie interaktívne výskumné analýzy priestorovo referencovaných dát. Do budúcnosti autorka plánuje rozvinúť tzv. „nástroj geografickej konverzie“, ktorý by sa následne použil na dokončenie atlasu. Atlas sa rozšíri o ďalšie geografické jednotky, ako zdravotné obvody a územia dochádzky za prácou (tzv. TTWA – Travel-To-Work Areas). Tieto jednotky však v priebehu sledovaného obdobia prekonal viaceré zmeny hraníc a ich hranice sú tak nekonzistentné s hranicami definovanými v On-line demografickom atlase pre roky 1971 – 1991. Analýza zmien v priebehu rokov 1971 – 2001 na tomto stupni geografických jednotiek si teda vyžaduje ďalšiu konverziu dát (Durham 2004).

Základné problémy priestorovej interpolácie dát načrtávajú Goodchild a Lam (1980) a Lam (1983). Podľa autorov Goodchild a Lam (1980) existuje množstvo prístupov, ktoré sú vhodné na riešenie problému areálovej interpolácie dát. Základnou a najčastejšie používanou metódou, je metóda priestorového váženia/prekryvu. Detailne sa ňou zaoberajú Markoff a Shapiro (1973) a Goodchild a Lam (1980). Ide o jednoduchú kartografickú metódu založenú na rozlohe územia prekryvu zdrojových a cieľových geografických jednotiek. Táto metóda predpokladá homogénne rozmiestnenie záujmovej premennej v zdrojových zónach, čo sa považuje za jej najväčšiu nevýhodu. Napriek tomu sa vyskytuje vo väčšine prác zaoberajúcich sa problematikou odhadu hodnôt vybranej premennej za odlišné zonálne systémy vďaka jednoduchej teórii, nenáročnej aplikácii a ľahkej implementácii v prostredí GIS.

Predpoklad homogénneho rozmiestnenia hodnôt premennej v zdrojových zónach je pre väčšinu dát humánno-geografickej povahy neprijateľný. Preto Tobler (1979) navrhol tzv. pyknofylaktickú metódu, ktorá odhaduje spojitý povrch premennej (konkrétne hustoty zaľudnenia) v uzloch jemnej mriežky na základe informácie o hustote zaľudnenia zdrojových jednotiek, pričom berie do úvahy efekt prilahlých zón. Vyhladený povrch hustoty zaľudnenia sa nakoniec použije na výpočet hodnôt tejto premennej v cieľových jednotkách. Metóda zároveň zachováva celkový objem informácie zdrojových zón. Túto podmienku pomenoval Tobler (1979) ako pyknofylaktickú vlastnosť. Lam (1983) v tejto súvislosti používa termín „objem/obsah, resp. rozsah zachovávací“ (z angl. *volume preserving*). V prípade areálovej interpolácie sa pojem „objem zachovávací“ vzťahuje na zachovanie celkovej hodnoty premennej každej zdrojovej zóny. Ak je napríklad interpolovanou premennou počet obyvateľov, potom sa musí na základe tohto kritéria v procese interpolácie zachovať súčet počtu obyvateľov zdrojových zón.

Toblerovu metódu vylepšil Rase (2001), keď namiesto tradičnej pravouhlej mriežky použil trojuholníkovú nepravidelnú sieť (TIN). Metóda pracuje podobne ako pyknofylaktický prístup, a to tak, že z pôvodných hraníc zdrojových jednotiek sa najskôr vytvorí TIN a následne sa iteratívnym procesom interpoluje vyhladený povrch. Proces vyhladenia a distribúcie dát sa opakuje, pokiaľ sa nedosiahne prah úplného vyhladenia, alebo pokiaľ sa nepresiahne maximálny počet opakovaní. Hlavným dôvodom modifikácie Toblerovej metódy bola možnosť zachovania pôvodných hraníc polygónov, ktoré pravouhlá mriežka nezachováva, zatiaľ čo flexibilnejší model TIN áno, pretože je schopný použiť body geografických hraníc ako geometrické vrcholy (Hawley a Moellering 2005). Hoci sa táto metóda v porovnaní s Toblerovou pyknofylaktickou metódou ťažšie aplikuje, je schopná vyvarovať sa chýb z konvertovania zdrojových polygónov do pravidelnej mriežky.

Pyknofylaktická metóda a metóda priestorového váženia teda patria do skupiny interpolačných metód, ktoré Lam (1983) nazval ako objem zachovávacíe metódy. K objem nezachovávacím metódam naopak zaradil tzv. bodové (alebo na bode založené, z angl. *point based*) priestorovo-interpolačné metódy, ktoré pracujú s bodmi ako s náhradami za zdrojové a/alebo cieľové geografické jednotky. Spoliehajú sa na matematické metódy, ktoré interpolujú hodnoty zo známych reprezentatívnych bodov (napr. centroidov) na základe funkcie, pri ktorej sa predpokladá, že dokáže popísať priestorovú diferenciáciu skrytú pod zdrojovými zónami (tzv. *underlying distribution*). Existujú však práce, v ktorých sa

využívajú zástupné body za jednotlivé zóny a pritom sa snažia zachovať pyknofylaktickú vlastnosť. Sú nimi štúdie nasledovných autorov: Martin (1989), Bracken (1991), Martin a Bracken (1991), Martin (1996), Okabe a Sadahiro (1997) a Kyriakidis (2004). Martin (1989) pracuje so štvorcovou sieťou a opisuje jednoduchý algoritmus, ktorý využíva ťažiská zdrojových jednotiek s plošnou funkciou na pridelenie počtu obyvateľov susedným štvorcovým mriežky. Tzv. kernel interpoláciu, ktorá je akousi obmenou interpolačnej metódy inverzných vzdialeností (modeluje klesajúcu hustotu zaľudnenia s rastúcou vzdialenosťou od centroidu) navrhli Bracken (1991) a Martin a Bracken (1991). Prechodnú metódu od bodových k areálovým interpolačným technikám predstavuje metóda „bod v polygóne“, ktorú použili vo svojej práci Okabe a Sadahiro (1997). Táto technika umiestňuje ťažisko zdrojových zón do hraníc cieľových zón. Kyriakidis (2004) naopak uvažuje o technike „územie do bodu“ a využíva geoštatistickú interpolačnú metódu kriging. Je to do budúcnosti pomerne sľubná technika, nakoľko kriging využíva aj rôzne pomocné informácie (Hawley a Moellering 2005).

Spomenuté metódy sa zvyknú označovať ako jednoduché areálovo interpolačné metódy. Alternatívou k nim sú prístupy, ktoré sa snažia lepšie porozumieť rozmiestneniu premennej, a to tak, že využívajú pomocné informácie na skvalitnenie výsledkov procesu odhadu dát. Väčšina autorov ich označuje ako tzv. inteligentné interpolačné metódy (Flowerdew a Green 1989, 1992 a 1994; Flowerdew et al. 1991; Sadahiro 2000b; Eicher a Brewer 2001; Mennis a Hultgreen 2005; Mennis a Hultgreen 2006a a 2006b; Langford 2007; Lin et al. 2013). Pomocné informácie sú, resp. by mali byť také údaje, o ktorých sa predpokladá, že odrážajú priestorové rozmiestnenie mapovaných premenných, prípadne s nimi nejakým spôsobom súvisia. Ako vhodné pomocné informácie vzťahujúce sa k obyvateľstvu (keďže to je premenná, s ktorou najčastejšie narába väčšina autorov zaoberajúcich sa problematikou prenosu dát medzi rôznymi zonálnymi systémami) sa najčastejšie používajú dáta z diaľkového prieskumu Zeme (DPZ) a z nich odvodené mapy využitia zeme a krajiny pokrývky, či už v rastrovej (Langford a Unwin 1994; Fisher a Langford 1996; Cocking et al. 1997; Yuan et al. 1997; Harvey 2000 a 2002b; Eicher a Brewer 2001; Gallego a Pedell 2001; Holt et al. 2004; Reibel a Agrawal 2007) alebo vektorovej podobe (Eicher a Brewer 2001; Mennis 2003; Bielecka 2005; Mennis a Hultgreen 2005; Mennis a Hultgreen 2006a a 2006b). Postupne sa objavilo niekoľko spôsobov realizácií metód odhadu dát s využitím údajov z DPZ ako pomocných informácií. Najvýznamnejšou technikou je tá, ktorá je založená na princípe dazymetrickeho mapovania. Dazymetrické mapovanie je formou kartografickej reprezentácie, ktorá predstavuje alternatívu ku klasickému kartogramu. Ten sa totiž z viacerých hľadísk ukázal ako nevhodný na detailnejšie priestorové analýzy dát humánno-geografickej povahy. Dazymetrická mapa je podľa Robinsona et al. (1984) taká mapa, v ktorej sú hranice zobrazovacích jednotiek nezávislé od sčítacích jednotiek. Moderná kartografia definuje dazymetrickú mapu ako mapu, ktorá zobrazuje dáta úplným (vyčerpávajúcim) rozdelením územia do zón, ktoré odzrkadľujú zmeny základného štatistického povrchu (Mennis 2006). Dazymetrická mapa našla uplatnenie predovšetkým v súvislosti s mapovaním obyvateľstva (termín dazymetrický vychádza z angl. *density measuring* a vzťahuje na meranie hustoty zaľudnenia, Langford 2006). Ako prvý vytvoril dazymetrickú mapu obyvateľstva európskej časti Ruska v roku 1922 ruský kartograf Se-

menov Ťan-Šanskij (Bielecka 2005; Mennis a Hultgreen 2005). Častokrát a mnohými autormi však býva za tvorcu dazymetrického mapovania nesprávne označovaný americký geograf John K. Wright, asi preto, že ho spopularizoval a ako prvý použil v USA. Wright (1936) rozvinul metódu mapovania hustoty zaľudnenia s využitím topografických máp ako pomocných informácií. Kým v minulosti bol dazymetrický prístup menej rozvinutý, postupne s rozvojom technológií, ako aj samotných geografických informačných systémov (GIS) a s tým spojenou lepšou dostupnosťou digitálnych dát sa čoraz viac rozvíjal. Práve dazymetrická mapa sa tak stala nielen jedným z riešení problému zobrazovania populačných dát agregovaných do priestorových jednotiek, ale všeobecné princípy dazymetrického mapovania sa začali aplikovať aj na problém areálovej transformácie geografických dát z navzájom odlišných zonálnych systémov. V oboch týchto prípadoch sa tento prístup zvykne označovať ako dazymetrická metóda (ako metóda dezagregácie obyvateľstva, aj ako metóda riešajúca problém areálovej interpolácie). Treba však podotknúť, že dazymetrická metóda býva častokrát stotožňovaná s tými dezagregačnými/intepolačnými metódami, ktoré ako pomocné informácie používajú práve dáta z topografických máp a údajov z DPZ. Správne by tak mali byť označené všetky metódy, ktoré sa snažia realistickejšie rozmiestniť zdrojové dáta do jemnejších zón a odvodiť hustotu zaľudnenia týchto regiónov prostredníctvom akýchkoľvek pomocných informácií, ktoré súvisia s rozmiestnením obyvateľstva. Preukázalo sa napríklad, že na počet a rozmiestnenie obyvateľstva vplyva sklon reliéfu, vzdialenosť od centra, prístup k dopravným systémom a mnoho ďalších faktorov (Liu a Clarke 2002). Zahnutie viacerých doplnkových informácií do procesu odhadu danej premennej nielenže zvyšuje komplexnosť modelov, ale zároveň zlepšuje aj ich presnosť a odolnosť. Je však potrebné zdôrazniť, že výber jednotlivých pomocných premenných si vyžaduje vysoké teoretické znalosti (napríklad z urbankej geografie). Príkladom sú práce využívajúce napr. nočné satelitné snímky (Sutton et al. 2001; Zandbergen a Ignizio 2010), dáta o nepriepustnosti povrchu (Sutton et al. 2009; Zandbergen a Ignizio 2010; Zandbergen 2011; Rosina et al. 2012), dáta z oblasti dopravy (ako napr. sieť ulíc, ciest a cestných uzlov: Xie 1995; Mrozinski a Cromley 1999; Voss et al. 1999; Reibel a Bufalino 2005; Bentley et al. 2013), parcelové dáta o využití zeme (Maantay et al. 2007), adresné body (Tapp 2010; Zandbergen 2011), budovy (Xie 2006; Wu et al. 2008) či v poslednom období aj lidarové dáta (Sridharan a Qui 2013).

Existuje niekoľko implementácií dazymetrickej metódy, bližšie sa im budeme venovať v podkapitole 1.2.2. Najjednoduchším prístupom je binárna dazymetrická metóda, v ktorej sa pomocné zóny klasifikujú buď ako zaľudnené, alebo nezaľudnené. V prípade použitia dát o krajinej pokrývke (CLC dáta) sa ako obývané najčastejšie označujú umelé povrchy alebo trieda urbanizovanej zástavby. Obyvateľstvo sa jednoducho redistribuje len do zaľudnených území (Langford a Unwin 1994; Fisher a Langford 1995; Cocking et al. 1997; Eicher a Brewer 2001; Holt et al. 2004; Mennis a Hultgreen 2005). Ďalšie prístupy rozlišujú viac kategórií pomocných zón, ako je len uvažovanie o obývatel'ných a neobývatel'ných územiach (v prípade CLC dát sú to jednotlivé triedy, resp. skupiny tried krajinej pokrývky). Takýto postup si vyžaduje špecifikovanie vzťahu medzi jednotlivými pomocnými zónami a odhadovanou premennou (CLC triedami a hustotou zaľudnenia). Hustota zaľudnenia jednotlivých dazymetrických zón môže byť určená viacerými spôsobmi: a) a priori (subjektívne,

napr. na základe terénneho výskumu alebo podrobných znalostí územia: Holloway et al. 1997; Eicher a Brewer 2001), b) prostredníctvom tréningových území a tzv. smplovania (ako tréningové územia sa napr. použijú tie zdrojové zóny, ktoré obsahujú len obývané CLC triedy, príp. sa stanoví hranica pre dominantné CLC triedy a takto vypočítaná hustota zaľudnenia sa použije ako relatívna váha na redistribúciu obyvateľstva: napr. Mennis 2003; Mennis a Hultgreen 2005; Mennis a Hultgreen 2006a a 2006b), alebo c) využitím rôznych štatistických modelov (regresná analýza: Fisher a Langford 1995; Yuan et al. 1997; Harvey 2000 a 2002b; Gallego a Peedel 2001; Bielecka 2005; Reibel a Agrawal 2007, technika EM algoritmu: Flowerdew a Green 1989, 1991, 1992, 1994).

Princípy štatistického modelovania sa na odhad počtu obyvateľov zvykli aplikovať už skôr, predovšetkým s cieľom hľadania určitej alternatívy k získavaniu údajov z populačných cenzov. Motiváciou bola hlavne snaha o odstránenie nedostatku dát, keďže populačné cenzy sa spravidla uskutočňujú len každých desať rokov. Rôzne štatistické prístupy sa používali aj na kontrolovanie spoľahlivosti cenzových výpočtov. Implementácia takýchto metód však bola časovo, dátovo aj technicky pomerne náročná (Wu et al. 2005; Kim 2006). Pri predikcii počtu obyvateľov sa od polovice 50. rokov 20. storočia ukázali ako mimoriadne vhodné práve údaje z DPZ. Postupne sa zaviedli rôzne metódy štatistického modelovania na odhad počtu obyvateľov v rôznych mierkach s rôznymi typmi snímok. Viacerí autori používajú na odhad počtu obyvateľov štatistický prístup založený na korelácii medzi počtom obyvateľov a bytovými jednotkami. Prvý, kto navrhol použiť počty bytových jednotiek získaných z leteckých snímok na odhad počtu obyvateľov, bol Green (1956). Prvý, ktorý túto metodológiu aj aplikoval, bol Porter (1956). Neskôr sa viacerí autori pokúsili identifikovať rôzne typy bytových jednotiek s cieľom získať presnejší podiel osôb na bytovú jednotku a tak vypočítať priemernú hustotu zaľudnenia za rôzne typy bytových jednotiek (Dueker a Horton 1972; Hsu 1973; Lo a Chan 1980) Ako uvádzajú Wu et al. (2005), v minulosti neexistoval efektívny spôsob automatického výberu obývaných bytových jednotiek či budov. Výskumníci sa spoliehali na manuálnu identifikáciu a výpočet bytových jednotiek z fotografií vysokého rozlíšenia. Táto vizuálna interpretácia však bola pracná a časovo náročná. Na zautomatizovanie časovo náročnej procedúry spočítavania bytových jednotiek použil Lo (1989) rastrový prístup a vypočítal maximálny možný výskyt bytových jednotiek v každej bunke mriežky vzhľadom na veľkosť bytovej jednotky. Na základe toho dokázal odhadnúť percento výskytu obývaných budov v každej bunke mriežky. V súčasnosti je však už možná automatická extrakcia bytových jednotiek zo satelitných snímok.

Okrem počtu bytových jednotiek podľa Lo (1986) rozlišujeme tri ďalšie prístupy, na základe ktorých je možné odhadnúť počet obyvateľov prostredníctvom údajov z DPZ. Sú nimi rozloha urbanizovaných území, rozloha rôznych tried využitia zeme a automatická analýza snímok. Vzťahy medzi mestskými plochami a veľkosťou populácie vo viacerých amerických mestách študoval Nordbeck (1965). Zistil, že zastavaná plocha sídla je úmerná počtu obyvateľov umocnenému na nejaké číslo (exponent b). Prvý, ktorý použil satelitné snímky na štúdium vzťahu medzi obyvateľstvom a mestskými plochami, bol Tobler (1979). Skúmal obyvateľstvo viacerých svetových miest a zistil, že ak majú mestá kruhový pôdorys a ak sa ich tvar výrazne v čase nemení, potom korelácia medzi polomerom a počtom obyvateľov nadobúda hodnotu $r = 0,87$ alebo vyššiu. Výsledky jeho štúdie taktiež naznačili, že

hodnoty korelačných koeficientov a exponentov b sú podobné pre mestá v USA, Kanade a Švédsku. Mestá v Japonsku a delte Nílu majú naopak koeficienty a exponenty, ktoré odzrkadľujú hustotu a kompaktnú štruktúru sídiel v Ázii a Strednom východe (Wu et al. 2005). So zvyšujúcou sa dostupnosťou a skvalitnením spracovania snímok z DPZ sa viacero autorov pokúsilo efektívnejšie študovať vzťahy medzi obyvateľstvom a mestskými plochami. Na základe výskumu miest v Číne našli Lo a Welch (1977) korelačný koeficient $r = 0,82$ a vyšší medzi obyvateľstvom a klasifikovanými mestskými plochami ako modifikovateľnú funkciu vzťahu Nordbecka (1965). Tento tzv. model alometrického rastu opisuje relatívnu mieru rastu obyvateľstva, ktorá je proporcionálna relatívnej miere rastu obývaného územia. Objavili sa aj štúdie, ktoré skúmali intenzitu mestského osvetlenia v noci ako indikátor populačnej veľkosti (Prosperie a Eyton 2000).

Najrozšírenejší štatistický prístup k odhadu počtu obyvateľov je však založený na korelácii medzi počtom obyvateľov a rôznymi kategóriami využitia zeme, prípadne krajinej pokrývky. Myšlienku použitia hodnôt pixelov snímky na odhad počtu obyvateľov rozvinul Hsu (1973), ako prví ju ale implementovali Iisaka a Hegedus (1982) pri odhade počtu obyvateľov Tokya. Ako identifikovať obývané územia zo satelitných snímok ukázali vo svojej práci Monmonier a Schnell (1984). Na to, aby bolo vôbec možné odhadnúť počet obyvateľov, sa kľúčovou otázkou stalo určenie hustoty zaľudnenia jednotlivých kategórií využitia zeme, a to buď prostredníctvom regresnej analýzy, výberového prieskumu alebo cenzových štatistík. Množstvo autorov uprednostňuje práve regresnú analýzu na určenie vzťahu medzi počtom obyvateľov a rôznymi kategóriami využitia zeme (Langford et al. 1991; Goodchild et al. 1993; Weber 1994; Lo 2003). V tejto súvislosti sa dáva prednosť skôr práci s rastrovou ako vektorovou formou údajov o využití zeme, pričom sa pracuje s počtom obyvateľov zdrojových zón a počtom pixelov každej kategórie využitia zeme, ktorý spadá do každej zdrojovej zóny. Integrovať cenzové dáta a dáta zo satelitných snímok sa prostredníctvom metód viacnásobnej regresie pokúsili Langford et al. (1991). Metóda pracuje s počtom obyvateľov zdrojových zón a počtom pixelov každej kategórie využitia zeme, ktorý spadá do každej zdrojovej zóny. Ten je možné určiť z klasifikácie satelitných snímok do jednotlivých kategórií využitia zeme. Na ich základe vypočíta regresné parametre, a to prostredníctvom troch modelov – všeobecného, ohniskového a jednoduchého modelu, ktoré sa použijú na redistribúciu počtu obyvateľov z cenzových obvodov do rastra veľkosti 1 km. Všeobecný model zahŕňa päť kategórií využitia zeme, ohniskový model naopak skúma len územia s vysokou a nízkou hustotou obývaných kategórií využitia zeme a jednoduchý model používa všetky obývané kategórie ako jednu nezávisle premennú.

Podľa Flowerdewa (1988) nie je na interpoláciu obyvateľstva vhodný lineárny regresný model, ale naopak, má sa použiť tzv. Poissonova distribúcia chýb. Flowerdew (1988) rozvinul metódu areálovej interpolácie dát, ktorá operuje ako Poissonov proces so skupinou binárnych premenných. V podstate tu rozvinutá metóda pracuje na základe vytvorenia regresného vzťahu medzi záujmovou premennou a jednou alebo viacerými pomocnými premennými, ktorý sa potom použije na odhad hodnôt záujmovej premennej pre cieľové zóny. Tento vzťah sa ale musí odhadnúť vhodným spôsobom vzhľadom na predpokladané rozdelenie záujmovej premennej (Poissonovo, binomické alebo normálne rozdelenie). V prípade jednotlivých tried využitia zeme ako binárnych premenných je táto metóda

podobná tej, ktorú použili Langford et al. (1991). Flowerdew a Green (1989) túto metódu rozšírili a zahrnuli do nej časť štatistickej metódy navrhutej Dempsterom (1977) a známej ako EM algoritmus. EM algoritmus je technika, ktorá pomáha vyrovať sa s chýbajúcimi dátami. Pozostáva z dvoch iteratívnych krokov, E a M. V kroku E (skratka anglického výrazu *expectation*) sa vypočítajú podmienené očakávané hodnoty chýbajúcich dát za podmienky, že je daný model a pozorované dáta. V kroku M (skratka anglického výrazu *maximization*) sa daný model kalibruje metódou maximálnej vierohodnosti na dátach z „úplného“ súboru vrátane odhadov z kroku E. Tieto kroky sa opakujú pokiaľ algoritmus neskonverguje. Práce Flowerdewa a Greena (1989 a 1991) a Flowerdewa et al. (1991) sa týkajú dát s diskretným rozdelením, a to hlavne dát, ktorých modelom je Poissonovo rozdelenie pravdepodobnosti (s takýmto typom dát sa pravidelne stretávame najmä v sčítaniach obyvateľstva) alebo dát s binomickým rozdelením. Neskôr Flowerdew a Green (1992 a 1994) túto metódu aplikovali na spojité premenné (dáta s normálnym rozdelením).

Prístupy štatistického modelovania sa dajú vhodným spôsobom využiť aj pri odhade dát z navzájom odlišných geografických jednotiek. Ako čiastkové analýzy sa často inkorporujú do inteligentných metód areálovej transformácie geografických dát, obzvlášť do dazymetrickej metódy, ktorá na odhad hustoty zaľudnenia pomocných zón používa práve regresné modely alebo techniku EM algoritmu. Mnohí autori dokonca členia inteligentné areálovo-interpoláčne metódy na tie, ktoré sú založené na dazymetrickom mapovaní a tie, ktoré využívajú princípy štatistického modelovania (Hawley a Moellering 2005; Wu et al. 2005; Langford 2006; Lin et al. 2013). Hranica medzi nimi je ale pomerne diskutabilná a spravidla nie je jasné, kedy možno danú metódu označiť za dazymetrickú (s využitím štatistických prístupov) a kedy za štatistickú. Najpodstatnejší rozdiel medzi nimi je v tom, že modely zo skupiny štatistických metód sú navrhnuté hlavne na odhad celkového počtu obyvateľov, zatiaľ čo v dazymetrickej metóde sa vybrané štatistické modely používajú na odhad hustoty zaľudnenia pomocných dazymetrických zón. Keďže sa však počet obyvateľov a hustota zaľudnenia dajú navzájom určiť jeden od druhého, cez rozlohu záujmového územia, metódy navrhnuté na odhad počtu obyvateľov sa môžu použiť aj na odhad rozmiestnenia obyvateľstva. Na druhej strane, dazymetrické metódy pracujú s rôzne členenou cenovou populáciou so zachovaním počtu obyvateľov pôvodných jednotiek, zatiaľ čo štatistické metódy len odhadujú celkový počet obyvateľov v intercenzovom období pre dané územie a nezachovávajú rozsah odhadovanej premennej. Výsledkom štatistických metód môžu byť aj negatívne počty obyvateľov, čo je v procese areálovej interpolácie neprípustné. Existujú však mnohé rozšírenia štatistických prístupov, ktoré sa práve s týmto problémom vyrovnávajú (Flowerdew a Green 1991, 1992 a 1994, Yuan et al. 1997). Hoci sú oba typy metód založené na hypotéze, že existuje vzťah medzi počtom obyvateľov zdrojovej zóny a pomocnými zónami, ktoré zdrojová zóna obsahuje, štatistické metódy sú relatívne komplexné a globálne zamerané, čo znamená, že sa kalibrujú použitím dát zo všetkých zdrojových jednotiek, ktoré tvoria študovaný región. Na odhad hodnôt pomocných zón teda používajú globálne informácie o celom študovanom území, zatiaľ čo dazymetrické techniky využívajú regionálnu regresiu na kalibráciu modelov. To znamená, že každá pomocná zóna/pixel určitej kategórie (krajinnej pokrývky, resp. využitia zeme) bude pri použití štatistických metód obsahovať rovnaký počet obyvateľov v celom území, bez ohľadu na to, v ktorej

zdrojovej zóne sa nachádza. Naopak pri použití dazymetrických metód budú mať síce tie isté triedy pomocných zón v určitej zdrojovej zóne rovnakú hustotu zaľudnenia, ale dve pomocné zóny rovnakej kategórie v rôznych zdrojových zónach budú mať pravdepodobne pridelený rôzny počet obyvateľov (Hawley a Moellering 2005). Wu et al. (2005) a Langford (2006) uvádzajú, že štatistické metódy sa viac sústreďujú na odvodenie vzťahu medzi interpolovanou a pomocnou premennou a vyžadujú si štatistické zhodnotenie tohto vzťahu.

Pokiaľ ide o samotnú dazymetrickú metódu, možnosť jej využitia v procese areálovej interpolácie načrtli Flowerdew a Openshaw (1987), no ako uvádzajú Fisher and Langford (1995), nikdy ju vo svojej práci nepoužili. Ako prví publikovali výsledky areálovej interpolácie použitím dazymetrickej metódy Fisher and Langford (1995). Ich metóda je variantom Wrightovej metódy v tom zmysle, že používa binárne delenie jednotlivých kategórií využitia krajiny, ktoré sa vzťahuje na prítomnosť a neprítomnosť obyvateľstva v území. Na základe princípov dazymetrickeho mapovania Fisher a Langford (1995) nadväzujúc na Langforda et al. (1991) aplikovali viacnásobnú regresiu na testovanie vzťahu medzi počtom obyvateľov a jednotlivými triedami krajiny pokrývky. Použili rovnaké tri modely (všeobecný, ohniskový a jednoduchý) a parametre týchto modelov následne využili na odhad počtu obyvateľov cieľových zón. Na rozdiel od Langforda et al. (1991) ale odvodili regionálne regresné modely, ktoré neskôr lokálne prispôbobi prostredníctvom rôznych techník, čím odstránili negatívne odhady počtu obyvateľov.

Okrem Fishera a Langforda (1995) sa medzi práce aplikujúce dazymetrické mapovanie zaraďujú štúdie Yuana et al. (1997), Harveyho (2000, 2002b), Gallega a Peedella (2001), Mennisa (2003), Mennisa a Hultgreena (2005), Mennisa a Hultgreena (2006a a 2006b), Bieleckej (2005), Reibela a Agrawala (2007), Rosinu et al. (2012) a Rosinu a Hurbánka (2013) a mnoho iných, ktoré síce využívajú regresnú analýzu na odhad hustoty zaľudnenia pomocných zón, ale sú taktiež regionálne špecifikované. Niekoľko interpolačných metód založených na regresných modeloch uviedli aj Goodchild et al. (1993). Práce Yuana et al. (1997) a Reibela a Agrawala (2007) nadviazali na Langforda et al. (1991), ale v konečnej fáze odhadnutý počet obyvateľov prispôbobi prostredníctvom rôznych techník tak, aby zachovali pyknofylaktickú vlastnosť odhadovanej premennej. Yuan et al. (1997) na rozdiel od Langforda et al. (1991) nepoužili globálny regresný model využívajúci údaje za všetky zdrojové geografické jednotky študovaného územia, ale aplikovali regionálny regresný model rozdelením sledovaného územia do štyroch nezávislých regiónov za účelom získania presnejších odhadov. Toto tvrdenie však nepreukázali testovaním a ani porovnaním svojej metódy s inými ostatnými používanými metódami. Pomerne inovatívny prístup, ktorý vo svojej podstate uplatňuje dazymetrické mapovanie na úrovni pixelov, obsahujú práce Harveyho (2000, 2002b). Tieto štúdie sa snažia vysporiadať s homogénnym charakterom samotných dazymetrických zón, a to prostredníctvom odhadu hustoty zaľudnenia každého pixela pomocou opakovanej regresie. Ide vlastne o aproximáciu k štatistickému prístupu modelovania dát – EM algoritmu, ktorý navrhol Dempster (1977) a neskôr použili Flowerdew a Green (1989, 1991, 1992, 1994) na kombináciu dát z rôznych zonálnych systémov. EM algoritmus a jeho vylepšenia v procese areálovej interpolácie neskôr predstavili aj Gallego (2010), Leyk et al. (2013), Schroeder a van Riper (2013), Sridharan a Qiu (2013).

Gallego a Peedell (2001) použili dazymetrickú metódu na vytvorenie mapy hustoty obyvateľstva Európy (tzv. PDG EU – *Population Density Grid of the EU*). Populačné dáta dezagregovali použitím databázy CORINE land cover, pričom predpokladali, že pomer hustoty zaľudnenia dvoch tried krajiny pokrývky je rovnaký pre ktorúkoľvek obec vo vtedajšej EÚ 15. Dostali však veľmi nepresné výsledky, ktoré sa podľa ich návrhu zlepšia, ak sa Európa rozdelí do regionálnych zoskupení podobného charakteru. V *Joint Research Centre (JRC) of European Commission* (Gallego 2010) bolo neskôr vytvorených viacero verzií PDG EU, ktoré sa odlišujú predovšetkým metódou dezagregácie (iteratívna metóda s využitím priestorového váženía, logistickej regresie či EM algoritmu) a čiastočne aj použitými pomocnými údajmi. Bielecka (2005) vytvorila dazymetrickú mapu obyvateľstva regiónu Podlasie na severe Poľska, pričom na dezagregáciu počtu obyvateľov z obcí do vybraných obývaných tried krajiny pokrývky použila najprv koeficienty vypočítané Gallegom a Peedellom (2001) pre Európu, ale neskôr ich prispôbila poľským podmienkam. Na určenie vzťahu medzi počtom obyvateľov každej obce a jednotlivými triedami krajiny pokrývky použila jednoduchý regresný model, na základe ktorého sa každá obec zaradila do jedného zo šiestich regiónov. Pre každý z týchto regiónov sa neskôr iteratívnym spôsobom vypočítal súbor koeficientov, prostredníctvom ktorých sa počet obyvateľov priradil jednotlivým triedam krajiny pokrývky. Vylepšenú dazymetrickú metódu priestorovej dezagregácie hustoty zaľudnenia predstavili Rosina a Hurbánek (2013) a Rosina (2015). Na spresnenie rozmiestnenia obyvateľstva s priestorovým rozlíšením 100 m použili voľne dostupné pomocné priestorové údaje. Iteratívny dezagregačný algoritmus pôvodne navrhnutý pre použitie s údajmi o krajiny pokrývke CORINE prispôbili tak, aby mohli byť využité aj ďalšie zdroje pomocných údajov – vrstva nepriepustných povrchov s vysokým rozlíšením a údaje z projektu OpenStreetMap (cestné a železničné úseky). Algoritmus vylepšili novým prístupom k regionálnemu odhadovaniu koeficientov hustoty zaľudnenia – predmetné regióny definovali v abstraktných n-rozmerných príznakových priestoroch odvodených z tých istých pomocných údajov, ktoré využili na priestorové spresnenie dezagregácie. Vzhľadom na potrebu referenčných údajov o obyvateľstve s vysokým rozlíšením pre porovnanie jednotlivých nastavení parametrov metódy uprednostnili ako skúmané územie Rakúsko a Slovinsko, kde boli takéto údaje dostupné. Posledné spomenuté práce sa však nezaoberajú využitím dazymetrickej metódy na odhad premennej z jednej skupiny geografických jednotiek do druhej, odlišnej skupiny územných jednotiek, ale venujú sa dezagregácii premennej do pomocných (dazymetrických) zón. Osobitne treba ešte spomenúť podobné projekty globálneho charakteru, ako databáza LandScan s priestorovým rozlíšením 30 sekúnd (pre USA existuje podrobnejšia verzia s rozlíšením 3 sekundy), kde sa zároveň využilo viacero druhov údajov ako krajinná pokrývka, intenzita nočného osvetlenia, sklonitosť reliéfu, cestné siete a i. (Dobson et al. 2000, Bhaduri et al. 2002), či projekt Gridded Population of the World (GPW) pôvodne vytvorený Toblerom, Deichmannom, Gottsegenom a Maloyovou v roku 1995 (Tobler et al. 1995) a momentálne dostupný vo verzii 4 a ďalej vyvíjaný v rámci centra CIESIN (*Center for International Earth Science Information Network*) na Kolumbijskej univerzite (CIESIN 2016).

Ako vidieť, postupne sa objavujú prístupy, ktoré nevyužívajú len jeden druh pomocných údajov, ale pracujú s kombináciou viacerých informácií. Wu et al. (2005) poznamenávajú,

že zahrnutím viacerých premenných do modelu sa zvyšuje jeho komplexnosť, a tak sa zároveň zlepšuje aj presnosť a robustnosť jednotlivých modelov. Prístup založený na využití viacerých pomocných informácií predstavili už Flowerdew a Green (1991, 1992, 1994), ich algoritmus však používa premenné cieľovej zóny na odhad záujmovej premennej. Ako indikátory sa za cieľové zóny použili viaceré premenné ako napr. volebné výsledky, počet voličov, počet áut pripadajúcich na jednu domácnosť, počet osôb pochádzajúcich z Pakistanu a i. Rôzne pomocné informácie (rozmiestnenie a veľkosť mestských území, riek, dopravných štruktúr, parkov a i.) rozvinul aj Deichmann (1996) a nazval ho *smart interpoláciou*. Je to na rastrí založená technika, ktorá vytvára série váh pre všetky bunky v zdrojovej zóne. Váhy sa určujú heuristicky na základe poskytnutých pomocných informácií (Hawley a Moellerling 2005). Tento typ interpolačnej metódy rozšírili Turner a Openshaw (2001), keď do odhadu parametrov modelu začlenili tzv. neurónové siete a tento prístup nazvali *clever spatial interpolation method*. Neurónové siete sú vlastne biologicky inšpirovaným prístupom k výpočtovým algoritmom inteligentného spracovania informácií. Vstupné poznatky dokážu zaznamenať, interpretovať, ale aj určiť vzťahy medzi poznatkami, zovšeobecniť ich a pod. Pomocou neurónových sietí je teda možné spracovať rôzne zložité dáta, ktoré nie je možné opísať jednoduchými pravidlami alebo ktorých zákonitosti nie je ľahké odhaliť napríklad štatistickými metódami. Predmetom tejto štúdie bola najmä tzv. dezagregovaná priestorová interpolácia – t. j. prípad, keď majú cieľové geografické jednotky vyššiu úroveň priestorového rozlíšenia ako zdrojové geografické jednotky (cieľové jednotky sú oveľa menšie ako zdrojové jednotky).

Navzájom kombinovať a prepájať možno aj samotné metódy areálovej transformácie dát (metódy štatistického modelovania s interpolačnými metódami, ako aj metódy bez využitia pomocných informácií a metódy s ich využitím). Napríklad Langford a Unwin (1994) aplikovali kernel metódu (kernel vyhladenie) na výsledky dazymetrickej metódy, a to s cieľom vytvorenia kartograficky prijateľnejšej a informatívnej mapy. Množstvo metód bez využitia pomocných informácií môže neskôr tieto informácie do seba inkorporovať, ak sa stanú dostupnými. Napríklad, ak sú v rámci pyknofylaktickej metódy dostupné informácie o obývaných územiach, môže sa obyvateľstvo zo zdrojových jednotiek prideliť najprv do tých obývaných polygónov, ktoré v nich ležia, pričom sa predpokladá, že neobývané polygóny majú nulový počet obyvateľov. Následne sa uskutoční interpolácia prostredníctvom vyhladenia. Kombináciu pyknofylaktickej a dazymetrickej metódy aplikovali napr. aj Kim a Yao (2010).

Ako vidieť z prehľadu literatúry zaoberajúcej sa problematikou areálovej transformácie geografických dát z jedného zonálneho systému do druhého, ani jedna metóda neposkytuje úplne presné výsledky, vždy ide len o odhady premennej. Spomínaní autori vo svojich štúdiách navrhli rôzne metódy, pričom niektorí sa venovali aj ich porovnaniu, ale nie všetci ich aj aplikovali v reálnej situácii a venovali pozornosť zhodnoteniu chýb jednotlivých odhadov. Presnosti vybraných metód sú venované napríklad práce Flowerdewa (1988), Flowerdewa a Greena (1989, 1991, 1992 a 1994), Flowerdewa et al. (1991), Goodchilda et al. (1993), Reibela a Bufalina (2004), Reibela a Agrawala (2007), Hawleyho a Moellerlinga (2005), Kima (2006 a 2007), Thiekena et al. (2006), Zandbergena a Ignizia (2010), Buttenfielda et al. (2015) či Ruthera et al. (2015), ktorí odhadnuté hodnoty porovnávajú

so známymi hodnotami alebo napr. Fishera a Langforda (1995), ktorí pracujú s ďalšími odhadmi. Okrem týchto spomenieme ešte práce Xiea (1995), Cockingsa et al. (1997), Mrozinskiho a Cromleyho (1999), Vossa et al. (1999), Sadahira (2000b), Eichera a Brewera (2001), Gregoryho (2002), Simpsona (2002), Bieleckej (2005), Mennisa a Hultgreena (2006a). Podľa Fishera a Langforda (1995) mali vedci v minulosti obmedzené možnosti ako porovnať výsledky a ako zhodnotiť chyby jednotlivých metód odhadu dát. Ich práca je prvým vážnym pokusom skúmať chyby vyplývajúce z areálovej transformácie geografických dát, hoci aj v nej ostali nezodpovedané viaceré otázky. Autori použili techniku Monte Carlo, ktorá pracuje s problémom modifikovateľných územných jednotiek. Algoritmus vytvára „pseudonáhodnú agregáciu n zón do m zón ... a to takú, že všetky n zóny v m zóne sú priestorovo súvislé“ (Openshaw 1977). Tým sa umožní násobná interpolácia vybranej premennej (počet obyvateľov) z jednej skupiny zdrojových jednotiek do viacerých skupín cieľových jednotiek (ktoré vznikli súčtom počtu obyvateľov m zón v každej cieľovej zóne). Vytvorí sa tak viacero testovacích situácií pre interpolačné metódy tých istých dát za to isté geografické územie. Tak sa preskúmajú vlastnosti chýb spojených s určitým interpolačným modelom. Fisher a Langford (1995) testovali metódu priestorového váženia, regresné metódy a binárnu dazymetrickú metódu. Ako však aj sami naznačujú, ich výsledky sú do istej miery limitované, pretože sa nezaoberali vyhodnotením výsledkov bodových interpolačných metód a ani EM algoritmu. Rovnako by bolo žiaduce testovať všetky spomínané metódy na väčšom území a preskúmať efekt prvotnej klasifikácie satelitných snímok. Z vykonaného testovania však vyplynulo, že priestorové váženie je najmenej presnou interpolačnou metódou. Naopak najpresnejšie výsledky v porovnaní s ostatnými metódami poskytla binárna dazymetrická metóda. Ukázala sa odolná voči zmenám hustoty zaľudnenia, ktoré sú spojené s určitými triedami využitia zeme, ako aj voči anomáliám vysokourbanizovaných, ale riedko osídlených území. Podľa tejto dvojice autorov to spôsobuje predovšetkým vlastnosť tejto metódy zachovať objem odhadovanej premennej. Dazymetrická metóda s použitím dát o využití zeme je podľa Fishera a Langforda (1995) odolnejšia v porovnaní s regresnými alebo plošnými metódami (bodové priestorovo-interpolačné metódy, pyknofylaktická metóda). Z výskumu ďalej vyplynulo, že presnosť všetkých interpolačných metód sa zvyšuje, ak množstvo cieľových jednotiek klesá a že presnosť výsledkov čiastočne zvyšuje aj komplexnosť (resp. obtiažnosť) regresných modelov. To kontrastuje s názormi Flowerdewa a Greena (1994), ktorí sú presvedčení, že jednoduchšie modely sú úspešnejšie. Sadahiro (2000b) predstavuje niekoľko detailných stochastických modelov chýb priestorového váženia. Aj podľa neho produkuje interpolácia z menších zdrojových do väčších cieľových jednotiek relatívne menšie množstvo chýb ako v opačnom prípade.

Práca Cockingsa et al. (1997) nadväzujúc na predchádzajúcu prácu Fishera a Langforda (1995) navrhuje opatrenia na parametrizáciu chýb areálovej interpolácie. Použila sa podobná technika Monte Carlo, ale testovali sa len dve metódy, a to dazymetrická metóda a priestorové váženie. Zistilo sa, že geometrické charakteristiky cieľových zón sú významným faktorom chýb v metóde priestorového váženia, ale nie chýb, ktoré sú spojené s dazymetrickou metódou.

Ostatní spomínaní autori sa zaoberali už len porovnaním vybraných metód a zhodnotením presnosti ich odhadov prostredníctvom štatistických analýz a mapovej vizuali-

zácie. Napríklad Mrozinski a Cromley (1999) poukázali na to, že odhady, ktoré poskytla dazymetrická metóda, sú presnejšie v porovnaní s tými, ktoré boli výsledkom priestorového váženía a pyknofylaktickej metódy. Eicher a Brewer (2001) zistili, že spomedzi binárnej dazymetrickej metódy, metódy troch tried a limitujúcej premennej poskytuje najpresnejšie výsledky metóda limitujúcej premennej. Medzi vektorovou a rastrovou formou dazymetrického mapovania sa neukázali podstatné rozdiely. Dazymetrické techniky založené na zónach produkovali síce o niečo menšie chyby ako rastrové dazymetrické metódy, ale rozdiely medzi nimi neboli štatisticky významné. Preto Eicher a Brewer odporúčajú používať postup založený na pixeloch pre ľahšiu a rýchlejšiu implementáciu vo forme skriptu. V práci Gregoryho (2002) sa podobne preukázali presnejšie výsledky binárnej dazymetrickej metódy kombinovanej s EM algoritmom v porovnaní s ostatnými interpolačnými metódami (priestorové váženie, pyknofylaktická metóda, regresné metódy). Mennis a Hultgreen (2006a) pracovali s počtom obyvateľov, Mennis a Hultgreen (2006b) ho doplnili o počet detí, domácnosti a obyvateľov španielskej národnosti. Štatistickou analýzou, ako aj vizuálnou interpretáciou porovnávali metódu priestorového váženía, binárnu dazymetrickú metódu a nimi vyvinuté dazymetrické metódy sámplovania a dospeli k záveru, že ich metódy dokážu výrazným spôsobom vylepšiť výsledky procesu odhadu dát oproti metóde priestorového váženía a s vhodne nastavenými parametrami aj binárnej dazymetrickej metódy. Xie (1995) porovnával a na základe opisných štatistik vyhodnocoval tri algoritmy metód cestnej siete. Zistil, že všetky jeho metódy poskytujú lepšie výsledky ako metóda priestorového váženía, pričom metóda hierarchicky váženej cestnej siete vykazuje najlepšie výsledky. Voss et al. (1999) pracovali s bodovou interpolačnou metódou, priestorovým vážením, metódou dĺžky cestnej siete a metódou zohľadňujúcou počet uzlov v cestnej sieti, ale takisto ich hodnotili len prostredníctvom štatistickej analýzy. Dospeli k záveru, že najmenej presnou metódou je bodová interpolačná metóda, za ňou nasledovali metódy priestorového váženía a dĺžky cestnej siete. Ako najlepšia sa ukázala metóda počtu uzlov. Reibel a Bufalino (2004), na rozdiel od predchádzajúcich štúdií, poskytli vo svojej práci nielen štatistické, ale aj vizuálne porovnanie metód cestnej siete a priestorového váženía.

Okrem počtu obyvateľov pracovali aj s počtom domácnosti a zistili, že metódy cestnej siete poskytujú lepšie a konzistentnejšie výsledky ako metóda priestorového váženía. Výsledky rôznych metód areálovej interpolácie (priestorové váženie, pyknofylaktická metóda, dazymetrická metóda a metóda hierarchicky váženej cestnej siete), ktoré ešte neboli dovtedy navzájom testované, zhodnotili Hawley a Moellering (2005). Ako zdrojové a cieľové geografické jednotky použili cenzové hierarchické zóny, čo umožnilo porovnať odhadnuté a skutočné hodnoty počtu obyvateľov. Najlepšie výsledky spomedzi všetkých testovaných metód vykázala metóda cestnej siete a ako sľubná sa javila aj dazymetrická metóda. Binárna dazymetrická metóda sa ukázala ako pomerne úspešná aj v práci Langforda (2006), ktorý ju porovnával s globálnymi regresnými modelmi. Naopak, Kim (2006) testoval niekoľko regresných modelov vytvorených Langfordom et al. (1991) a tri typy dazymetrickej metódy (binárnu dazymetrickú metódu, metódu troch tried a metódu limitujúcej premennej) a zistil, že hoci je korelácia medzi obývanými triedami krajiny pokrývky a hustotou zaľudnenia jasne preukázateľná, výsledky troch dazymetrických metód nie sú presvedčivé.

V reálnej situácii (povodne v Nemecku v roku 1999 a 2002) aplikovali Thieken et al. (2006) mapu hustoty zaľudnenia vytvorenú Gallegom a Peedelom (2001). Výsledky poskytli pomerne realistický obraz, až na značne podhodnotenú odhadu počtu obyvateľov v neurbanizovaných triedach krajínnej pokrývky. Dazymetrické zjemnenie vykázalo presnejšie výsledky ako metóda priestorového váženía, pyknofylaktická metóda či metóda EM algoritmu aj v prácach Buttenfielda et al. (2015) a Ruthera et al. (2015). K zaujímavým výsledkom dospeli Zandbergen a Ignizio (2010) a Zandbergen (2011), ktorí v rámci dazymetrickej metódy testovali viacero pomocných informácií (triedy krajínnej pokrývky, vrstvu nepriepustnosti povrchu, cestnej siete, nočných satelitných snímok a adresných bodov). Najpresnejšie odhady sa dosiahli použitím adresných bodov, vrstvy nepriepustnosti povrchu a až potom krajínnej pokrývky a cestnej siete. Najmenej presnou bola okrem priestorového váženía metóda využívajúca nočné satelitné snímky, čo autori zdôvodňujú ich nízkym priestorovým rozlíšením. Adresné body naopak vykázali relatívne presné odhady aj v nízko zaľudnených rurálnych oblastiach, ktoré pre ostatné dazymetrické prístupy zvyčajne predstavujú značnú výzvu. K podobným zisteniam dospeli aj Owens et al. (2010) a Tapp (2010). Podľa Simpsona (2002), jedinou možnosťou, ako sa dá vyhnúť aproximáciám v procese transformácie dát z jedného systému geografických jednotiek do druhého, odlišného systému geografických jednotiek, je zachovať zdrojové dáta ako individuálne referencované bodové záznamy. Takéto individuálne kódované dáta, ako sú napríklad záznamy poštových smerových čísel či adresné body, ktoré obsahujú prepojenia na územno-správne jednotky, sa však pri areálovej transformácii dát používajú zatiaľ len v niektorých krajinách (Veľká Británia, škandinávské krajiny).

Ako konštatujú Fisher a Langford (1995) a Gregory (2002), väčšina prác, ktoré sa zaoberajú presnosťou interpolačných techník, je zameraná na údaje typu celkového počtu obyvateľov a len málo autorov sa venuje komplikovanejším premenným. Podstatné však je, že ani jedna metóda neposkytuje úplne presné odhady cieľových dát. Zásadne by mala byť preto každá metóda doplnená kvantitatívnou mierou spoľahlivosti. Podľa Fishera a Langforda (1995) výsledkom testovania iba jednej skupiny známych hodnôt je miera presnosti vzťahujúca sa iba na jednu určitú geografickú situáciu, čo veľmi málo napovedá o globálnej vhodnosti použitých metód. Jedným z naliehavších problémov areálovej transformácie je teda aj stanovenie akejsi systematickej metodológie zhodnotenia chýb. Ako uvádza Gregory (2002), je potrebné si uvedomiť, že každá z metód produkuje chyby a jej presnosť vždy závisí od konkrétnej situácie. Je preto ťažké kvantifikovať „typické“ chyby vytvorené jednotlivými metódami, a tak býva každá z interpolačných metód spravidla porovnávaná so základnou metódou – metódou priestorového váženía. Doposiaľ sa ani jedna z metód areálovej interpolácie nepreukázala ako všeobecne vhodná. Existuje však niekoľko faktorov, ktoré dokážu ovplyvniť presnosť jednotlivých metód. Sú nimi relatívna veľkosť a priestorové usporiadanie zdrojových a cieľových geografických jednotiek a charakter a kvalita pomocných informácií použitých v procese transformácie dát z jedného systému geografických jednotiek do druhého, odlišného zonálneho systému (Sadahiro 2000a, Gregory 2002, Zandbergen 2011).

1.2.1 Klasifikácia metód areálovej transformácie geografických dát

Ako vyplynulo z uvedeného prehľadu, existuje pomerne veľké množstvo metód areálovej transformácie geografických dát, ktoré sa navzájom líšia z viacerých hľadísk. Vo všeobecnosti sa však v dostupnej literatúre stretávame len s minimom prác, ktoré poskytujú ucelený pohľad na rôzne techniky umožňujúce prenos a odhad dát za nekompatibilné geografické jednotky. Autori spravidla podávajú návrh jednej (či už priestorovo-interpoláčnej, priamo areálovo-interpoláčnej alebo štatistickej) metódy, prípadne uvažujú o viacerých prístupoch, z ktorých jedným je takmer vždy priestorové váženie. Delenie jednotlivých metód podľa určitých kritérií implicitne vidieť vo väčšine spomenutých prác, nikde však nie je uvádzaná ich komplexná klasifikácia. Prehľad dovedy známych a používaných metód zaoberajúcich sa riešením problému priestorovej inkompatibility uvádza vo svojej práci Lam (1983) a určité triedenie jednotlivých techník načrtávajú aj Goodchild a Kemp (1990), Fisher a Langford (1995), Hawley a Moellering (2005), Wu et al. (2005), Zandbergen a Ignizio (2010) či Lin et al. (2013).

Lam (1983) rozdeľuje priestorovo-interpoláčnej metódy na základe toho, či sú schopné uchovať rozsah interpolovanej premennej. Rozlišuje tzv. **objem/rozsah nezachovávajúce a zachovávajúce metódy**. K prvým menovaným zaraďuje *bodové (point-based) interpoláčnej metódy*, k objem zachovávajúcim zase *areálove (areal-based) interpoláčnej metódy: metódu priestorového váženia a pyknofylaktickú metódu*.

Podľa Goodchilda a Kempa (1990) existuje veľké množstvo algoritmov navrhnutých na riešenie problému areálovej interpolácie. Tieto sa vo všeobecnosti klasifikujú na základe dichotómie do šiestich párov: bodové vs. plošné, globálne vs. lokálne, exaktné vs. aproximujúce, stochastické vs. deterministické, súvislé vs. prerušované a objem zachovávajúce vs. objem nezachovávajúce. Jednotlivé algoritmy sa navzájom fundamentálne odlišujú v spôsobe prerozdelenia hodnôt premennej zo zdrojových jednotiek. Na základe toho táto dvojica autorov identifikuje tri všeobecné typy priestorovo-interpoláčnej techník. Prvú skupinu tvoria **priestor vyhladzujúce** (z ang. *spatial smoothing*) metódy, ktoré sú založené na bodoch a/alebo rastroch. Zaraďujú sa sem metódy, ktoré pracujú s bodmi (centroidami) zdrojových, pravidelnou mriežkou a jednou z globálnych (*kriging, analýza trendového povrchu, Furierova analýza, analýza klzavého priemeru*), alebo lokálnych vyhladzujúcich techník (*B spline, proximálna analýza či Toblerov pyknofylaktický prístup*). Ďalšou skupinou sú **priestorovo vážené** (z ang. *areal weighting*) metódy založené na ploche alebo vektore, ku ktorým autori zaraďujú *metódu priestorového váženia*. Poslednou skupinou metód sú **modelovacie al. štatisticky založené metódy** (z ang. *modeling, resp. statistics-based*). Sú to metódy, ktoré pracujú s matematickými a štatistickými modelmi, ale na rozdiel od predchádzajúcich techník využívajú aj iné premenné, ktoré sú v nejakom vzťahu k interpolovanej premennej.

Podobne Fisher a Langford (1995) identifikujú **kartografické, regresné a povrchové** (z ang. *surface*) interpoláčnej metódy. Ku kartografickým metódam zaraďujú *metódu priestorového váženia a binárnu dazymetrickú metódu*. Ako regresné metódy uvádzajú *metódu s využitím troch typov lineárne regresných modelov* na odhad hustoty zaľudnenia jednotlivých tried krajiny pokrývky ako kontrolných premenných (Langford et al. 1991), *regresnú me-*

tódu navrhnutú Goodchildom et al. (1993) a *metódu EM algoritmu* pracujúcu s Poissonovu distribúciou a rôznymi demografickými charakteristikami ako kontrolnými premennými, ktorú rozvinuli Flowerdew (1988) a Flowerdew a Green (1989 a 1991). Povrchové metódy sú založené na matematickom tvrdení, že na hustotu zaľudnenia je potrebné nazeráť ako na spojito sa meniacu funkciu. Patrí sem Toblerova *pyknofylaktická metóda a bodové priestorovo-interpolačné metódy* založené na využití ťažísk zdrojových zón a funkcie vzdialenosti (Martin 1989, Bracken 1991; Martin a Bracken 1991). Ako však uvádzajú Fisher a Langford (1995), kartografické metódy možno vnímať ako špeciálne typy povrchových metód, a to v takých prípadoch, keď je na tomto povrchu premenná rozmiestnená diskkrétne a nie spojito.

Pomerne utriedený pohľad na metódy, ktoré riešia problém areálovej interpolácie, podávajú Hawley a Moellering (2005). Jednotlivé metódy rozdeľujú podľa toho, či v procese prenosu a odhadu dát za odlišné zonálne systémy využívajú pomocné informácie alebo nie. Tie, ktoré pracujú s pomocnými dátami ďalej delia na *metódy areálovej interpolácie s využitím dát z DPZ, metódy areálovej interpolácie s využitím dát o cestnej sieti a ostatné areálovo-interpolačné metódy s využitím pomocných dát*. Medzi **metódy areálovej interpolácie bez využitia pomocných informácií** zaraďujú *metódu prekryvu, pyknofylaktickú metódu a bodové interpolačné metódy*. K **metódam areálovej interpolácie s využitím dát z DPZ** zaraďujú *dazymetrickú metódu a globálne regresné metódy, ktoré pracujú s údajmi z DPZ*. Medzi **interpolačné metódy s využitím dát cestnej siete** radia tri algoritmy pre areálovú interpoláciu s využitím dát o cestnej sieti, ktorých autorom je Xie (1995) – *metódu dĺžky siete, metódu hierarchicky váženej siete a metódu bytovej zataženej siete*. K tzv. ostatným **metódam areálovej interpolácie s využitím pomocných dát** zaraďujú metódy využívajúce štatistické techniky a komplexnejšie pomocné informácie. Patrí sem *EM algoritmus Flowerdewa a Greena (1989, 1991, 1992 a 1994), metóda kontrolných zón* (Goodchild et al., 1993) a tzv. *smart interpolačné metódy* využívajúce rôzne sady pomocných informácií (Deichmann 1996 a Turner a Openshaw 2001).

Práca Wu et al. (2005) na rozdiel od uvedených triedení rozlišuje **metódy areálovej interpolácie dát a metódy štatistického modelovania**. Metódy areálovej interpolácie ďalej delí do dvoch kategórií podľa využitia pomocných informácií. Metódy, ktoré nevytvárajú pomocné informácie člení podobne ako Lam (1983) na *bodové a areálové interpolačné metódy* (sem radí priestorové váženie a pyknofylaktickú metódu). V rámci metód, ktoré využívajú pomocné informácie, uvažuje o dazymetrickej metóde, kam zaraďuje aj regresné metódy či metódu EM algoritmu ako špecifické prístupy dazymetrického mapovania, ktoré sa aplikujú na určenie vzťahu medzi hustotou zaľudnenia a jednotlivými pomocnými zónami. Metódy štatistického modelovania delí na základe vzťahu medzi obyvateľstvom a 1) mestskými územiai, 2) využitím zeme, 3) bytovými jednotkami, 4) charakteristikami pixelov obrázka a 5) inými fyzickými a sociálno-ekonomickými charakteristikami.

Mierne odlišný pohľad na prístupy riešiacie problém priestorového nesúladu podávajú Zandbergen a Ignizio (2010). Jednotlivé metódy rozlišujú podľa komplexnosti na **základné prístupy uplatňujúce pravidlá inklúzie a exklúzie** (napr. klasická metóda centroidu) a **metódy areálovej interpolácie**, v rámci ktorých odlišujú *metódu priestorového váženia a metódy, ktoré sa snažia zmierniť predpoklad homogénneho rozmiestnenia premennej*

v zdrojových zónach. Tieto ďalej členia na viacero skupín: povrchové metódy (z angl. *surface fitting*) využívajúce inferečnú štatistiku (Tobler 1979, Bracken 1991 a pod.), *zonálne metódy* využívajúce pomocné informácie za sadu externých (kontrolných) alebo cieľových zón (regresné techniky vyvinuté Goodchildom et al. 1993, Langfordom et al. 1991, EM algoritmus Flowerdewa a Greena 1994) a skupinu metód súhrnne nazvanú ako *dazymetrické mapovanie* (pod ktorým rozumejú proces rozmiestnenia premennej zdrojovej zóny do podrobnejších zón analýzy využívajúc pomocné informácie, ktorými sú zväčša údaje o krajinnej pokrývke).

Novšie štúdie zaoberajúce sa problematikou prenosu dát z navzájom nekompatibilných územných systémov v podstate delia areálovo-interpolačné metódy do dvoch kategórií: na **jednoduché** (uvažujú o homogénnom rozmiestnení údajov v zdrojových zónach) a **inteligentné metódy areálovej interpolácie** (vyžívajúce pomocné informácie, prostredníctvom ktorých uvažujú o priestorovom rozmiestnení záujmovej premennej). Inteligentné metódy sa ďalej členia na tie, ktoré sú založené na dazymetrickom mapovaní alebo na štatistických metódach (Lin et al. 2013). Okrem toho Lin et al. (2013) predstavujú aj ďalšiu klasifikačnú schému, ktorá jednotlivé interpolačné metódy delí na základe použitého dátového modelu, konkrétne na **metódy založené na vektore** (ak všetky priestorové dáta použité v procese areálovej interpolácie, vrátane zdrojových, pomocných aj cieľových zón, sú vo vektorovom formáte) a **metódy založené na rastrí** (ak sa v hociktorom kroku interpolačného procesu použijú rastrové dáta).

Z uvedených príkladov vidieť, že existuje niekoľko kritérií, podľa ktorých je možné deliť metódy areálovej transformácie geografických dát z jedného systému územných jednotiek, do druhého, odlišného zonálneho systému. V princípe sa jednotlivé metódy od seba odlišujú vo svojej komplexnosti, jednak na základe odlišných predpokladov o rozmiestnení premennej v danom území, využitia pomocných informácií, či type použitých dát, ale aj na základe stupňa generalizácie, zachovania pyknofylaktickej vlastnosti, či využitia princípov štatistického modelovania v procese prenosu a odhadu hodnôt príslušnej premennej. Je nutné konštatovať, že nie všetky predstavené klasifikačné schémy zachovávajú logické pravidlá klasifikácie a používajú rovnakú/podobnú terminológiu. Niektorí autori napríklad považujú jednoduché interpolačné metódy za také, ktoré uvažujú o homogénnom rozmiestnení premennej, pre iných sú to tie, ktoré v procese prenosu a odhadu dát nevyužívajú pomocné informácie. Ako už bolo spomenuté, odlišný a nejasný postoj zaujímajú jednotlivé práce aj v prípade dazymetrickej metódy, resp. metódy dazymetrického mapovania. Kým jedny súhrnne nazývajú všetky metódy využívajúce pomocné informácie a snažiac sa zjemniť priestorové rozmiestnenie záujmovej premennej ako dazymetrické, iné tento termín používajú len v prípade využitia údajov z DPZ, alebo vtedy, keď je v procese interpolácie nutné odvodiť hustotu zaľudnenia. Nejednoznačná je aj hranica medzi dazymetrickými a štatistickými metódami, ktoré sa používajú na získanie hustoty zaľudnenia jednotlivých pomocných zón.

Na základe dostupnej literatúry a najdôležitejších charakteristík jednotlivých metód používaných na odhad vybranej premennej za rôzne geografické jednotky budeme rozlišovať niekoľko rovín a v rámci nich hierarchické stupne, na základe ktorých delíme metódy areálovej transformácie geografických dát nasledovne (upravené podľa Madajová 2010 a 2011):

A. Podľa obsahu:

1. **rozsah/objem zachovávajúce metódy areálovej transformácie,**
2. **rozsah/objem nezachovávajúce metódy areálovej transformácie.**

B. Podľa formy:

1. **priestorovo-interpoláčne metódy (PIM) areálovej transformácie,**
2. **metódy areálovej transformácie založené na štatistickom modelovaní,**
3. **kombinované metódy areálovej transformácie dát.**

Rozdiel medzi B1 a B2 možno definovať v súlade s Wu et al. (2005). PIM sú v prvom rade určené na problém prerozdelenia dát z jednej skupiny geografických jednotiek do druhej pomocou určitých interpolačných operácií. Štatistické modelovanie sa na druhej strane viac zaujíma o odvodenie vzťahu medzi interpolovanou premennou a inými, zväčša sociálno-ekonomickými premennými s cieľom odhadu tejto premennej prostredníctvom rôznych teórií známych v urbánnej geografii. Tento prístup bol pôvodne navrhnutý na odhad veľkosti populácie v medzicenzovom období alebo veľkosti populácie územia odlišného od toho, za ktoré sa údaje publikovali z cenzu. Tieto prístupy sa môžu, ale nemusia začleniť aj do procesu prenosu a odhadu dát za rôzne zonálne systémy.

Keďže v procese areálovej transformácie geografických dát z odlišných zonálnych systémov sa metódy štatistického modelovania často aplikujú v kombinácii s nejakou priestorovo-interpoláčnou metódou ako samotné (t. j. do priestorovej interpolácie sa inkorporujú rôzne matematické a štatistické modely), možno toto delenie ešte modifikovať a ako tretiu kategóriu uviesť kombinované metódy areálovej transformácie dát.

PIM možno ďalej deliť na:

- 1.1. bodové interpolačné metódy (na bodoch založené metódy areálovej interpolácie – napr. metóda inverzných vzdialeností,
- 1.2. líniové interpolačné metódy (na líniách založené metódy areálovej interpolácie) – metódy cestnej siete,
- 1.3. plošné/polygónové interpolačné metódy (na území založené metódy areálovej interpolácie) – priestorové váženie, pyknofylaktická metóda, dazymetrická metóda.

a na ďalšom stupni na:

- I. rastrové
- II. vektorové metódy,

tak ako ich vníma Lin et al (2013).

C. Podľa využitia pomocných informácií:

1. **jednoduché metódy areálovej transformácie (metódy bez využitia pomocných informácií),**
2. **inteligentné metódy areálovej transformácie (metódy s využitím pomocných informácií):**

- 2.1. špecifické – s využitím pomocných informácií za zdrojové jednotky,
- 2.2. flexibilné – s využitím viacerých (komplexnejších) pomocných informácií alebo využitím pomocných informácií za cieľové jednotky,
- 2.3. hybridné – zložené z 2.1. a 2.2. alebo z jednoduchých a inteligentných metód.

K jednoduchým transformačným metódam patria napr. bodové interpolačné metódy, priestorové váženie a pyknofylaktická metóda, k špecifickým možno zaradiť dazymetrickú metódu či metódy cestnej siete. Flexibilné metódy zahŕňajú napr. kombináciu dazymetrickej a pyknofylaktickej metódy (Kim a Yao 2010) alebo metódy s využitím EM algoritmu či neurónových sietí.

V tejto súvislosti je ešte potrebné zdôrazniť, že pod dazymetrickou metódou rozumieme metódu založenú na princípe dazymetrického mapovania, pričom termín dazymetrický stotožňujeme v súlade s Wrightom (1936) s meraním hustoty zaľudnenia. Za dazymetrické metódy budeme teda považovať také techniky, ktoré sa snažia zmierniť predpoklad homogénneho rozmiestnenia premennej a v rámci zdrojových geografických jednotiek pracujú s jemnejšími (podrobnejšími) zónami, ktorých hustotu zaľudnenia sa snažia odvodiť prostredníctvom rôznych prístupov (a priori cez prahové hodnoty, prostredníctvom štatistického modelovania).

1.2.2 Stručná charakteristika a všeobecné porovnanie základných metód areálovej transformácie geografických dát

V nasledujúcej podkapitole je podaná stručná charakteristika jednotlivých metód areálovej transformácie geografických dát s cieľom čo najlepšie objasniť spôsob ich použitia a poukázať na základné rozdiely medzi nimi. Dôraz sa kladie predovšetkým na tie metódy, ktoré sú vhodné, resp. sa najčastejšie používajú pri riešení problému analýzy, kombinácie či porovnania údajov dostupných za rôzne geografické jednotky. Pri vysvetľovaní základných princípov fungovania jednotlivých techník vychádzame z delenia metód areálovej transformácie geografických dát podľa využitia pomocných informácií, keďže práve táto rovina delenia transformačných metód sa javí ako najpodstatnejšia v rámci riešenia problému odhadu dát za nekompatibilné priestorové jednotky. Ako uviedli Goodchild a Gopal (1989, str. 219): „... *the more one knows about the area, the better estimates*“ (čím viac vieme o území, tým lepšie sú odhady). Okrem toho podávame aj stručnú charakteristiku geografických konverzných tabuliek ako špeciálneho prístupu transformácie údajov z jedného zonálneho systému do druhého, odlišného systému geografických jednotiek.

A. Jednoduché metódy areálovej transformácie

Jednoduché metódy areálovej transformácie geografických dát nevyužívajú pomocné informácie v procese prenosu a odhadu dát za nekompatibilné geografické jednotky. Radíme k nim bodové priestorovo-interpolačné metódy, pyknofylaktickú metódu a priestorové váženie.

A1. Bodové interpolačné metódy

Väčšina bodových interpolačných techník odhaduje hodnoty bodov rozmiestnených na vopred definovanej pravidelnej sieti z nepravidelne rozmiestnených údajových bodov, ktoré nesú hodnoty modelovaného javu. Tento proces sa označuje v anglicky písanej literatúre ako *gridding* a jeho cieľom je vytvoriť kvázi spojité povrch hodnôt študovanej premennej (Hlásny 2007). Zvyknú sa deliť na tzv. globálne a lokálne, a to podľa toho, či uvažujú naraz o všetkých údajových hodnotách alebo len o hodnote vnútri preddefinovaného susedstva každého bodu (Goodchild a Lam 1980). Ďalšou možnosťou je rozdeliť ich na exaktné a aproximačné (Lam 1983; Xie 1995; Wu et al. 2005), podľa toho, či pracujú s pôvodnými hodnotami bodov, alebo určujú funkcie celkového povrchu. Globálne metódy zvyčajne produkujú približné interpolačné výsledky, lokálne naopak poskytujú presnú interpoláciu. Exaktné metódy zahŕňajú proximálnu analýzu (výsledkom sú Thiessenove polygóny), splajny (výsledkom je spojité povrch bez zakrivenia), kriging a iné. Medzi aproximačné metódy sa zase radia trendové modely alebo Fourierove modely a i.

Bodové interpolačné metódy sa používajú častejšie v súvislosti s interpoláciou dát fyzickogeografickej povahy (napríklad na tvorbu modelu reliéfu), isté uplatnenie však našli aj v humánnogeografickej sfére pri odhade hodnôt za nekompatibilné zonálne systémy. V takomto prípade reprezentuje každú zdrojovú zónu tzv. kontrolný bod, ktorý je spravidla ťažiskom tejto zóny, ale môžu to byť aj zástupné body alebo niektoré typy vážených stredov. Z kontrolného (známeho, reprezentatívneho bodu) sa odhadnú (interpolujú) hodnoty do ľubovoľného iného bodu na základe funkcie, o ktorej sa predpokladá, že vhodne popisuje priestorovú reprezentáciu premennej zdrojovej zóny. Príkladom sú štúdie Martina (1989), Brackena (1991), Martina a Brackena (1991), v ktorých sa použil druh bodovej interpolačnej techniky nazvaný *kernel interpolácia*. Ako kontrolný bod sa použije ťažisko zdrojovej zóny. Nad každým kontrolným bodom sa následne striedavo umiestni tzv. filtrovacie okno, pričom obyvateľstvo zdrojovej zóny sa prideliť len tým bunkám mriežky, ktoré ležia v danom okne na základe váženého prostredníctvom funkcie poklesu vzdialenosti medzi ťažiskom zdrojovej zóny a bunkou mriežky. Táto metóda je teda variantom klasickej interpolačnej metódy inverzných vzdialeností. Ďalší typ metódy bod v polygóne – presúva hodnoty premennej zo zdrojovej zóny do cieľovej, ak jej „reprezentatívny bod“ leží v cieľovej zóne (Okabe a Sadahiro 1997). Hoci táto metóda nie je časovo náročná, jej výsledky sa pre tú istú skupinu dát môžu značne líšiť, a to v závislosti od lokalizácie vybraných reprezentatívnych bodov.

Výsledky bodových interpolačných metód sú teda ovplyvnené jednak kvalitou pôvodných dátových bodov (ich hustotou a priestorovým usporiadaním), ale aj zložitou povrchu, preto výber jednotlivých metód závisí predovšetkým od typu dát či stupňa požadovanej presnosti. Každá z bodových interpolačných metód má svoje výhody aj nevýhody. Podľa Lam (1983) možno exaktné bodové interpolačné metódy v dôsledku formálnej jednoduchosti a flexibility považovať za spoľahlivejšie ako aproximujúce.

Vo všeobecnosti sa však bodové interpolačné metódy nepovažujú za príliš vhodné na riešenie problému odhadu dát za rôzne geografické jednotky. Spôsobuje to jednak problematický predpoklad homogénneho rozmiestenia skúmanej premennej v zdrojovej zóne, čím sa vo väčšine bodových metód nemusí zachovať celková hodnota interpolovanej pre-

mennej a jednak vplyv lokalizácie ťažísk zdrojových zón na interpolované hodnoty. Použitie kontrolných bodov na reprezentáciu celej zdrojovej zóny totiž produkuje chyby. Výpočet ťažiska alebo umiestnenie stredu daného územia závisí od súradníc bodov, ktoré definujú hranice zdrojovej zóny (t. j. od geometrie zóny). Ak je zdrojová zóna symetrická alebo relatívne jednoduchá, ťažisko alebo stred územia je vhodným kontrolným bodom a odhadnutá hodnota pre každú bunku mriežky bude relatívne spoľahlivá. Ak však tvar zdrojovej zóny nie je symetrický, alebo jej hranice sú veľmi generalizované, lokalizácia ťažiska ovplyvní výsledky odhadov dát. V skutočnosti sú cenzové jednotky len zriedkakedy symetrické, pričom sa situácia komplikuje aj tým, že počet obyvateľov alebo iná humánnogeografická premenná v nich nie je rovnomerne rozmiestnená. Najväčším problémom väčšiny bodových interpolačných metód je však fakt, že môžu vytvárať dramatické vizualizácie rozmiestnenia odhadovanej premennej a nezachovávajú celkovú hodnotu premennej v zdrojovej zóne, čím je automaticky narušená spoľahlivosť metódy. Objavili sa preto určité pokusy, ktoré sa pri použití bodovej interpolačnej metódy snažia uchovať obsah interpolovanej premennej. Príkladom sú práce Martina (1996), ktorý modifikoval kernel interpolačný algoritmus, aby zabezpečil, že počet obyvateľov vzťahujúci sa k cieľovým zónam bude súčtom počtu obyvateľov zdrojových zón a Kyriakidis (2004), ktorý naopak uvažuje o interpolačnej technike „územie do bodu“, ktorá využíva geoštatistickú metódu krigingu doplnenú o začlenenie pomocných informácií do procesu interpolácie dát.

A2. Pyknofylaktická metóda

Iný variant, ktorý vlastne predstavuje opak predchádzajúceho, najlepšie ilustruje práca amerického geografa Walda R. Toblera (1979). Tobler uvažuje o diskretných zónach a dáta za ne dostupné považuje za prejavy súvislého povrchu danej premennej. Pre populačné dáta stanovil existenciu súvislého povrchu hustoty zaľudnenia nad celým študovaným územím a počet obyvateľov špecifických zón vidí ako súčasť tohto povrchu nad regiónom, ktorý je definovaný zonálnymi hranicami (Flowerdew a Green 1992). Toto kritérium nazýva „pyknofylaktickou vlastnosťou“, ktorá vlastne znamená, že integrál funkcie hustoty vo vnútri každého regiónu je rovný počtu obyvateľov regiónu, čím sa pyknofylaktická metóda zaraďuje medzi objem zachovávajúce metódy. Úhrny dát za zdrojovú skupinu geografických jednotiek sa teda počas areálovej transformácie uchovávajú pre druhú skupinu geografických jednotiek.

Tobler odhaduje funkciu vybranej premennej (v jeho prípade hustoty zaľudnenia) v uzloch jemnej mriežky z agregovaných populačných štatistík pre ľubovoľne tvarovaný región. Cieľom tejto metódy je modelovať povrch tak, aby bol dokonale hladký, ale bez toho, aby sa v procese areálovej transformácie dát nejaké údaje stratili. Tobler vyvinul postup na vytvorenie takéhoto vyhladeného povrchu z dát založených na polygónoch. Na zdrojové zóny sa naloží dostatočne hustá mriežka (v závislosti od veľkosti zdrojových zón, spravidla však 100 x 100 m). Každéj bunke v mriežke sa prideli číslo (od 1 po m) identifikujúce zdrojovú zónu, ktorá danú bunku obsahuje. Každéj bunke mriežky sa následne prideli hodnota tej zdrojovej zóny, v ktorej bunka leží a vydeli sa počtom buniek v danej zdrojovej zóne. Potom sa zavedie vyhladzovacia podmienka, že hodnota každej bunky sa nahradí novou hodnotou (novým počtom obyvateľov) vypočítanou ako vážený priemer

hodnôt jej štyroch susedov. Odhady pre každý región sa uchovávajú ako súčty rozdielu medzi pôvodnou a novou vyhladenou hodnotou bunky pre každú bunku v regióne i . Tento priemerný odhad pre každý región sa následne pridá k novej vyhladenej hodnote bunky v regióne, až pokiaľ nebude mať žiadna bunka hodnotu menšiu ako 0. Potom sa vypočíta celková hodnota premennej pre každý región a určí sa priemerný rozdiel hodnôt danej premennej, ktorý sa pridá každej bunke v regióne k . Toblerov proces sa opakuje dovtedy, pokiaľ všetky odhady nedosiahnu používateľom definovaný prah alebo maximálny počet opakovaní. Počas každého opakovania sa súhrn prispôsobí stanovenému počtu obyvateľov v zdrojovej zóne. Nakoniec sa hodnota premennej z každej bunky realokuje do cieľových zón pomocou prvej mriežky pre skupinu cieľových zón. Výsledkom je rastrová mriežka cieľovej zóny. Počet najbližších susedov a počet opakovaní určuje celkový stupeň vyhladenia a je subjektívny.

Tento postup je však efektívny v tom prípade, ak je zrejmé, že modelované dáta možno považovať za spojito sa meniace v priestore (Flowerdew a Green 1992). Pre väčšinu sociálnych a ekonomických dát je však tento prístup menej efektívny, pretože pre takéto dáta sú typické skôr nespojitosti v ich priestorovom rozložení.

A3. *Priestorové váženie*

Metóda prekryvu alebo priestorového váženia sa považuje za štandardnú a zároveň najjednoduchšiu areálovú interpolačnú metódu. Túto techniku zaviedli Markoff a Shapiro vo svojej práci z roku 1973, detailnejšie ju však rozpracovali Goodchild a Lam (1980). Metóda je založená na predpoklade, že údaje sú v zdrojových zónach rozmiestnené rovnomerne, takže dáta platné pre skupinu zdrojových geografických jednotiek sa úmerne pridelia skupine cieľových geografických jednotiek. Premenná sa interpoluje na základe rozlohy územia prieniku medzi zdrojovou a cieľovou zónou. Zóny prieniku sa vytvoria prekryvom zdrojových a cieľových zón. Hodnoty premennej pre cieľovú zónu sa potom odhadnú na základe známych hodnôt premennej v zdrojovej zóne a z časti prieniku so zdrojovou zónou – z tzv. váhy:

$$\hat{y}_c = \sum_z \frac{A_{zc} \cdot y_z}{A_z}$$

kde:

\hat{y}_c = odhad hodnoty premennej v cieľovej zóne,

y_z = hodnota premennej v zdrojovej zóne,

A_z = rozloha zdrojovej zóny,

A_{zc} = rozloha prieniku oboch zón.

Hodnoty premennej v zdrojových zónach 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 humánogeografickej povahy je však predpoklad homogenity vnútri zdrojových zón nerealistický. Okrem toho sú zdrojové zóny často vymedzené s iným zámerom

ako cieľové geografické jednotky, a teda nie vždy sú vhodným nástrojom distribúcie informácií pre cieľové zóny. Ako hovorí Lam (1983), je preto potrebné kontrolovať spoľahlivosť odhadov hlavne charakterom a stupňom homogenity zdrojovej zóny a veľkosťou cieľovej zóny vo vzťahu k zdrojovej zóne.

Metóda priestorového váženía poskytuje lepšie výsledky v porovnaní s inými metódami, keď je možné považovať skúmaný povrch za nespojitý. Naopak, ak je spojitosť skutočnou vlastnosťou povrchu, potom je lepšie použiť pyknofylaktickú metódu. V prípadoch, keď nie je možné charakter povrchu považovať za spojitý a ani za nespojitý, musia sa pre spoľahlivé výsledky zaviesť rôzne vedľajšie podmienky. Niektorí geografi sa preto pokúsili zmierniť predpoklad homogenity použitím iných typov metód, ktoré sa snažia narušiť predpoklad rovnomerného rozmiestnenia premenných v zdrojových zónach tým, že berú do úvahy určité pomocné informácie o rozmiestnení premenných vnútri zdrojových, prípadne cieľových zón. Povedané inými slovami, pokúšajú sa použiť nejaké externé poznatky o danej lokalite. Tieto metódy areálovej transformácie geografických dát označujeme ako inteligentné metódy.

B. Inteligentné metódy areálovej transformácie

Predpoklad homogénneho rozmiestnenia premenných vo vnútri zdrojových zón (prípád priestorového váženía) je opodstatnený iba v tom prípade, ak nemáme žiadne informácie o rozmiestnení údajov v zdrojových, prípadne cieľových zónach. Veľakrát však takéto pomocné informácie k dispozícii máme, a to buď priamo alebo nepriamo prostredníctvom rozmiestnenia iných dát, o ktorých predpokladáme, že sú v nejakom vzťahu so záujmovou premennou. Poznatky o takýchto údajoch sú totiž schopné ovplyvniť naše predpoklady o tom, ako môžu byť vnútri zón rozmiestnené ostatné premenné. Zapojenie doplnkových informácií do procesu areálovej transformácie geografických dát zabezpečuje lepšie výsledky a celkovo napomáha k skvalitneniu transformačného procesu.

V minulosti sa pomocné informácie získavali napríklad z topografických máp (rozloha obývaného územia) alebo zo štatistických zdrojov (počet bytových jednotiek), v posledných rokoch sa však ako vhodné doplnkové údaje zaužívali predovšetkým dáta z diaľkového prieskumu zeme (satelitné snímky), hlavne z nich odvodené mapy využitia zeme (land use), mapy krajiny pokrývky (land cover), nepriepustnosti povrchu či intenzity nočného osvetlenia. Okrem nich môžu pri odhadoch vybraných premenných asistovať aj rôzne kombinácie iných informácií, napríklad z oblasti fyzickej geografie (sklonitosť reliéfu), demogeografie, dopravy, alebo aj lidarové dáta. Ich využitie v procese odhadu hodnôt vybranej premennej súvisí hlavne s rozvojom geografických informačných systémov a s tým spojeným čoraz prístupnejším množstvom digitálnych dát. Takéto informácie potom možno použiť aj v procese areálovej transformácie dát, pretože sú schopné výskumníkovi poskytnúť lepšie informácie o rozmiestnení odhadovanej premennej. Napríklad je zrejmé, že mestské oblasti budú hustejšie zaľudnené ako zalesnená krajina, rovnako možno predpokladať, že hustota zaľudnenia (a teda samotný počet obyvateľov) súvisí s hustotou cestnej siete.

Existuje viacero implementácií metód areálovej transformácie, ktoré využívajú takéto pomocné informácie. Sú nimi jednak špecifické, ale aj flexibilné a hybridné metódy are-

álovej transformácie. Bližšie sa budeme venovať rôznym formám dazymetrickej metódy (aj v kombinácii so štatistickými metódami), metódam cestnej siete a technike EM algoritmu.

B1. Dazymetrická metóda

Dazymetrická metóda pracuje na princípe tzv. dazymetrického mapovania, ktoré ako forma kartografickej reprezentácie predstavuje alternatívu ku klasickému kartogramu. Ten sa totiž z viacerých hľadísk ukázal ako nevhodný pre detailnejšie priestorové analýzy dát humánnogeografickej povahy. Podľa Eichera a Brewera (2001) sa prostredníctvom dazymetrického mapovania znázorňujú kvantitatívne priestorové dáta prostredníctvom hraníc, ktoré rozdeľujú mapované územie do zón relatívnej homogenity s cieľom lepšieho zobrazenia štatistického povrchu. Dazymetrické mapovanie teda prezentuje povrch omnoho realistickejšie ako tradičný kartogram a umožňuje lepšie porozumieť rozmiestneniu danej premennej. Práve dazymetrická mapa sa preto stala jedným z riešení problému zobrazovania populačných dát agregovaných do priestorových jednotiek a dá sa vhodným spôsobom využiť aj v procese areálovej transformácie geografických dát. K zjemneniu rozmiestnenia agregovaných dát v zdrojových geografických jednotkách dochádza prostredníctvom pomocných informácií týkajúcich sa študovaného územia (zvyčajne sú to údaje z topografických máp a DPZ vo forme máp využitia zeme alebo krajinej pokrývky). Dazymetrická metóda si teda vyžaduje buď dáta klasifikovať do jednotlivých kategórií (tried) pomocných zón, alebo využiť už klasifikovanú skupinu údajov (mapy využitia zeme, krajinej pokrývky). Prekryvom zdrojových geografických jednotiek a vrstvy pomocných informácií (či už vo forme polygónov alebo v rastrovej forme) vzniknú pomocné, tzv. dazymetrické zóny, ktoré obsahujú prepojenia na oba zonálne systémy. Ak odhadujeme počet obyvateľov a ako pomocné informácie použijeme jednotlivé triedy krajinej pokrývky, potom je premenná (počet obyvateľov) prerozdelená na základe vzťahu medzi jednotlivými kategóriami krajinej pokrývky a hustotou zaľudnenia. Použité pomocné údaje teda slúžia ako váha na odhad rozmiestnenia obyvateľstva v zdrojových zónach. Takto vážené hodnoty sa následne reagregujú do cieľových zón.

Najjednoduchšou formou vzťahu medzi triedami krajinej pokrývky a hustotou zaľudnenia je uvažovanie o obývaných a neobývaných územiach. Zvyčajne sa ako zaľudnené definujú triedy urbanizovanej zástavby, prípadne celá kategória umelých povrchov, zvyšným územiám je pripísaná nulová hustota zaľudnenia. Ide o tzv. binárnu dazymetrickú metódu, pri ktorej sa obyvateľstvo prerozdelí do cieľových zón na základe rozlohy obývaných území (namiesto rozlohy celých prienikových území ako v prípade metódy priestorového váženia). Okrem tohto prístupu existuje niekoľko ďalších implementácií, ktoré rozlišujú viac kategórií pomocných zón. Spravidla sa pracuje so štyrmi typmi území (urbanizované, poľnohospodárske, zalesnené a zamokrené územia, resp. územia s vysokou, strednou, nízkou a nulovou hustotou zaľudnenia). Kľúčovým sa v tomto smere stáva spôsob určenia vzťahu medzi jednotlivými dazymetrickými zónami a odhadovanou premennou. Tento vzťah môže byť určený buď *a priori*, alebo štatistickým odhadom. V prvom prípade sa pre rôzne kategórie pomocných zón používajú vopred definované hodnoty určené na základe nejakej váženej schémy – buď subjektívne (napr. v pomere 70 % – 20 % – 10 % – 0 %), alebo prostredníctvom tréningových území (takých, v ktorých je dominantná jedna kategória

krajinnej pokrývky). Tieto hodnoty sa následne použijú na redistribúciu premennej zo zdrojových do dazymetrických zón. Tento prístup dazymetrickej metódy si teda vyžaduje dobré znalosti skúmaného územia. Za jeho najväčšie nevýhody sa však považuje práve subjektívne určenie váh na pridelenie hodnôt premennej, ako aj fakt, že sa nezohľadňuje individuálna rozloha každej dazymetrickej zóny v rámci jednotlivých kategórií, čo môže spôsobovať ďalšie chyby. Ak totiž napríklad jedna zdrojová zóna obsahuje málo urbanizovaných zón, aj tak sa im prideli rovnaké percento hodnoty premennej ako tým zónam, ktoré môžu obsahovať niekoľkonásobne väčšie množstvo urbanizovaných zón.

Aby sa vyhlo chybám vyplývajúcim z predchádzajúceho dazymetrického prístupu, vznikli nové spôsoby dazymetrického mapovania, ktoré sú založené na princípe limitujúcej premennej, tzv. empirického výberu (samplingu) alebo na štatistických odhadoch (regresné metódy, metóda EM algoritmu). V prípade metódy limitujúcej premennej (Hollowaya et al. 1997, Eicher a Brewer 2001) sa premenná prideli na základe istého obmedzenia jej hodnoty. Tento prah sa nastaví zvlášť pre každú kategóriu pomocných zón z hodnôt dostupných za zdrojové jednotky (pre poľnohospodárske dazymetrické zóny napríklad na 50 ob. km², zalesnené na 15 ob./km²). Flexibilný prístup, ktorý kombinuje viacero techník, rozvinuli Mennis (2003), Mennis a Hultgreen (2005) a Mennis a Hultgreen (2006a,b). Vybraná premenná je rozdelená do pomocných dazymetrických zón na základe priestorového váženého a odhadnutej hustoty každej triedy pomocnej zóny. Táto hustota sa odvodí prostredníctvom výberového procesu z tých zdrojových zón, ktoré sú spojené s danou pomocnou triedou. Váhy možno pre každú kategóriu krajinnej pokrývky vypočítať rôznymi spôsobmi. Prvý (Mennis 2003) pracuje s hodnotami všetkých zdrojových jednotiek, ktoré boli celé obsiahnuté v rámci danej urbanizovanej triedy (metóda obsahu). Nevýhodou tejto metódy však bolo to, že v reálnych situáciách je pomerne ťažké harmonizovať priestorové rozlíšenie oboch skupín geografických jednotiek, to znamená, že zväčša nie je možné disponovať takými zdrojovými jednotkami, ktoré by celé ležali v jednotlivých triedach pomocných zón. Ďalšími možnosťami (Mennis a Hultgreen 2005 a Mennis a Hultgreen 2006a,b) je výpočet na základe tých zdrojových zón, ktorých ťažiská ležia v danej kategórii pomocných zón (centroidná metóda) alebo sa nastaví prah percentuálnej hodnoty pokrytia, na základe ktorého sa potom vyberajú tie zdrojové zóny, ktoré sú pokryté danou pomocnou triedou v rozsahu rovnom alebo väčšom ako je tento prah (metóda „percentuálneho pokrytia“).

Dazymetrická metóda môže pri odhade neznámych hodnôt premennej v cieľových zónach využiť aj spomínané štatistické prístupy, ktoré sa použijú na preskúmanie vzťahu medzi jednotlivými pomocnými dazymetrickými zónami a odhadovanou premennou, a to hlavne prostredníctvom regresnej analýzy alebo techniky EM algoritmu.

B2. Metódy cestnej siete

Podobný prístup ako dazymetrická metóda s využitím dát z DPZ, ktorý využíva rastúce a dostupnejšie množstvo digitálnych databáz, navrhli Xie (1995) a Voss et al. (1999). Ak totiž pri probléme transformácie dát z navzájom odlišných zonálnych systémov pracujeme s premennými vzťahujúcimi sa na obyvateľstvo alebo bývanie, je vhodné ako pomocnú informáciu využiť cestnú sieť. Možno vychádzať z predpokladu, že domy sú takmer vždy

lokalizované pri cestách, a preto aj rozmiestnenie obyvateľstva je v úzkom vzťahu s cestnou sieťou. Z toho možno napríklad usudzovať, že čím väčšia je hustota ciest, tým vyššia je hustota zaľudnenia.

Xie (1995) rozvinul tri tzv. „prekrývajúce sa sieťové algoritmy“ (z angl. *overlaid network algorithms*), na základe ktorých sa premenná zdrojových zón pridelí jednotlivým cestným úsekom. Ide o metódu dĺžky cestnej siete, hierarchicky váženej cestnej siete a bytovej zaťažnosti cestnej siete (priradenie premennej jednotlivým cestným úsekom na základe dĺžky, tried a bytovej zaťažnosti cestných úsekov). Všetky sú založené na využití techník GIS a relevantných pomocných informácií, ktoré sa priamo vzťahujú k odhadovanej premennej. Termín *sieť* v tomto prípade znamená, že prerozdelenie atribútových hodnôt vybranej premennej (napr. obyvateľstva) zdrojovej zóny je založené na cestných segmentoch. V terminológii GIS 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 skupine jednorozmerných objektov a o zdrojových a cieľových geografických jednotkách ako o dvojrozmerných objektoch – polygónoch. Termín *sieť* 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 skupiny geografických jednotiek
- prekryvom prienikových zón a vrstvy cestnej siete (prostredníctvom váh vypočítaných napríklad na základe dĺžky či počtu uzlov v cestnej sieti) nastáva transformácia údajov do cieľových zón.

Najjednoduchšou metódou cestnej siete je **metóda dĺžky cestnej siete**, ktorá predpokladá homogénne rozmiestnenie premennej pozdĺž cestnej siete. Vybraná premenná cieľovej zóny sa odhadne na základe sčítania dĺžky ciest v zónach prieniku, ktoré jej prislúchajú, a podielu hodnoty danej premennej vzťahujúcej sa na jednotku dĺžky ciest korešpondujúcich zdrojových zón.

Metóda hierarchicky váženej cestnej siete (alebo jednoduchšie **metóda typov ciest**) berie do úvahy 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 každému typu ciest pridelí váha na základe predpokladanej hustoty zaľudnenia. Hodnota váhy je tým väčšia, čím hustejšie je daný typ ciest zaľudnený. Súčet váh v každej zdrojovej zóne sa musí rovnať 1, váhy sú určené subjektívne (napr. na základe vedomostí o danom území, terénneho výskumu alebo podľa priemerných hodnôt za nejaké väčšie referenčné územie). O tejto metóde sa teda predpokladá, že zohľadnenie jednotlivých druhov cestných úsekov vylepší výsledky predchádzajúcej metódy.

Hoci sa metóda hierarchicky váženej cestnej siete snaží presnejšie vystihnúť rozmiestnenie obyvateľstva pozdĺž jednotlivých cestných úsekov, je založená len na určitých predpokladoch, ktoré aproximujú reálnu situáciu, a nie na skutočných pozorovaniach. Množstvo

dostupných digitálnych informácií a moderné techniky GIS však prispeli k rozvoju ďalšieho prístupu – **metódy bytovej zaťažnosti cestnej siete**. Táto metóda využíva informácie o počte domácností v každej zdrojovej zóne a digitálne dáta zahŕňajúce rozsahy jednotlivých adries, na základe ktorých je možné určiť, koľko domov je lokalizovaných pozdĺž jednotlivých cestných úsekov. Základným princípom tejto metódy je, že počty obyvateľov sú vzťahujú na domy a domy môžu byť pripojené k častiam ulice podľa rozsahu adresy v každej zóne.

Voss et al. (1999) doplnili tieto tri algoritmy sieťových metód o techniku, ktorá uvažuje o počte uzlov v cestnej sieti. Túto metódu 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 priesečníku s ostatnými cestnými úsekmi. Premenná je cieľovým zónam pridelená na základe časti celkového počtu uzlov zdrojovej zóny, ktoré sú lokalizované v každej zóne prieniku.

B3. „Flexibilné“ inteligentné metódy areálovej transformácie geografických dát

Rozoberané metódy využívajú špecifické pomocné informácie, aby lepšie vyjadrili rozmiestnenie skúmanej premennej. Existujú však metódy flexibilnejšie, využívajúce komplexné pomocné údaje, ktoré nepracujú iba s jedným druhom dát, ale s viacerými pomocnými informáciami, prípadne s pomocnými informáciami dostupnými za cieľové geografické jednotky. Príkladom sú areálovo-transformačné metódy založené na EM algoritme, rozvinuté Flowerdewom a Greenom (1989, 1991, 1992 a 1994) a neskôr Gallegom (2010), Griffithom (2013), Schroederom a van Riperom (2013) či Sridharanom a Quiom (2013).

Metódy areálovej transformácie založené na EM algoritme

Patria do skupiny inteligentných metód, ktoré využívajú viacero pomocných informácií dostupných za cieľové geografické jednotky. Za ich zakladateľov sa považuje dvojica autorov Flowerdew a Green (1989, 1991, 1992 a 1994), ktorí rozšírili štatistickú metódu na odhad počtu obyvateľov subzón v rámci zdrojových zón navrhnutú Flowerdewom (1988) a doplnili ju o časť štatistickej teórie známej ako EM algoritmus (Dempster et al. 1977).

Algoritmus EM (EM je akronym vytvorený spojením začiatočných písmen anglických termínov *expectation a maximization*) je všeobecná štatistická technika určená na riešenie problému chýbajúcich dát. Algoritmus sa skladá z dvoch krokov, ktoré sa striedavo opakujú. V kroku E sa vypočíta podmienená očakávaná hodnota chýbajúcich dát za podmienky, že je daný určitý model a pozorované dáta. V kroku M sa daný model kalibruje metódou maximálnej vierohodnosti tak, aby vyhovoval „úplnému“ súboru dát vrátane odhadov získaných v kroku E. Obidva kroky sa potom opakujú dovtedy, pokiaľ algoritmus neskonverguje.

Ak sa problém areálovej interpolácie (resp. areálovej transformácie dát) chápe ako problém odhadu hodnôt y_{zc} pre každú podzónu vytvorenú ako prienik zdrojových a cieľových zón, potom sa môže považovať aj za problém chýbajúcich dát. Hodnota y_{zc} bude vo všeobecnosti chýbajúci údaj, ale riadkové súčty y_z budú známe. Ak sa hodnoty y_{zc} majú odhadnúť na základe rozlohy podzón A_{zc} a pomocnej informácie x_c o cieľových zónach, potom

na odvodenie odhadov hodnôt y_{zc} sa dá použiť algoritmus EM. Samozrejme, v praxi by hodnoty y_{zc} nemali byť rovné nule pre tie kombinácie zdrojových a cieľových zón, ktoré sa skutočne pretínajú. Ako uvádzajú Flowerdew a Green (1991), efektívne použitie pomocnej premennej si v procese interpolácie nevyžaduje jednoznačný kauzálny vzťah so záujmovou premennou.

Ak premenná Y vyjadruje napríklad počet obyvateľov, potom môžeme opodstatnene predpokladať, že y_{zc} má Poissonovo rozdelenie s parametrom μ_{zc} a že tento parameter je funkciou x_c, A_{zc} a množiny neznámych parametrov β . Krok E potom zahŕňa výpočet podmienenej očakávanej hodnoty premennej y_{zc} za podmienky, že sú dané hodnoty y_z a platný model parametra μ_{zc} podľa rovnice:

$$\hat{y}_{zc} = E(y_{zc} | \hat{\mu}_{zc}, y_z) = \frac{\hat{\mu}_{zc} y_z}{\sum_k \hat{\mu}_{zk}}$$

Krok M pozostáva z kalibrácie modelu $\mu = \mu(\beta, x_c, A_{zc})$ metódou maximálnej vierohodnosti, pričom sa predpokladá, že odhady sú nezávislé veličiny s Poissonovým rozdelením. Takto získame hodnoty koeficientov, ktoré podávajú informáciu o tom, ako sú parametre μ_{zc} Poissonovho rozdelenia spojené s pomocnými údajmi a rozlohou územia podzón. Hodnoty μ_{zc} získané v kroku M sa môžu potom využiť v kroku E na odvodenie lepších odhadov hodnôt premennej y_{zc} . Tieto odhady sa zasa môžu opakovane použiť v kroku M a priniesť lepšie odhady parametrov μ_{zc} , pokiaľ algoritmus neskonverguje. Začiatkové hodnoty pre odhady y_{zc} môže poskytnúť metóda priestorového vázenia.

Ak je pomocnou informáciou pre cieľové zóny napríklad geologická charakteristika v tom zmysle, že každú cieľovú zónu budujú buď ílovce ($x_c = 1$) alebo vápence ($x_c = 2$), potom počet obyvateľov môžeme považovať za premennú, ktorá má Poissonovo rozdelenie s parametrom λ_1 pre ílovcové zóny, alebo λ_2 pre vápencové zóny. Vhodný model pre počet obyvateľov podzóny by mohol mať tvar $\hat{\mu}_{zc} = \lambda_c A_{zc}$, kde $c = 1$, ak zóna c je ílovcová a $c = 2$, ak je vápencová. Krok E by sa potom mohol uskutočniť v tvare:

$$\hat{y}_{zc} = y_z \frac{\lambda_c A_{zc}}{\sum_k \lambda_k A_{zk}}$$

pre všetky cieľové zóny k , ktoré pretínajú zdrojovú zónu z . Krok M potom pozostáva z kalibrácie modelu $\mu_{zc} = \lambda_c A_{zc}$, využívajúc \hat{y}_{zc} ako dáta. Po dosiahnutí konvergencie opakovaním krokov E a M môžeme odhadnúť celkový počet obyvateľov cieľovej zóny ako súčet odhadov počtu obyvateľov príslušných podzón:

$$\hat{y}_c = \sum_z \hat{y}_{zc}$$

Túto metódu môžeme v rámci algoritmu EM ľahko generalizovať pre rôzne typy pomocných údajov. Napríklad pomocná premenná X by mohla nadobúdať viac ako dve možné hodnoty. Mohla by byť spojitou premennou alebo by mohla existovať kombinácia niekoľkých

diskrétnych alebo spojitych premenných, ktoré by spoločne poskytovali zlepšené odhady premennej y_{zc} . Ďalšie informácie o týchto možnostiach poskytuje Green (1989).

Doposiaľ prediskutované modely vychádzajú z predpokladu, že záujmová premenná má Poissonovo rozdelenie. To je často odôvodnený predpoklad v prípade sociálnych a demografických dát, ktoré vyjadrujú počty osôb alebo udalostí v každej zóne. V mnohých situáciách však záujmová premenná nevyjadruje celkový počet, ale podiel osôb alebo udalostí, ktoré patria do určitej kategórie. Mohol by nás zaujímať napríklad podiel ekonomicky aktívnych osôb, ktoré sú nezamestnané, alebo podiel registrovaných kriminálnych činov, ktoré sa spáchali za použitia motorových vozidiel. Pre problémy tohto typu je adekvátny predpoklad binomického rozdelenia. Metódy areálovej interpolácie aplikovateľné v tomto kontexte rozvinuli Green (1990) a Flowerdew a Green (1991). V uvedených prípadoch znovu predpokladáme, že dáta o podieloch osôb vyznačujúcich sa určitými charakteristikami sú dostupné pre súbor zdrojových zón, ale odhady sú potrebné pre súbor cieľových zón. V týchto prípadoch nie je dôležitá rozloha, ale celkový počet obyvateľov N_{zc} podzón, ktoré vzniknú ako prieniky zdrojových a cieľových zón. Tu sa usilujeme odhadnúť hodnoty y_{zc} vyjadrujúce počet „úspechov“ (t. j. prípadov, podiel ktorých nás zaujíma), ktoré podľa predpokladu majú binomické rozdelenie s parametrami N_{zc} a p_{zc} , kde p_{zc} je pravdepodobnosť, že prvok skúmaného súboru bude „úspech“. Ak je dostupná pomocná informácia x_c o cieľových zónach, parameter p_{zc} sa odhadne ako funkcia x_c a množiny neznámych parametrov β .

Pre úplnosť dodávame, že Flowerdew a Green (1992 a 1994) navrhli sériu inteligentných metód založených na EM algoritme aj pre spojité premenné s normálnym rozdelením. V tomto prípade predpokladajú, že spojité premenné predstavujú priemernú hodnotu z pozorovaní v každej zdrojovej zóne. Bližšie sa však tomuto prístupu nebudeme venovať.

C. Geografické konverzné tabuľky

Techniku geografických konverzných tabuliek (ďalej aj GKT) možno považovať za určitý nástroj, prostredníctvom ktorého možno aplikovať vybrané metódy areálovej transformácie geografických dát mimo prostredia GIS. Metódu GKT možno použiť obzvlášť v takých prípadoch, keď máme k dispozícii nejaké databázové záznamy, z ktorých sa každý odvoláva na viac než jeden zonálny systém a ktoré disponujú údajmi o zdrojových geografických územiach prieniku medzi oboma zonálnymi systémami (neberú teda do úvahy len rozlohu územia ako metóda priestorového váženía).

GKT predstavujú akýsi zoznam záznamov, ktorý spája zdrojové jednotky s cieľovými. Je to vlastne súbor dát, prostredníctvom ktorého môžu byť prepojené rôzne skupiny geografických jednotiek, a to pomocou položky, ktorá je pre oba súbory spoločná. Niekedy sa GKT zvyknú označovať ako look-up (vyhľadávacie) tabuľky, čo je však všeobecný, ale zavádzajúci termín pre geografické konverzné tabuľky. V rámci GKT možno totiž rozlíšiť jednak look-up, look-down alebo look-across tabuľky (Simpson 2002). Jednotlivé typy GKT uvádzame v ich pôvodnom – anglickom znení, pretože sa nám nepodarilo nájsť ich vhodný jednoslovný slovenský ekvivalent. Keďže GKT vo svojej podstate vyjadrujú vzťah medzi dvoma zonálnymi systémami, potom:

- a) Look-up tabuľky sa používajú na agregáciu menších jednotiek do väčších jednotiek. Zdrojové jednotky sú tu menšie ako cieľové, zdrojová jednotka je celkom obsiahnutá v cieľovej jednotke. Preto ich možno nazvať ako „vyhľadávacie“ tabuľky.
- b) Look-down tabuľky sa používajú na dezagregáciu väčších jednotiek do menších jednotiek. Zdrojové jednotky sú tu väčšie ako cieľové, a teda jedna zdrojová geografická jednotka sa môže rozdeliť medzi viacero cieľových geografických jednotiek.
- c) Look-across tabuľky sa používajú na konverziu do veľkostne podobnej kategórie, možno ich nazvať ako „prehľadávacie“ tabuľky.

Podstatou GKT je odhad veľkosti prieniku medzi zdrojovou a cieľovou jednotkou tak, aby mohli byť údaje prerozdelené medzi oba zonálne systémy. Pod pojmom „veľkosť“ prieniku sa nemusí rozumieť iba skutočná rozloha prienikového územia, ako v prípade metódy priestorového váženía, ale ako vážené kritérium sa môžu použiť aj iné premenné, ktoré určitým spôsobom súvisia s odhadovanou premennou. Môže to byť napríklad počet obyvateľov, domácností alebo adries. GKT môžu mať rôzne formáty a spôsoby konštrukcie, dajú sa však opísať prostredníctvom určitého všeobecného rámca. Vo všeobecnosti pozostávajú z troch stĺpcov, ktoré identifikujú zdrojovú geografickú jednotku, cieľovú geografickú jednotku a tzv. váhu. Váha, ako už bolo spomenuté, udáva na základe nejakého váženého kritéria podiel zdrojovej jednotky, ktorý leží v cieľovej jednotke. Váha môže nadobúdať hodnoty väčšie ako 0 a menšie alebo rovné 1. Váhy pre jednu a tú istú geografickú jednotku môžu po sčítaní nadobúdať maximálnu hodnotu rovnú jednej, pretože každá geografická jednotka môže byť pridelená len raz. Ak je tento súčet vždy rovný jednej, konverzná tabuľka je „vyčerpávajúca“ (úplná) a žiadne dáta sa v procese konverzie nestratili (Simpson 2002).

Príklady geografickej konverznej tabuľky sú uvedené v tab. 1a a 1b. V tab. 1a prekrývajú hypotetické zdrojové jednotky (zóny A, B a C) viac než jednu cieľovú geografickú jednotku (P, Q, R a S), a preto majú viac než jeden záznam. Posledné dve zdrojové geografické jednotky z tab. 1b (zóny C a D) sú celé obsiahnuté v cieľových jednotkách (Q a P), ich váha je rovná 1.

Ak sú zdrojové geografické jednotky oveľa menšie ako cieľové, veľa z nich je v konverznej tabuľke prezentovaných jedným záznamom s váhou rovnou 1 (zdrojová jednotka je celkom obsiahnutá v cieľovej jednotke). Ak sú všetky váhy v konverznej tabuľke rovné 1, tabuľka je hierarchická.

Percentuálny podiel zdrojových jednotiek s váhou rovnou 1 sa nazýva *stupeň hierarchie*. Získame ho sčítaním všetkých váh w_{zr} , ktoré sa rovnajú 1 a vydelením tohto súčtu počtom všetkých zdrojových jednotiek (Simpson 2002). Pokiaľ je stupeň hierarchie rovný 0, tabuľka je vyhľadávacia. Príkladom takejto tabuľky je tab. 1a pre tab. 1b je stupeň hierarchie $100.2/4 = 50 \%$.

Na druhej strane sa rozlišuje aj tzv. *stupeň prispôsobenia*, ktorý sa definuje ako súčet maximálnych (tých najväčších) váh všetkých zdrojových zón vydelený ich počtom a vyjadrený v percentách. Pre tab. 1a je potom stupeň prispôsobenia $100(0,5 + 0,5 + 0,5)/3 = 4,16 \%$. Pre tab. 1b je to $100(0,6905 + 0,7667 + 1 + 1)/4 = 86,4 \%$.

Váha sa aplikuje na zdrojové dáta tak, že sa ňou tieto údaje vynásobia. Hodnoty premennej y_z pre zdrojové geografické jednotky sa skonvertujú na odhad hodnôt premennej y_c pre cieľové geografické jednotky použitím váh w_{zc} zaznamenaných v konverznej tabuľke podľa vzťahu:

$$\hat{y}_c = \sum_z w_{zc} y_z$$

Tab. 1. Príklady geografickej konverznej tabuľky

a)

Zdrojové jednotky z	Cieľové jednotky c	Váha w_{zc}
Zóna A	Zóna P	0.5
Zóna A	Zóna Q	0.25
Zóna A	Zóna R	0.25
Zóna A	Zóna Q	0.5
Zóna A	Zóna S	0.5
Zóna A	Zóna R	0.5
Zóna A	Zóna S	0.5

Upravené podľa Durhamovej (2004)



b)

Zdrojové jednotky z	Cieľové jednotky c	Váha w_{zc}
Zóna A	Zóna P	0.3095
Zóna A	Zóna Q	0.6905
Zóna B	Zóna P	0.7667
Zóna B	Zóna Q	0.2333
Zóna C	Zóna Q	1.000
Zóna D	Zóna Q	1.000

Upravené podľa Simpsona (2002)



Hypotetický príklad konverzie dát je uvedený v tab. 2.

Tab. 2. Príklad konverzie dát použitím údajov z tab. 1b

Zdrojové jednotky z	Hypotetické údaje za zdrojovú jednotku (y_z)	Cieľové jednotky c	Odhadnuté údaje za cieľovú jednotku (y_c)
A	20	P	29
B	30	Q	36
C	5		
D	10		

Upravené podľa Simpsona (2002)

Z tabuľky vyplýva, že máme štyri zdrojové geografické jednotky a dve cieľové geografické jednotky:

$$\begin{aligned} z_1 &= A & c_1 &= P \\ z_2 &= B & c_2 &= Q \\ z_3 &= C \\ z_4 &= D \end{aligned}$$

K dispozícii máme údaje za jednotlivé zdrojové geografické jednotky a z tab. 1b váhy, ktoré určujú mieru prekryvu medzi zdrojovými a cieľovými geografickými jednotkami:

$$\begin{aligned} y_{z1} &= 20 & w_{z1c1} &= 0,6905 & w_{z3c2} &= 1 \\ y_{z2} &= 30 & w_{z1c2} &= 0,6905 & w_{z4c2} &= 1 \\ y_{z3} &= 5 & w_{z2c1} &= 0,7667 \\ y_{z4} &= 10 & w_{z2c2} &= 0,2333 \end{aligned}$$

Hodnoty pre cieľové geografické jednotky odhadneme podľa uvedeného vzorca nasledovne:

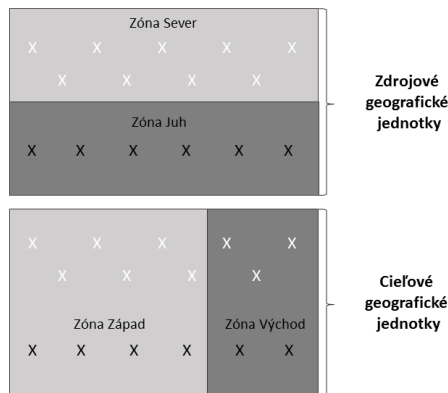
$$\hat{y}_{c1} = w_{z1c1} \cdot y_{z1} + w_{z2c1} \cdot y_{z2} + w_{z3c1} \cdot y_{z3} + w_{z4c1} \cdot y_{z4} = 0,3095 \cdot 20 + 0,7667 \cdot 30 + 0 + 0 = 29$$

$$\hat{y}_{c2} = w_{z1c2} \cdot y_{z1} + w_{z2c2} \cdot y_{z2} + w_{z3c2} \cdot y_{z3} + w_{z4c2} \cdot y_{z4} = 0,6905 \cdot 20 + 0,2333 \cdot 30 + 1 \cdot 5 + 1 \cdot 10 = 36$$

Odhad dát z jedného systému geografických jednotiek do iného, odlišného systému geografických jednotiek si však vždy vyžaduje istú mieru aproximácie. Výsledky pre cieľové geografické jednotky sa preto považujú za odhady a stupne hierarchie a prispôsobenia potom vyjadrujú „veľkosť“ tohto odhadu. Keď zdrojová geografická jednotka celá neleží v cieľovej geografickej jednotke, vždy bude v procese konverzie dát prítomná aproximácia. Vo všeobecnosti sa však preukázalo, že chyby sú menšie, keď sú zdrojové jednotky relatívne menšie ako cieľové jednotky. Potom je väčšina zdrojových jednotiek celkom obsiahnutá v cieľových jednotkách.

Jediná cesta, ako sa vyhneme aproximáciám v procese konverzie dát, spočíva podľa Simpsona (2002) v tom, že zachováme zdrojové dáta ako individuálne záznamy referencované do bodov. Tieto záznamy môžu byť potom hierarchicky zoskupené do akejkoľvek skupiny geografických jednotiek (s váhou rovnou 1), a tak navzájom porovnateľné. Z praktických dôvodov je však väčšina geografických dát geografickými agregátmi. V niektorých krajinách (Veľká Británia, škandinávské krajiny) sa však v rámci sústavy GKT takéto individuálne kódované dáta používajú. Sú nimi zväčša záznamy poštových smerových čísel (PSČ záznamy), ktoré obsahujú prepojenia na administratívne jednotky. Základom tohto prístupu je dezagregácia dát za zdrojové geografické jednotky do skupiny PSČ, ktoré pokrývajú každú zónu prieniku a následná reagregácia údajov do cieľových geografických jednotiek, práve prostredníctvom PSČ, ktoré pokrývajú každú cieľovú jednotku. V GIS predstavujú PSČ bodové entity lokalizované v priestore pomocou mriežkovej referencie ich ťažísk (Norman, 2006c). Príklad využitia PSČ záznamov v procese transformácie dát z jedného zonálneho systému do druhého uvádzame podľa Normana (2006c) nižšie:

Obr. 3 znázorňuje zmenu hypotetickej skupiny geografických jednotiek. Zdrojové geografické jednotky pozostávajú z dvoch zón, zóny *Sever* a zóny *Juh*. Za každú zónu máme k dispozícii skupinu PSČ (9 bielych krížikov predstavuje 9 PSČ v zóne *Sever*, 6 čiernych krížikov zase 6 PSČ v zóne *Juh*). Cieľové geografické jednotky pozostávajú z dvoch zón – zóny *Západ* a zóny *Východ* a zaberajú územie rovnakej rozlohy ako zdrojové geografické jednotky, pričom obsahujú tie isté PSČ (6 PSČ z pôvodnej zóny *Sever* a 4 PSČ zo zóny *Juh* teraz obsahuje zóna *Západ*, zóna *Východ* má 3 PSČ zo zóny *Sever* a 2 zo zóny *Juh*).



Obr. 3. Spojenie geografických jednotiek cez PSČ a konverzia vážených dát
(Upravené podľa Normana 2006c)

Údaje zo zóny *Sever* môžu byť prerozdelené použitím časti PSČ, ktoré „padli“ do zóny prieniku zóny *Sever* a zóny *Západ* (6/9) a zóny *Sever* a *Východ* (3/9). Podobne postupujeme aj pri údajoch zo zóny *Juh*.

Ak budeme uvažovať napríklad o počte obyvateľov, pričom v zóne *Sever* žilo 9 000 a v zóne *Juh* 6 000 osôb, potom sa tieto údaje môžu do novovytvorených zón prerozdeliť nasledovne:

- zóna *Západ* bude mať 10 000 obyvateľov [časť zóny *Sever* (9 000 ob. x 6/9) + časť zóny *Juh* (6 000 x 4/6)],
- zóna *Východ* bude mať 5 000 obyvateľov [časť zóny *Sever* (9 000 ob. x 3/9) + časť zóny *Juh* (6 000 x 2/6)].

Základným predpokladom tohto prístupu je, že rozmiestnenie PSČ sa zhoduje s rozmiestnením obyvateľstva. Keďže však obyvateľstvo nie je podľa PSČ rovnomerne rozmiestnené, rozširujú sa dezagregačné váhy prostredníctvom adries, jednotlivých osôb alebo počtov domácností.

Ako však zdôrazňujú Norman et al. (2003), ak sú údaje humánogeografického charakteru konvertované z jednej skupiny geografických jednotiek do druhej, musia sa konvertovať ako prvotné (tzv. hrubé) dáta, nie ako miery, percentá alebo inak odvodené údaje. Uvažujme napríklad o dvoch zónach *A* a *B*, ktoré sú zdrojovými geografickými jednotkami a miera nezamestnanosti je 20 % pre zónu *A* a 15 % pre zónu *B*. Ak sa časť zóny *A* prideliť zóne *B* a ak je na základe nejakého váženého kritéria $w_{zc} = 0,25$, nemôžeme v tomto prípade postupovať takým spôsobom, že od zóny *A* odčítame 5 % nezamestnanosť (čo je $0,25 * 20\%$) a prideliť ju k zóne *B*. Postupujeme tak, že skonvertujeme hrubé dáta a až potom vyčítame vlastné miery (pozri tab. 3). Týmto spôsobom sa 25 % z čitateľa a menovateľa preniesie zo zóny *A* do zóny *B*. Tento predpoklad je však samozrejme založený na tom, že aj zamestnané a nezamestnané obyvateľstvo je podobne rozmiestnené na stupni podzón (Norman 2006c).

Tab. 3. Konverzia mier medzi rôznymi zonálnymi systémami

	Zdrojové geografické jednotky		Váha konverzie	Cieľové geografické jednotky	
	Zóna A	Zóna B	0,25 (A do B)	Zóna A	Zóna B
Nesprávny prístup Miera nezamestnanosti	20 %	15 %	$0,25 * 20\% = 5\%$	15 %	20 %
Správny prístup Nezamestnaní (v osobách)	40	15	$0,25 * 40 = 10$	30 (40 – 10)	25
Ekonomicky aktívni (v osobách)	200	100	$0,25 * 200 = 50$	150 (200 – 50)	150
Miera nezamestnanosti	20 %	15 %	-	20 % (30/150*100)	16,67 % (25/150*100)

Upravené podľa Normana (2006c)

Na záver treba dodať, že GKT by sa mali vždy datovať k obdobiu, ku ktorému sa vzťahujú. Dôležitá je aj otázka ich údržby. Údržba jednej dobre zdokumentovanej aktuálnej GKT medzi každou dvojicou geografických jednotiek je podľa Simpsona (2002) dôležitejšia ako pokus zahrnúť do nej široký výber vážených kritérií, alebo produkovať mnohonásobné konverzné tabuľky, z ktorých nie všetky sú vždy dostupné.

2. APLIKAČNÁ ČASŤ

V predchádzajúcej časti sme poskytli prehľad množstva metód používaných na transformáciu dát z jedného systému geografických jednotiek do druhého, odlišného systému geografických jednotiek. Ako vyplynulo z charakteristiky tých najpoužívanejších, každá z metód má svoje výhody aj nevýhody a je pomerne ťažké určiť tú najvhodnejšiu. Ak napríklad uvažujeme o počte obyvateľov ako o záujmovej premennej, potom metóda priestorového váženía predpokladá jej homogénne rozmiestnenie. Hoci je to metóda, ktorá sa vyznačuje pomerne jednoduchou aplikáciou, záujmová premenná je v skutočnosti len zriedkavo rovnomerne rozmiestnená. Použitie metódy priestorového váženía sa preto odporúča v tom prípade, keď nemáme k dispozícii žiadne pomocné údaje o rozmiestnení premennej v zdrojových zónach, prípadne keď rozdiel medzi zdrojovými a cieľovými jednotkami nie je až taký veľký.

Pyknofylaktická metóda naopak uvažuje o heterogénnom rozmiestnení údajov, tento postup je však efektívny v tom prípade, ak je zrejmé, že modelované dáta možno považovať za spojito sa meniace v priestore (Flowerdew a Green 1992). Pre väčšinu sociálnych a ekonomických dát je však tento prístup menej efektívny, pretože pre takéto dáta sú typické skôr nespojitosti v ich priestorovom rozložení.

Vo všeobecnosti nemožno ani bodové interpolačné metódy považovať za vhodné na riešenie problému odhadu humánno-geografických dát za rôzne geografické jednotky. Väčšina bodových metód nezachováva celkovú hodnotu odhadovanej premennej a napokon je tu negatívny dosah lokalizácie ťažísk zdrojových zón na hodnoty odhadovanej premennej. Použitie kontrolných bodov na reprezentáciu celej zdrojovej zóny produkuje ďalšie chyby. Bodové interpolačné metódy môžu navyše vytvárať pomerne dramatické vizualizácie rozmiestnenia odhadovanej premennej (Langford et al. 1991) a podobne ako štatistické metódy obsahujú spravidla interaktívne a iteratívne zisťovania priestorového správania študovaného fenoménu. Implementácia týchto prístupov si preto vyžaduje presné a jasné vedomosti o dátach a dokonalé porozumenie priestorovej štatistike.

Metódy, ktoré využívajú princípy štatistického modelovania síce aplikujú uznávané formuly na odhad hodnôt záujmovej premennej, ale používajú náhodný spôsob na vyriešenie neurčitosti problému transformácie geografických dát z navzájom odlišných zonálnych systémov. Predpokladajú, že rozmiestnenie záujmovej premennej možno úplne presne opísať

matematickým jazykom alebo definovať vzorcom. Tento predpoklad síce zvyšuje komplexnosť problematiky, ale na druhej strane limituje jej aplikáciu, pretože s takýmito predpokladmi o štatistických vlastnostiach dát sa len zriedka stretávame v praxi (Wu et al. 2005). Tieto metódy okrem toho prijali aproximujúci prístup, ktorý stanovuje jednoduchú funkciu definovanú pre celé študované územie. Z jednotlivých pozorovaní sa teda odvodí všeobecný vzorec na základe nejakých predpokladaných štatistických pravidiel, ktorý sa aplikuje na individuálne študované jednotky. Počas tohto procesu sa ale zvyčajne stratí pôvodný rozsah odhadovanej premennej, čím sa zredukuje aj presnosť odhadu pre cieľové geografické jednotky v porovnaní so zdrojovými jednotkami.

Metódy založené na EM algoritme síce berú do úvahy hodnoty pomocných premenných, ktoré nemusia byť v priamom vzťahu s odhadovanou premennou, rozvinuté sú však skôr prístupy, ktoré používajú dodatočné informácie dostupné za skupinu cieľových geografických jednotiek. Za cieľové zóny však zvyčajne nemáme k dispozícii vôbec žiadne premenné. Metódy neuronových sietí zase využívajú množstvo rozmanitých údajov na zlepšenie presnosti odhadu hodnôt, ich aplikácia si však vyžaduje presné a jasné vedomosti o dátach, technickú zručnosť a pokročilé teoretické vedomosti.

Pokiaľ ide o ďalšie inteligentné metódy areálovej transformácie geografických dát, (tie ktoré využívajú pomocné informácie), dazymetrické techniky, ktoré využívajú údaje z DPZ ako pomocné informácie, predpokladajú, že rozmiestnenie obyvateľstva sa vzťahuje na jednotlivé kategórie využitia územia, prípadne krajinnej pokrývky. Tieto metódy očakávajú vyššiu hustotu zaľudnenia v mestských a prímestských zónach ako napríklad v poľnohospodárskych alebo zalesnených územiach. Presnosť odhadu v ich prípade však závisí aj od toho, ako boli územia rôzneho využitia zeme resp. krajinnej pokrývky odvodené zo satelitných snímok a následne klasifikované. Nevýhodou niektorých aplikácií dazymetrickej metódy však môže byť aj to, že v niektorých prípadoch je pomerne ťažké harmonizovať priestorové rozlíšenie zdrojových a pomocných skupín geografických jednotiek na to, aby bolo možné uskutočniť vyhovujúce pridelenie hodnoty premennej pomocným dazymetrickým zónam, ak sa chceme vyhnúť ich subjektívnemu určeniu (z tohto dôvodu nemožno v našej modelovej situácii použiť napr. techniku smplovania rozvinutú Mennisom a Hultgreenom (2005)). Metódy cestnej siete zase predpokladajú, že obyvateľstvo súvisí s hustotou cestnej siete. Hierarchická metóda cestnej siete berie okrem toho do úvahy aj triedy jednotlivých ciest, kde sa opäť stretávame s problémom subjektivity určenia váh jednotlivých kategórií cestných úsekov.

2.1. Výber metód areálovej transformácie geografických dát

Prístupy k riešeniu problematiky transferu a odhadu hodnôt premennej za odlišné zonálne systémy, ako aj výber konkrétnej metódy teda závisí od rôznych faktorov, predovšetkým však od územných jednotiek, požadovaných údajov a od prijatých predpokladov o týchto entitách. Nemalou mierou však k rozhodnutiu, akú metódu použiť, prispievajú bohužiaľ aj také skutočnosti, akými sú dostupnosť dát, softvérových možnosti, ale aj časová náročnosť, ľahká implementácia či odhadovaná presnosť. Aj vzhľadom na tieto faktory sme sa rozhodli pre nasledovné metódy:

Z jednoduchých metód areálovej transformácie geografických dát sme uprednostnili priestorové vázanie ako tradičnú a štandardnú areálovo-interpoláčnú techniku, ktorá patrí medzi najviac používané metódy odhadu dát za nekompatibilné územné jednotky. Jej algoritmus je jasne formulovaný a jednoducho aplikovateľný. Výber pyknofylaktickej metódy by si vyžadoval porovnanie s jej vylepšenou verziou, ktorá namiesto pravidelnej mriežky používa triangulačnú sieť, čo má, ako uvádza Rase (2001), viacero výhod oproti klasickej Toblerovej metóde.

Pokiaľ ide o inteligentné prístupy, z rôznych implementácií dazymetrických metód sme sa z podobných príčin rozhodli pre jej základnú formu – binárnu dazymetrickú metódu, ktorá sa v mnohých štúdiách ukázala ako pomerne presná a dokonca poskytla lepšie výsledky ako niektoré jej „vylepšené“ verzie. V rámci binárnej dazymetrickej metódy sme zvolili dva, resp. tri typy pomocných informácií (pokiaľ o metódach cestnej siete uvažujeme ako o dazymetrickej technike): údaje o krajinej pokrývke Slovenska z databázy CLC (CORINE Land Cover), podrobnejšie údaje o budovách z databázy ZBGIS (Základná báza údajov pre geografický informačný systém) a vrstvu ciest z Digitálneho mapového diela SVM 50 (Spojitá vektorová mapa Slovenska). Z ostatných dazymetrických techník sme pracovali len s jednoduchou verziou v rámci metódy typov cestnej siete, ktorá pracuje so subjektívne určenými váhami. Zložitejšie techniky v predkladanej monografii nepoužívame z viacerých dôvodov: existujúce prístupy sú zväčša navrhnuté na prácu s menšími územnými jednotkami než s akými sme pracovali my a pri ich realizácii by vznikalo množstvo dodatočných otázok, ktoré by bolo potrebné zodpovedať. Medzi inými aj to, ako presne sú pomocné zóny klasifikované, ako ich vhodne zaradiť do jednotlivých kategórií, alebo na základe čoho im prisúdiť jednotlivé váhy. Štatistické metódy, či už samotné globálne regresné metódy, prípadne ich lokálne aplikácie v rámci dazymetrickej metódy, EM algoritmus a pod. sme sa rozhodli nepoužiť vzhľadom na množstvo dodatočných krokov potrebných na ich realizáciu (výber vhodných modelov, ich aplikácia, zahrnutie dodatočných techník na zachovanie pyknofylaktickej vlastnosti, nevhodné dáta). Novému prístupu k regionálnemu odhadovaniu koeficientov hustoty zaľudnenia v rámci dazymetrickej metódy (iteratívny dezagregačný algoritmus) na Slovensku sú venované práce Rosina (2015) a Rosina a Hurbánek (2016). Vzhľadom na zložitosť a žiaduce vysoké odborné, ako aj teoretické vedomosti sme zatiaľ upustili aj od modernej priestorovo-interpoláčnej metódy neurónových sietí.

V rámci nášho výskumu sme použili aj prístup, na ktorom sú založené geografické konverzné tabuľky a načrtli sme možnosť ich využitia v takom prípade, keď existujú prepojenia medzi oboma zonálnymi systémami a zároveň aj vhodné databázové záznamy súvisiace s odhadovanou premennou dostupné za skupinu zdrojových geografických jednotiek a čo je najpodstatnejšie, aj za územia ich prienikov s cieľovými geografickými jednotkami. To je možné v praxi uskutočniť, ale len v takom prípade, keď sa používajú bodovo-referencované rozsiahle záznamy s prepojeniami na množstvo zonálnych systémov, na základe ktorých je potom pomerne jednoducho určiť váhy na odhad neznámych údajov. Údaje takéhoto typu, ako sú napríklad záznamy PSČ vo Veľkej Británii, u nás zatiaľ použiť nemôžeme. Napriek tomu sa aspoň v teoretickej rovine pokúsime načrtnúť možnosti takéhoto prístupu a zároveň poukázať na fakt, že ak disponujeme vhodnými údajmi, ktoré úzko súvisia s odhadovanou premennou, výsledky procesu odhadu dát môžu byť pomerne presné.

Namiesto PŠČ použijeme obce, pretože pracujeme so zdrojovými a cieľovými jednotkami založenými na vnorenej hierarchii. To znamená, že ak vieme, ktoré obce ležia v prieniku zdrojových a cieľových zón, potom môžeme využiť známe hodnoty za obce na určenie váhy potrebnej na odhad vybraných premenných. GKT pracujú v princípe rovnako ako metóda priestorového váženia, premenné za zdrojové geografické jednotky sú však vážené nie na základe rozlohy územia, ale podľa ľubovoľnej inej premennej, ktorá je v nejakom vzťahu s odhadovanou premennou.

2.2. Dáta použité na testovanie vybraných metód areálovej transformácie

Väčšina dát humánogeografickej povahy sa na Slovensku zhromažďuje za cenzové štatistické jednotky (osoby, rodiny, byty, domy a pod.), ale z rôznych dôvodov (zredukovanie rozsahu publikovaných údajov, ale predovšetkým v dôsledku zachovania anonymity jednotlivých subjektov – ochrany osobných údajov) sa publikujú za územné jednotky určitého stupňa priestorovej agregácie (obce, okresy, kraje). Pre náš výskum bolo vybrané celé územie Slovenska, na ktorom sme sa pokúsili aplikovať vybrané metódy areálovej transformácie geografických dát za navzájom odlišné zonálne systémy. Zdrojové geografické jednotky, teda územia, za ktoré sme mali k dispozícii údaje, predstavovali územnosprávne jednotky – okresy s hranicami vo vymedzení v rokoch 1991 a 2001. Za cieľové geografické jednotky, teda zóny, pre ktoré tieto dáta potrebujeme, sme zvolili systém funkčných mestských regiónov (FMR) 91-B navrhnutý Bezákom (2000). Vybrané metódy sme sa rozhodli použiť zvlášť za okresy z roku 1991 a 2001, teda v dvoch modelových situáciách, v snahe čo najlepšie testovať a systematicky zhodnotiť presnosť metód, keďže v roku 1996 nastala zmena územnosprávneho členenia Slovenskej republiky a počet okresov sa zmenil z 38 v roku 1991 na 79 okresov v roku 2001. Bude preto zaujímavé sledovať, či a ako vplyva veľkosť/rozsah zdrojových a cieľových geografických jednotiek na proces areálovej transformácie dát, teda s akou presnosťou sa nám podarí odhadnúť dáta z menšieho počtu zdrojových geografických jednotiek (38 okresov z roku 1991) do väčšieho počtu zón (66 FMR) a naopak, z väčšieho počtu zdrojových jednotiek (79, resp. 72² okresov z roku 2001) do menšieho počtu cieľových zón (66 FMR). Zdrojové a cieľové geografické jednotky sú znázornené na obr. 4.

Pre FMR ako pre cieľové geografické jednotky, do ktorých budeme transformovať vybrané dáta z okresov, sme sa rozhodli, pretože:

1. Na testovanie presnosti jednotlivých metód sme na jednej strane potrebovali regióny, ktorých hranice neboli kompatibilné s okresmi. Zároveň to však museli byť také územia, za ktoré sme poznali skutočné hodnoty odhadovaných dát. FMR 91-B predstavujú systém regiónov, za ktoré nemáme primárne k dispozícii žiadne údaje. Keďže je to však systém, ktorý sa dá vyskladať z obcí a dáta za obce poznáme a máme ich očistené o územné zmeny, ktoré jednotlivé obce v období rokov 1991 – 2001 prekonali, bude možné odhadnuté hodnoty premenných za FMR porovnať so skutočnými hodnotami (teda takými, kto-

² mestské časti Bratislavy I – V a Košíc I – IV boli zlúčené do okresu Bratislava a Košice

ré získame agregáciou dát za obce) a tým zhodnotiť presnosť jednotlivých metód areálovej transformácie za rôzne zonálne systémy;

2. Sú to geografické jednotky, ktoré uvádzame ako príklad vhodnejšieho nástroja geografickej analýzy než akým sú samotné okresy, a to z viacerých dôvodov (podľa Bezáka 2000):

- systém FMR umožňuje prezentáciu výsledkov populačných cenzov a bežnej štatistickej evidencie v omnoho realistickejšej forme ako územnosprávne jednotky – je cenným analytickým nástrojom,
- je to fixný regionálny systém nezávislý od akýchkoľvek zmien územnosprávnej organizácie,
- poskytuje príležitosť preniknúť do vnútroregionálnej štruktúry na konzistentne definovanej báze,
- umožňuje korektné medziregionálne porovnávanie (FMR sú definované na základe kritérií, ktoré sa dôsledne aplikujú v rámci celej krajiny),
- existuje pomerne ľahká aktualizácia FMR so cieľom odzrkadlenia prípadných zmien regionálnej štruktúry.

Pre vzájomné porovnanie spoľahlivosti každej metódy ponúkame v predkladanej monografii okrem v literatúre tradične interpolovanej premennej – počtu obyvateľov (PO) – aj odhad ďalších dvoch premenných – počtu obyvateľov v predproduktívnom veku (POPRED) a počtu ekonomicky aktívnych obyvateľov pracujúcich v poľnohospodárstve (EAOPOL). Hodnoty uvedených premenných za okresy z roku 1991 a 2001 sme získali z jednotlivých sčítaní obyvateľstva.

Pri dazymetrickej metóde sme pracovali s vektorovými údajmi o krajinskej pokrývke Slovenska z databázy CORINE Land Cover (CLC) z rokov 1990 a 2000 (SAŽP 1996 a 2004), ďalej s triedou objektov *budova* z databázy ZBGIS z roku 2008 a s vrstvou *cesty* z Digitálneho mapového diela SVM 50 (roky 1991 – 1997).

Pod krajinnou pokrývkou (land cover) sa rozumejú tak prírodné, ako aj človekom pretvorené či vytvorené objekty zemského povrchu rôznej hierarchickej dimenzie, ktoré možno rozpoznať na satelitných snímkach prostredníctvom fyziognomických a morfoštruktúrnych znakov. Krajinná pokrývka teda predstavuje zhmotnený priemet prírodných priestorových daností a zároveň súčasného využívania krajiny človekom resp. človekom pretvorenej alebo vytvorenej krajiny (Feranec a Oťaheľ 2001). Tematická vrstva CLC90 Slovenska bola vytvorená vizuálnou interpretáciou obrazových záznamov LANDSAT TM z obdobia 1989 – 1992. Jej aktualizáciou na podklade satelitných snímkov LANDSAT 7 ETM vznikla vrstva CLC2000 (Feranec a Oťaheľ 2003). Identifikáciou objektov zo satelitných snímkov sa získali informácie o polohe a rozlohe areálov 44 tried krajinskej pokrývky Európy, z ktorých sa na Slovensku vyskytuje 31. Tieto triedy sú hierarchicky zaradené do štyroch úrovní zostavených podľa mierky mapovania. Prvá úroveň obsahuje päť základných tried – umelé povrchy, poľnohospodárske areály, lesné a poloprírodné areály, zamokrené areály a vody, ktoré sa ďalej členia do 15 tried druhej úrovne. Ostatné úrovne sú detailnejšie rozpracované, najpodrobnejšia je legenda 4. úrovne, ktorá akceptuje národné zvláštnosti krajinskej pokrývky Českej republiky, Maďarska, Poľska a Slovenska. V rámci binárnej dazymetrickej metódy sme pracovali s kategóriou *1.1. urbanizovaná zástavba* z druhej hierarchickej úrovne.

Zdrojové geografické jednotky

Okresy 1991

Okresy 2001



Cielové geografické jednotky

FMR 91-B



1 Bánovce nad Bebravou	23 Martin	45 Sobrance
2 Banská Bystrica	24 Medzilaborce	46 Spišská Nová Ves
3 Banská Štiavnica	25 Michalovce	47 Spišská Stará Ves
4 Bardejov	26 Myjava	48 Stará Ľubovňa
5 Bratislava	27 Námestovo	49 Stropkov
6 Brezno	28 Nitra	50 Svidník
7 Čadca	29 Nová Baňa a Žarnovica	51 Šahy
8 Doľný Kubín	30 Nové Mesto nad Váhom	52 Šaľa a Galanta
9 Dubnica nad Váhom	31 Nové Zámky	53 Stúrovo
10 Dunajská Streda	32 Piešťany	54 Topoľčany
11 Gelnica a Prakovce	33 Poprad	55 Tornala
12 Giraltovce	34 Považská Bystrica	56 Trebišov
13 Hlohovec	35 Prešov	57 Trenčín
14 Hnúšťa	36 Prievidza	58 Trnava
15 Humenné	37 Púchov	59 Tvrdošín a Nížná
16 Komárno	38 Revúca	60 Veľké Kapušany
17 Košice	39 Rimavská Sobota	61 Veľký Krtíš
18 Kráľovský Chlmec	40 Rožňava	62 Vranov nad Topľou
19 Krupina	41 Ružomberok	63 Zlaté Moravce
20 Levice	42 Senica	64 Zvolen
21 Liptovský Mikuláš	43 Skalica a Holíč	65 Žiar nad Hronom
22 Lučenec	44 Snina	66 Žilina

Obr. 4. Zdrojové a cieľové geografické jednotky – okresy a FMR 91-B

Okrem údajov o urbanizovanej zástavbe disponujeme aj presnejšími pomocnými informáciami vzťahujúcimi sa na obyvateľstvo z databázy ZBGIS. ZBGIS je priestorovou objektovo orientovanou bázou údajov, ktorú tvorí a zabezpečuje Úrad geodézie kartografie a katastra SR na základe zákona č. 215/1995 Z. z. o geodézii a kartografii v znení neskorších predpisov. ZBGIS tvoria údaje a metaúdaje o objektoch krajiny, ich priestorových a tematických atribútoch a vzájomných väzbách. Túto databázu sme využili len v druhej modelovej situácii, keďže sme pracovali s katalógom tried objektov z verzie 2008 (staršie dáta viažuce sa približne k roku 2001 sme nemali k dispozícii). Z databázy sme vyselektovali triedu *budova* a v rámci nej typ 16 *rodinný dom* (RD) a typ 17 *bytový dom* (BD). Pri odhade premennej EAOPOL sme okrem toho použili aj typ 19 *poľnohospodárska budova* (PB) a typ 20 *skleník*. Vychádzali sme pritom z predpokladu, že medzi danou premennou a touto váhou existuje užší vzťah než s univerzálnou pomocnou informáciou (typ 16 a 17).

Pokiaľ ide o pomocné informácie v rámci testovaných metód cestnej siete, využili sme údaje z Digitálneho mapového diela SVM 50, ktoré vzniklo na báze Základnej mapy Slovenskej republiky v mierke 1:50 000 – ZM50 (Geodetický a kartografický ústav Bratislava). Geografické objekty, ktoré sa na ZM50 nachádzali, boli rozdelené do jednotlivých vrstiev. Pracovali sme s vrstvou cesty, ktorá rozlišuje dve triedy prvkov – línie a uzly. Vrstva línie obsahuje deväť typov ciest – diaľnica, cesta 1. triedy, cesta 2. triedy, cesta 3. triedy, poľná a lesná cesta udržiavaná, poľná a lesná cesta neudržiavaná, chodník, hlavná cesta v obci a vedľajšia cesta v obci. Z nich sme v rámci metód cestnej siete pre obe modelové situácie použili triedy *diaľnica, cesta 1. triedy, cesta 2. triedy, cesta 3. triedy, hlavná cesta v obci a vedľajšia cesta v obci*, ktoré sa vzťahujú na obdobie rokov 1991 až 1997.

V rámci geografických konverzných tabuliek sme použili údaje o počte *cenзовých domácností* (CD) z jednotlivých sčítaní obyvateľstva ako univerzálnu pomocnú informáciu vhodnú na odhad akýchkoľvek premenných vzťahujúcich sa na obyvateľstvo. Okrem nich sme sa ako vážené kritérium rozhodli aplikovať aj špecifickejšie dáta, konkrétne počet *ekonomicky aktívnych obyvateľov* (EAO) a *rozlohu zastavanej plochy* (ZP) z databázy Úhrnné hodnoty druhov pozemkov (ÚHDP) 1987 a 2001 s cieľom otestovať, ako sa budú takéto, zdanlivo (ne)súvisiace dáta v jednotlivých situáciách správať. Domnievame sa totiž, že medzi vybranými premennými a váhami existuje vzťah rôznej intenzity, ktorý má vplyv na proces odhadu dát. Tento počiatočný predpoklad sme sa pokúsili podložiť prostredníctvom matice korelácie medzi odhadovanými premennými a jednotlivými váhami. Koeficienty (Pearsonov korelačný koeficient) sa vypočítali za zdrojové geografické jednotky a sú uvedené v tab. 17 na str. 95. Medzi premennou PO a váhou EAO dosahoval Pearsonov korelačný koeficient v oboch prípadoch (modelových situáciách) hodnotu 1; vysoká korelácia sa javí aj pri váhe CD (0,99). Premenná PO a váha ZP sú naopak v slabšom vzťahu (0,55; resp. 0,69). Podobne vystupuje aj premenná POPRED (silný vzťah s EAO – 0,95, resp. 0,98 a CD – 0,94, resp. 0,96; slabší so ZP – 0,7, resp. 0,5). Naproti tomu, premenná EAOPOL vykazuje veľmi slabú koreláciu s váhami CD a EAO (0,09; 0,27, resp. 0,08 a 0,37) a vysokú s váhou ZP (0,79 a 0,83). Neskôr, pri samotnom hodnotení úspešnosti jednotlivých metód GKT, sa pokúsime overiť, či metódy vážiace podľa informácií, o ktorých sa predpokladalo tesnejšie prepojenie medzi nimi a odhadovanými premennými, poskytnú presnejšie odhady daných premenných než ostatné metódy. Napríklad pri premenných PO a POPRED by mala byť metóda GKT vážiaca podľa EAO a CD porovnateľná a zároveň úspešnejšia ako metóda GKT používajúca ZP. Pri premennej EAOPOL by mala naopak metóda GKT vážiaca podľa ZP poskytnúť oveľa presnejšie odhady ako metóda GKT vážiaca podľa EAO a univerzálnej pomocnej informácie CD.

2.3. Implementácia vybraných metód areálovej transformácie

Všetky aplikované metódy pracujú na podobnom princípe, líšia sa len v tom, akú váhu používajú na prerozdelenie premennej zo zdrojových do cieľových geografických jednotiek. Základom každej metódy je prekryv zdrojových a cieľových zón. Naložením vrstvy 38 okresov z roku 1991 a 72 okresov z roku 2001 (okresy Bratislava I-V a Košice I-IV sú

zlúčené a uvádzané ako jeden, teda ako Bratislava a Košice) a 66 FMR 91-B sme získali v prvom prípade 93 a v druhom prípade 122 zón prieniku, ktoré možno vidieť na Obr. 5 a 6 (v prílohe).

Zóny prieniku obsahujú ID kódy – prepojenia na zdrojové aj cieľové geografické jednotky. V závislosti od použitej metódy sa vypočítala váha, ktorá sa aplikovala na premenné zdrojových geografických jednotiek. Reagregáciou takto získaných dát sa nakoniec získal odhad daných premenných pre cieľové zóny. Venujme sa však postupne každej metóde zvlášť.

2.3.1. Priestorové váženie

Metóda priestorového váženia odhaduje vybranú premennú na základe územia prieniku A_{zc} medzi zdrojovou (okres) a cieľovou (FMR 91-B) zónou. Osobitne za každú testovanú situáciu (okresy 1991 → FMR 91-B a okresy 2001 → FMR 91-B) sa v prostredí ArcGIS 9.3 vytvorili zóny prieniku a za každú zónu sa vypočítala jej rozloha. Hodnoty premennej za FMR 91-B sa potom odhadli na základe známych hodnôt premennej y_z v okresoch a z tzv. váhy w_{zc} udávajúcej podiel rozlohy zdrojovej zóny z , ktorá leží v cieľovej zóne c (rozloha A_{zc}) a celkovej rozlohy zdrojovej zóny A_z . Vypočíta sa ako:

$$w_{zc} = \frac{A_{zc}}{A_z}$$

Následne sa váha aplikovala na hodnotu premennej zdrojovej zóny y_z , čím sa získali hodnoty premennej y_{zc} za zóny prieniku. Ich reagregáciou do cieľových geografických jednotiek sa získal celkový odhad premennej vzťahujúci sa k príslušnému FMR 91-B:

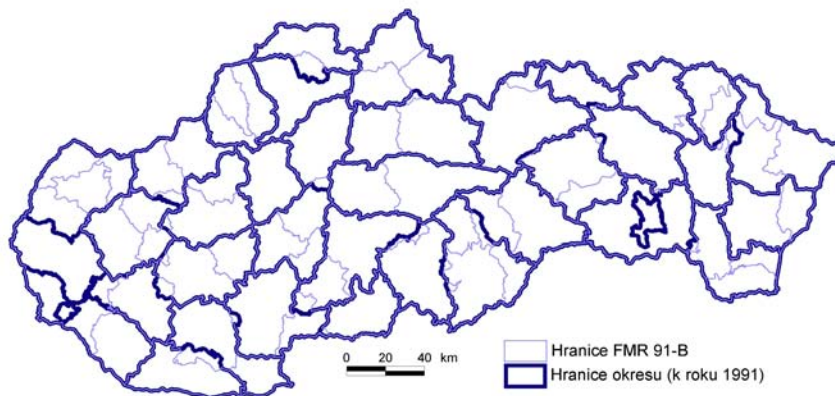
$$\hat{y}_c = \sum_z \hat{y}_{zc} = \sum_z w_{zc} y_z$$

Princíp aplikácie metódy priestorového váženia na odhad hodnôt premennej demonštrujeme na príklade tabuliek Tab. 5a (prvá testovaná situácia) a 5b (druhá testovaná situácia). Tabuľky neobsahujú všetky zdrojové a cieľové geografické jednotky, rovnako ani obe odhadované premenné. V dôsledku ich veľkého rozsahu uvádzame len vzorové tabuľky, ktoré na vybraných zdrojových geografických jednotkách načrtávajú spôsob odhadu na príklade premennej PO.

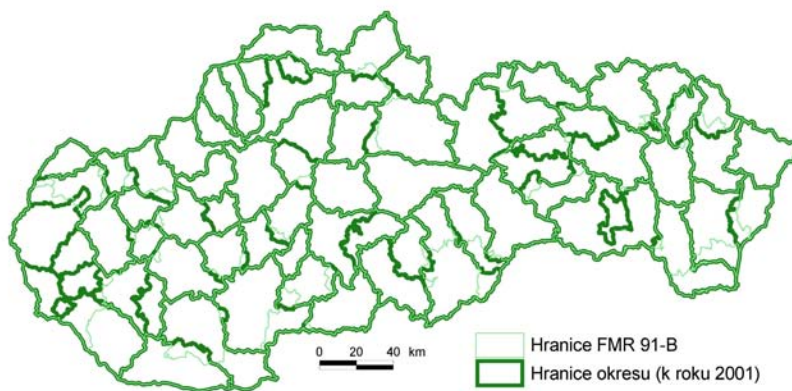
2.3.2. Binárna dazymetrická metóda

Pri binárnej dazymetrickej metóde sme v prvom rade pracovali s územiaми súvislej a nesúvislej zástavby z databázy CLC 1990 a 2000, ktoré sme prekryli so zdrojovými geografickými jednotkami – okresmi, čo nám umožnilo identifikovať zastavané polygóny v každej zdrojovej zóne (obr. 7 a 8). Keďže okrem údajov o urbanizovanej zástavbe disponujeme aj presnejšími pomocnými informáciami vzťahujúcimi sa na obyvateľstvo, rozhodli sme sa tieto dáta využiť a otestovať v druhej modelovej situácií (odhad z okre-

Obr. 5. Hranice okresov v roku 1991 a FMR 91-B: zóny prieniku



Obr. 6. Hranice okresov v roku 2001 a FMR 91-B: zóny prieniku



Tab. 5. Vzorová tabuľka odhadu počtu obyvateľov metódou priestorového váženía za vybrané zóny priteniku:

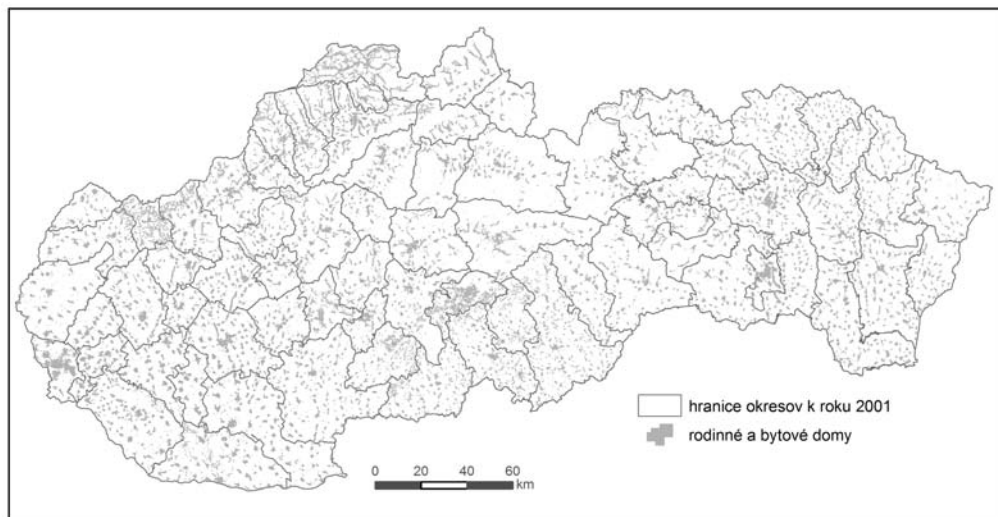
a) rok 1991

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ y _{...}		CIEĽOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY w		ODHAD y _{...}	
Kód okresu	Názov okresu	Rozloha okresu (ha)	PO	Kód FMR 91-B	Názov FMR 91-B	Rozloha FMR 91-B (ha)	Rozloha priteniku (ha)	Váha	w*PO
5100	Bratislava-mesto	36753,67	442197	B5	Bratislava	258763,00	36753,67	1	442197
5201	Bratislava-vidiek	124110,76	145301	B5	Bratislava	258763,00	124110,76	1	145301
5202	Dunajská Streda	107507,24	109345	B5	Bratislava	258763,00	32150,37	0,299053	32700
5202	Dunajská Streda	107507,24	109345	B10	Dunajská Streda	75356,87	75356,87	0,700947	76645
5203	Galanta	98582,56	143846	B5	Bratislava	258763,00	15200,87	0,154194	22180
5203	Galanta	98582,56	143846	B52	Šaľa a Galanta	90903,48	82937,19	0,841297	121017
5203	Galanta	98582,56	143846	B58	Trnava	76934,63	444,51	0,004509	649
5204	Komárno	109994,67	109279	B16	Komárno	94074,38	94074,38	0,855263	93462
5204	Komárno	109994,67	109279	B31	Nové Zámky	103789,22	15920,30	0,144737	15817
5205	Levice	155033,19	120703	B20	Levice	124835,35	122461,07	0,789902	95344
5205	Levice	155033,19	120703	B51	Šahy	44034,33	32572,120	0,210098	25359

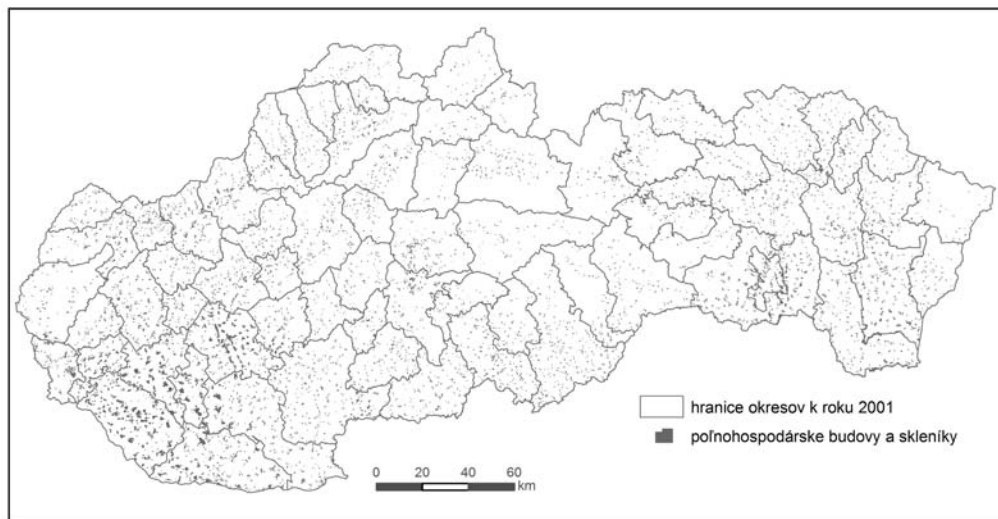
b) rok 2001

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ y _{...}		CIEĽOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY w		ODHAD y _{...}	
Kód okresu	Názov okresu	Rozloha okresu (ha)	PO	Kód FMR 91-B	Názov FMR 91-B	Rozloha FMR 91-B (ha)	Rozloha priteniku (ha)	Váha	w*PO
100	Bratislava (I-V)	36753,67	428672	B5	Bratislava	258763,00	36753,67	1	428672
106	Malacky	95010,62	64354	B5	Bratislava	258763,00	95010,62	1	64354
107	Pezinok	37549,20	54164	B5	Bratislava	258763,00	37549,20	1	54164
108	Senec	35902,90	51825	B5	Bratislava	258763,00	35902,90	1	51825
201	Dunajská Streda	107507,24	112384	B5	Bratislava	258763,00	32150,37	0,299053	33609
201	Dunajská Streda	107507,24	112384	B10	Dunajská Streda	75356,87	75356,87	0,700947	78775
202	Galanta	64230,60	94533	B5	Bratislava	258763,00	8492,84	0,1322243	12500
202	Galanta	64230,60	94533	B52	Šaľa a Galanta	90903,48	55293,25	0,8608552	81379
202	Galanta	64230,60	94533	B58	Trnava	76934,63	444,51	0,0069205	654
203	Hlohovec	26736,31	45351	B13	Hlohovec	28216,80	23402,70	0,875315	39696
203	Hlohovec	26736,31	45351	B32	Piešťany	47515,08	931,211363	0,0348295	1580
203	Hlohovec	26736,31	45351	B58	Trnava	76934,63	2402,404938	0,0898555	4075
204	Piešťany	38083,04	63928	B32	Piešťany	47515,08	38083,037754	1	63928

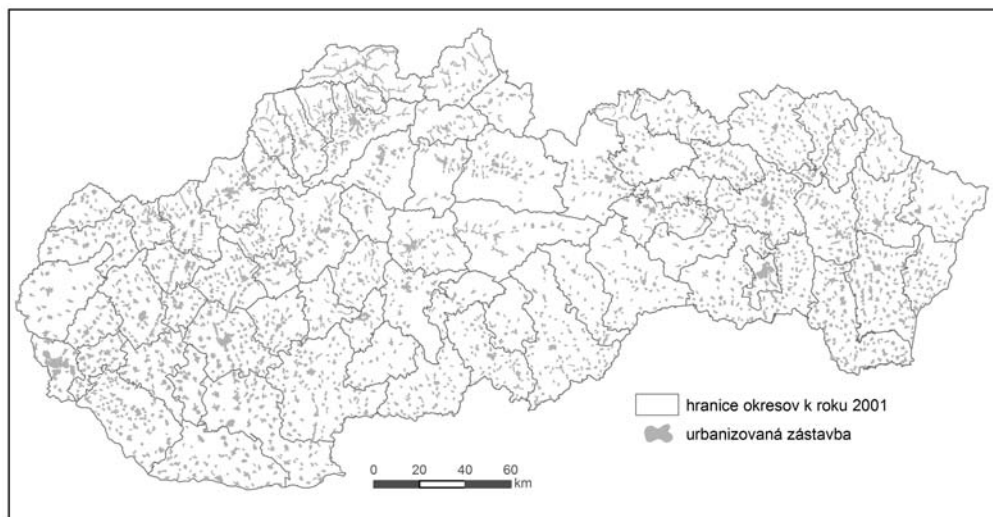
sov 2001 do FMR 91-B). Obývané územie v tomto prípade reprezentujú budovy typu rodinný a bytový dom z databázy ZBGIS (obr. 9). Pri odhade premennej EAOPOL sme okrem toho z už vysvetlených príčin využili aj typ poľnohospodárska budova a skleníky (obr. 10).



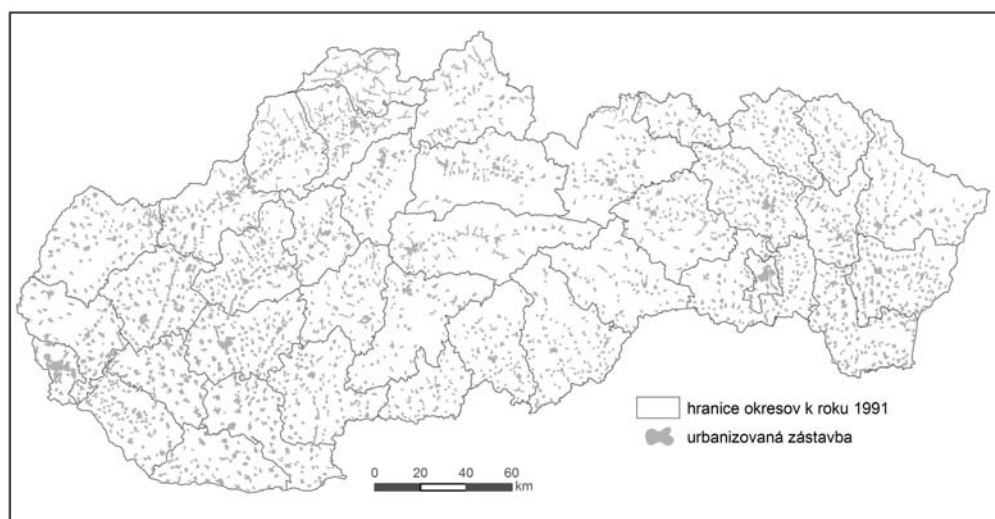
Obr.7. Prekrýv okresov 1991 a triedy 1.1. urbanizovaná zástavba z databázy CLC90



Obr. 8. Prekrýv okresov 2001 a triedy 1.1. urbanizovaná zástavba z databázy CLC00



Obr. 9. Prekryv okresov 2001 s vrstvou rodinných a bytových domov z databázy ZBGIS



Obr. 10. Prekryv okresov 2001 s vrstvou poľnohospodárskych budov a skleníkov z databázy ZBGIS

Prekryvom vrstvy pomocných zón a vrstvy cieľových geografických jednotiek (FMR 91-B) sme získali vrstvu zón prieniku. Za každú zónu prieniku sa podobne ako v metóde priestorového váženía vypočítala váha – namiesto celkovej rozlohy prienikového územia sa ale pracovalo s rozlohou zastavaných polygónov. Aplikáciou váhy na hodnoty premenných

zdrojových zón a následnou reagregáciou hodnôt premenných za zóny prieniku sa získali odhady premenných za cieľové geografické jednotky.

Odhad hodnôt premenných demonštrujeme na príklade odhadu počtu obyvateľov za vybrané geografické jednotky v tab. 6a,b. Vo vzorových tabuľkách neuvádzame vždy tie isté zdrojové zóny a zóny prieniku, ale jednotlivé zóny postupne striedame, aby sme zobrazili všetky možné vzťahy medzi zdrojovými a cieľovými geografickými jednotkami.

2.3.3. Metódy cestnej siete – metóda dĺžky a typov ciest

Oba typy metód cestnej siete pracujú na základe váh zostrojených prekryvom okresov s vrstvou cestnej siete získanej zo SVM 50 Slovenska (obr. 11 a 12). Pri aplikácii týchto metód sa bralo do úvahy 6 typov ciest: diaľnice, cesty 1., 2. a 3. triedy, hlavné a vedľajšie cesty v obci. Poľné cesty, lesné cesty a chodníky sa nepoužili.

V prípade metódy dĺžky ciest sa pracovalo len s dĺžkou jednotlivých cestných úsekov. Metóda pozostávala z niekoľkých krokov:

- a) v prvom rade sa pomocou jednoduchšej funkcie ArcGIS vypočítala dĺžka ciest za zdrojové zóny D_z ,
- b) následne sa jednotlivým cestným úsekom v zdrojovej zóne pridelila hodnota odhadovanej premennej na jednotku dĺžky ciest zdrojovej zóny y_z / D_z ,
- c) na základe prekryvu zdrojových a cieľových geografických jednotiek na jednej strane a vrstvou ciest na strane druhej sa určila dĺžka ciest každej zóny prieniku D_{zc} ,
- d) vynásobením dĺžky ciest každej zóny prieniku hodnotou premennej prepočítanej na jednotku dĺžky ciest korešpondujúcej zdrojovej zóny sa odhadla hodnota premennej za zónu prieniku y_{zc} .
- e) Výsledný odhad hodnoty premennej za cieľovú zónu sa vypočítal podľa rovnice:

$$\hat{y}_c = \sum_z \frac{y_z}{D_z} D_{zc} = \sum_z y_z w_{zc} = \sum_z \hat{y}_{zc}$$

V rámci metódy hierarchicky váženej cestnej siete (alebo jednoduchšie metódy typov ciest) sa vytvorila jednoduchá hierarchia, na základe ktorej sa každému typu cesty expertným odhadom priradila váha, ktorá odráža 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. Súčet váh v každej zdrojovej jednotke musí byť rovný 1. Ďalší postup bol potom podobný ako v prípade metódy dĺžky ciest, a to:

- a) prekryvom vrstiev zdrojových zón a cestných úsekov sa určila dĺžka každého typu cesty v každej zdrojovej zóne D_i^z ;

Tab. 6. Vzorová tabuľka odhadu počtu obyvateľov binárnou dazymetrickou metódou za vybrané zóny prieniku:

a) rok 1991

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ y_z		CIELOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY w		ODHAD y_x
Kód okresu	Názov okresu	Rozloha okresu (ha)	PO	Kód FMR 91-B	Názov FMR 91-B	Rozloha triedy 11 v prieniku (ha)	Váha w	w*PO
5206	Nitra	11359,01	211517	B13	Hlohovec	154,52	0,0136031	2877
5206	Nitra	11359,01	211517	B28	Nitra	8081,59	0,71147	150488
5206	Nitra	11359,01	211517	B52	Šaľa a Galanta	592,41	0,0521536	11031
5206	Nitra	11359,01	211517	B63	Zlaté Moravce	2530,48	0,2227733	47120
5207	Nové Zámky	8416,83	153466	B20	Levice	132,60	0,0157547	2418
5207	Nové Zámky	8416,83	153466	B31	Nové Zámky	6059,25	0,7198972	110480
5207	Nové Zámky	8416,83	153466	B58	Štúrovo	2224,97	0,2643481	40568
5208	Senica	7258,74	147119	B5	Bratislava	1268,22	0,1747158	25704
5208	Senica	7258,74	147119	B26	Mýjava	1855,21	0,2555883	37601
5208	Senica	7258,74	147119	B42	Senica	2050,74	0,28252	41564
5208	Senica	7258,74	147119	B43	Skalica a Holíč	2084,57	0,2871812	42250
5209	Topoľčany	8159,71	160767	B1	Bánovce n. B.	2005,09	0,2457301	39505
5208	Topoľčany	8159,71	160767	B13	Hlohovec	142,03	0,0174065	2798
5205	Topoľčany	8159,71	160767	B54	Topoľčany	6012,60	0,7368634	118463

b) rok 2001

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ y_z		CIELOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY w		ODHAD y_x
Kód okresu	Názov okresu	Rozloha triedy (ha)	PO	Kód FMR 91-B	Názov FMR 91-B	Rozloha prieniku (ha)	Váha	w*PO
100	Senica	3280,43	60891	B5	Bratislava	567,13	0,172884	10527
106	Senica	3280,43	60891	B26	Mýjava	255,80	0,077977	4748
107	Senica	3280,43	60891	B42	Senica	2121,32	0,646661	39376
108	Senica	3280,43	60891	B43	Skalica a Holíč	336,17	0,102479	6240
201	Skalica	1779,80	46791	B43	Skalica a Holíč	1779,80	1,000000	46791
201	Trnava	5019,51	127125	B58	Trnava	5019,51	1,000000	127125
202	Bánovce n. Bebr.	2004,50	38640	B1	Bánovce n. B.	2004,50	1,000000	38640
202	Iľava	1505,48	62042	B9	Dubnica n. Váhom	1505,48	1,000000	62042
202	Mýjava	1926,79	29243	B26	Mýjava	1620,78	0,841184	24599
203	Mýjava	1926,79	29243	B30	Nové Mesto n.V.	306,00	0,158816	4644

b) následne sa vypočítala tzv. vážená dĺžka ciest každej zdrojovej zóny, t. j. každému typu cesty sa pridelila vopred prisúdená váha:

$$\text{vážená } D_z = \sum_t D_t^z w_t$$

c) rovnako sa určila vážená dĺžka ciest každej zóny prieniku medzi zdrojovou a cieľovou zónou:

$$\text{vážená } D_{zc} = \sum_t D_t^{zc} w_t$$

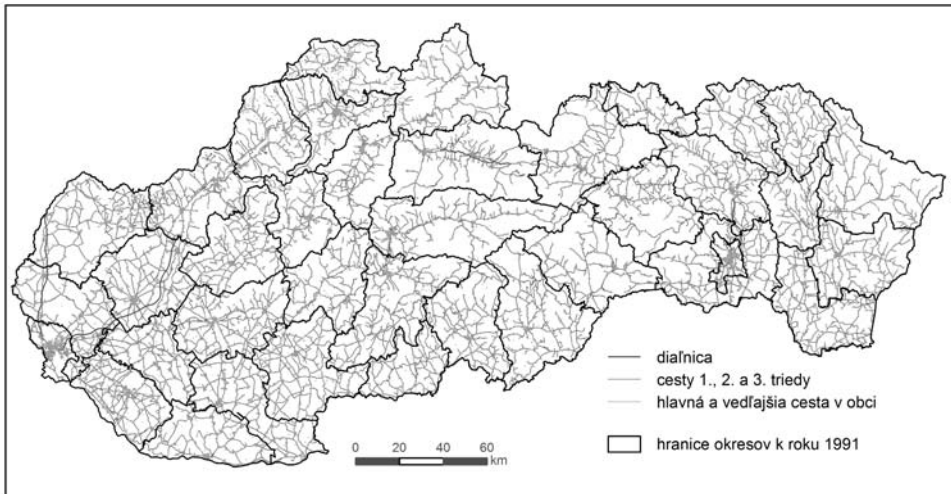
d) vynásobením hodnoty premennej v príslušnej zdrojovej zóne y_z a hodnoty váhy w_{zc} (t.j. podielom váženej dĺžky danej zóny prieniku D_{zc} na váženej dĺžke príslušnej zóny D_z) sa vypočítal odhad hodnoty premennej v danej zóne prieniku:

$$\hat{y}_{zc} = \frac{\text{vážená } D_{zc}}{\text{vážená } D_z} \cdot y_z$$

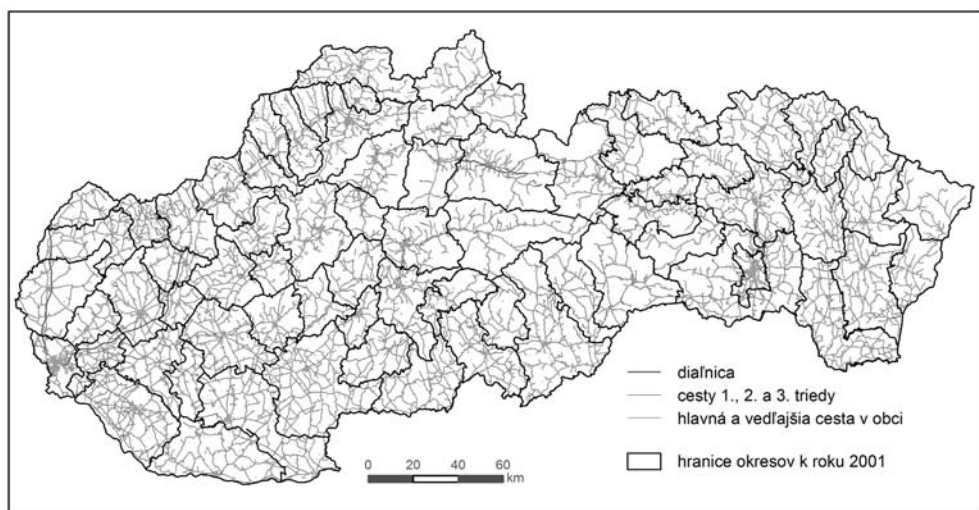
e) odhad hodnoty premennej za jednotlivé cieľové zóny sa následne vypočítal ako:

$$\hat{y}_c = \sum_z \hat{y}_{zc} = \sum_z y_z \frac{\text{vážená } D_{zc}}{\text{vážená } D_z} = \sum_z y_z \frac{\sum_t D_t^{zc} w_t}{\sum_t D_t^z w_t}$$

Princíp aplikácie metódy dĺžky a typov ciest na odhad hodnôt premenných demonštrujeme na príklade vzorových tabuliek tab. 7a,b a tab.8 a,b, ktoré opisujú odhad počtu obyvateľov za vybrané zóny prieniku.



Obr. 11. Prekryv okresov 1991 a vybraných cestných úsekov z databázy SVM50



Obr. 12. Prekryv okresov 2001 a vybraných cestných úsekov z databázy SVM50

2.3.4. Geografické konverzné tabuľky

Ako vidieť z predchádzajúcich príkladov, mimo prostredia GIS sa dajú potrebné výpočty a samotná reagregácia vybraných premenných zo zón prieniku do jednotlivých FMR 91-B uskutočniť prostredníctvom geografických konverzných tabuliek. Prístup GKT sme však uplatnili aj samostatne s cieľom potvrdiť alebo vyvrátiť predpoklad, že ak sa ako váhy použijú informácie, ktoré sú v úzkom vzťahu s odhadovanými premennými, potom výsledky procesu odhadu dát za nekompatibilné zonálne systémy by mali byť oveľa presnejšie. Pri uplatnení tohto prístupu sme využili vnorenú hierarchiu zdrojových a cieľových geografických jednotiek. Keďže oba systémy jednotiek pozostávajú z obcí a z prostredia GIS vieme, ktoré obce (resp. časti obce) ležia v prieniku okresov a FMR 91-B, mohli sme známe hodnoty za obce použiť na určenie váhy potrebnej na odhad vybraných premenných.

Vynásobením hodnôt každého záznamu zdrojových geografických jednotiek – okresov príslušnou váhou (počtom cenзовých domácností, počtom ekonomicky aktívnych obyvateľov a rozlohou zastavanej plochy) sme získali odhad premenných za skupinu prienikových zón. Keďže prispôsobenie dvoch párov geografických jednotiek nie je dokonalé, konverzné tabuľky majú zvyčajne pre každú zdrojovú jednotku viac než jeden záznam s hodnotou váhy, ktorá ukazuje, ako je zdrojová jednotka rozdelená medzi cieľovú jednotku. Ako príklad uvádzame tab. 10a za prvú testovanú situáciu a tab. 10b za druhú testovanú situáciu, ktoré načrtávajú odhad premennej počet obyvateľov s použitím počtu cenзовých domácností ako váženého kritéria.

Tab. 7. Vzorová tabuľka odhadu počtu obyvateľov metódou dĺžky cestnej siete za vybrané zóny priteniku:

a) rok 1991

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ y_z	CIELOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY w		ODHAD y_{zc}	
Kód okresu	Názov okresu	Dĺžka ciest (m)	PO	Kód FMR 91-B	Názov FMR 91-B	Dĺžka ciest v priteniku (m)	Váha w	w*PO
5210	Trenčín	934761,50	179499	B30	Nové Mesto n.V.	387205,24	0,414229	74354
5210	Trenčín	934761,50	179499	B32	Piešťany	47973,23	0,051321	9212
5210	Trenčín	934761,50	179499	B57	Trenčín	499583,03	0,534450	95933
5211	Tinava	962650,70	233882	B13	Hlohovec	180949,42	0,187970	43963
5211	Tinava	962650,70	233882	B32	Piešťany	267980,48	0,278378	65108
5211	Tinava	962650,70	233882	B58	Tinava	513720,80	0,533652	124812
5301	Banská Bystrica	896297,24	177654	B2	Banská Bystrica	457570,28	0,510512	90694
5301	Banská Bystrica	896297,24	177654	B6	Brezno	438726,96	0,489488	86960
5302	Čadca	500036,29	122687	B7	Čadca	383866,47	0,767677	94184
5302	Čadca	500036,29	122687	B66	Žilina	116169,82	0,232323	28503

b) rok 2001

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ y_z	CIELOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY w		ODHAD y_{zc}	
Kód okresu	Názov okresu	Dĺžka ciest (m)	PO	Kód FMR 91-B	Názov FMR 91-B	Dĺžka ciest v priteniku (m)	Váha	w*PO
304	Nové Mesto n.V.	391200,54	63530	B30	Nové Mesto n.V.	334519,05	0,8551089	54325
304	Nové Mesto n.V.	391200,54	63530	B32	Piešťany	56681,49	0,1448911	9205
305	Partizánske	203848,14	48005	B50	Topoľčany	203848,14	1	48005
306	Považská Bystrica	269905,78	65150	B34	Považská Bystrica	269905,75	1	65150
307	Prievidza	576713,63	140444	B35	Prievidza	576713,63	1	140444
308	Púchov	229301,89	45761	B37	Púchov	229301,89	1	45761
309	Trenčín	510345,08	112767	B57	Trenčín	510345,08	1	112767
401	Komárno	684017,44	108556	B16	Komárno	580164,44	0,848172	92074
401	Komárno	684017,44	108556	B31	Nové Zámky	103852,99	0,151828	16482
402	Levice	986666,30	120021	B20	Levice	792437,85	0,8031468	96394
402	Levice	986666,30	120021	B51	Šahy	194228,4442	0,1968532	23627

Tab. 8. Vzorová tabuľka počtu obyvateľov metódou typov cestnej siete za vybrané zóny prieniku

a) rok 1991

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ _{y_z}	CIELOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY _w		ODHAD _{y_z}
Kód okresu	Názov okresu	Vážená dĺžka ciest (m)	Kód FMR 91-B	Názov FMR 91-B	Vážená dĺžka ciest v prieniku (m)	Váha _w	w*PO
5303	Dolný Kubín	303307,50	B8	Dolný Kubín	111003,27	0,365976	43684
5303	Dolný Kubín	303307,50	B27	Námestovo	111175,46	0,366544	43751
5303	Dolný Kubín	303307,50	B59	Tvrdošín a Nižná	81128,77	0,267480	31927
5304	Liptovský Mikuláš	381155,52	B21	Liptovský Mikuláš	218067,65	0,572122	75447
5304	Liptovský Mikuláš	381155,52	B41	Ružomberok	155414,38	0,407745	53770
5304	Liptovský Mikuláš	381155,52	B59	Tvrdošín a Nižná	7673,49	0,020132	2655
5305	Lučenec	282694,98	B22	Lučenec	276385,31	0,977680	93847
5305	Lučenec	282694,98	B64	Zvolen	6309,67	0,022320	2142
5306	Martín	290479,22	B23	Martín	285112,01	0,981523	110898
5306	Martín	290479,22	B65	Žiar nad Hronom	5367,21	0,018477	2088

b) rok 2001

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ _{y_z}	CIELOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY _w		ODHAD _{y_z}
Kód okresu	Názov okresu	Vážená dĺžka ciest (m)	Kód FMR 91-B	Názov FMR 91-B	Vážená dĺžka ciest v prieniku (m)	Váha	w*PO
403	Nitra	373641,14	B13	Hlohovec	6240,64	0,016702	2731
403	Nitra	373641,14	B28	Nitra	367400,50	0,983298	160809
404	Nové Zámky	470592,64	B20	Levice	8796,14	0,018692	2796
404	Nové Zámky	470592,64	B31	Nové Zámky	328792,75	0,698678	104518
405	Nové Zámky	470592,64	B53	Štúrovo	133003,75	0,282630	42280
405	Šala	146767,62	B52	Šala a Galanta	146767,62	1	54000
406	Topoľčany	222810,65	B13	Hlohovec	9124,95	0,040954	3034
406	Topoľčany	222810,65	B54	Topoľčany	213685,70	0,959046	71055
407	Zlaté Moravce	149353,18	B28	Nitra	6774,02	0,045356	1979
407	Zlaté Moravce	149353,18	B29	Nová Baňa a Žarnovica	5598,58	0,037486	1635
407	Zlaté Moravce	149353,18	B63	Zlaté Moravce	136980,58	0,917159	40008
501	Bytča	83038,64	B66	Zlínka	83038,64	1	30788

Takto získané vážené hodnoty sme následne agregovali a pridelili každej cieľovej zóne – FMR 91-B, čím sme dospeli k celkovému odhadu vybraných premenných za cieľové geografické jednotky.

V zmysle triedenia GKT nie sú tieto tabuľky ani vyhľadávacie a ani hierarchické. Stupeň hierarchie je pre každú tabuľku rovnaký, a to 18,4 % za rok 1991 a 47,2 % za rok 2001. Stupeň prispôsobenia sa líši v závislosti od použitej váhy a udáva ho Tab. 9.

Tab. 9. Porovnanie stupňa prispôsobenia geografických konverzných tabuliek

Váha	Stupeň prispôsobenia (%)	
	rok 1991	rok 2001
Počet cenzových domácností	77,3	94,0
Počet ekonomicky akt.obyv.	77,9	94,2
Rozloha zastavanej plochy	75,7	92,6

Tab. 10. Vzorová tabuľka odhadu počtu obyvateľov metódou GKT pri použití počtu cenzových domácností ako váženého kritéria – a) rok 1991

Kód okresu	Názov okresu	CD	PREMENNÁ y _t	PO	CIEĽOVÉ GEOGRAFICKÉ JEDNOTKY 91-B	Názov FMR 91-B	CD	VÝPOČET VÁHY w	Váha w	ODHAD y _{pc} w*PO
5312	Žiar nad Hronom	33274	94052	94052	B3	Banská Štiavnica	6173	6173	0,18552	17449
5312	Žiar nad Hronom	33274	94052	94052	B29	Nová Baňa a Žarnovica	11081	11081	0,333023	31321
5312	Žiar nad Hronom	33274	94052	94052	B65	Žiar nad Hronom	16242	16020	0,481457	45282
5313	Žilina	60794	181864	181864	B66	Žilina	71470	60794	1	181864
5401	Bardejov	23115	112067	112067	B4	Bardejov	20206	20206	0,874151	69095
5401	Bardejov	23115	112067	112067	B12	Giraltovce	2909	2909	0,125849	9947
5402	Humenné	36094	95989	95989	B15	Humenné	20869	20000	0,554109	62097
5402	Humenné	36094	95989	95989	B24	Medzilaborce	3809	3809	0,10553	11826
5402	Humenné	36094	95989	95989	B44	Snina	12285	12285	0,340361	38143
5403	Košice-mesto	85168	112986	112986	B17	Košice	118654	85168	1	235160
5404	Košice-vidiek	32392	99292	99292	B17	Košice	118654	32392	1	99292
5405	Michalovce	37401	111376	111376	B25	Michalovce	31476	31476	0,841582	93732
5405	Michalovce	37401	111376	111376	B45	Sobrance	5925	5925	0,158418	17644
5406	Poprad	47981	154129	154129	B33	Poprad	46273	46102	0,960839	148093
5406	Poprad	47981	154129	154129	B47	Spisská Stará Ves	1879	1879	0,039161	6036
5407	Prešov	63105	200245	200245	B17	Košice	118654	286	0,004532	908
5407	Prešov	63105	200245	200245	B35	Prešov	62834	62689	0,993408	198925
5407	Prešov	63105	200245	200245	B48	Stará Ľubovňa	14014	130	0,00206	413
5408	Rožňava	30500	86311	86311	B38	Revúca	10016	9337	0,306131	26422
5408	Rožňava	30500	86311	86311	B40	Rožňava	21163	21163	0,693869	59889
5409	Spisská Nová Ves	45770	145481	145481	B11	Gelnica a Prakovce	7850	7850	0,17151	24951
5409	Spisská Nová Ves	45770	145481	145481	B33	Poprad	46273	171	0,003736	544
5409	Spisská Nová Ves	45770	145481	145481	B35	Prešov	62834	145	0,003168	461
5409	Spisská Nová Ves	45770	145481	145481	B46	Spisská Nová Ves	37604	37604	0,821586	119525
5410	Stará Ľubovňa	13884	46513	46513	B48	Stará Ľubovňa	14014	13884	1	46513
5411	Svidník	13399	44289	44289	B49	Stropkov	5886	5787	0,431898	19128
5411	Svidník	13399	44289	44289	B50	Svidník	7612	7612	0,568102	25161
5412	Trebišov	39155	118524	118524	B17	Košice	118654	808	0,020636	2446
5412	Trebišov	39155	118524	118524	B18	Kraľovský Chlmec	13478	13476	0,344222	40799
5412	Trebišov	39155	118524	118524	B56	Trebišov	18973	18973	0,484561	57432
5412	Trebišov	39155	118524	118524	B60	Veľké Kapušany	5896	5896	0,150581	17847
5413	Vranov nad Topľou	22058	73681	73681	B15	Humenné	20869	869	0,039396	2903
5413	Vranov nad Topľou	22058	73681	73681	B49	Stropkov	5886	99	0,004488	331
5413	Vranov nad Topľou	22058	73681	73681	B62	Vranov nad Topľou	21090	21090	0,956116	70448

ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ y _j		CIELOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY w		ODHAD y _j	
Kód okresu	Názov okresu	CD	PO	Kód FMR 91-B	Názov FMR 91-B	CD	CD v priteniku	Váha w	w*PO
509	Turčianske Teplice	6553	16866	B23	Martin	44852	6322	0,964749	16271,45613
509	Turčianske Teplice	6553	16866	B65	Žiar nad Hronom	19150	231	0,035251	594,543873
510	Tvrdošín	11410	35062	B59	Tvrdošín a Nižná	11891	11410	1	35062
511	Žilina	57602	156361	B66	Žilina	81227	57602	1	156361
601	Banská Bystrica	45139	111984	B2	Banská Bystrica	45139	45139	1	111984
602	Banská Štiavnica	7079	17151	B3	Banská Štiavnica	7079	7079	1	17151
603	Brezno	25912	65909	B6	Brezno	25912	25912	1	65909
604	Dereva	15333	33514	B64	Zvolen	44836	15333	1	33514
605	Krupina	8994	22885	B19	Krupina	6324	6324	0,703135	16091,25417
605	Krupina	8994	22885	B51	Šahy	10350	1878	0,208806	4778,522348
605	Krupina	8994	22885	B64	Zvolen	44836	792	0,088059	2015,223482
606	Lučenec	30432	72837	B22	Lučenec	39490	30432	1	72837
607	Poltár	9481	23666	B22	Lučenec	39490	8868	0,935344	22135,85993
607	Poltár	9481	23666	B39	Rimavská Sobota	23678	613	0,064656	1530,14007
608	Revúca	15635	40918	B38	Revúca	10986	10934	0,699328	28615,12069
608	Revúca	15635	40918	B55	Tornala	7769	4701	0,300672	12302,87931
609	Rimavská Sobota	32709	83124	B14	Hnúšťa	6757	6757	0,206579	17171,69183
609	Rimavská Sobota	32709	83124	B22	Lučenec	39490	190	0,005809	482,8505916
609	Rimavská Sobota	32709	83124	B38	Revúca	10986	52	0,00159	132,148583
609	Rimavská Sobota	32709	83124	B39	Rimavská Sobota	23678	23065	0,705158	58615,5205
609	Rimavská Sobota	32709	83124	B55	Tornala	7769	2645	0,080865	6721,788499
610	Veľký Krτίš	19423	46741	B51	Šahy	10350	262	0,013489	630,4969366
610	Veľký Krτίš	19423	46741	B61	Veľký Krτίš	19161	19161	0,986511	46110,50306
611	Zvolen	28711	67633	B64	Zvolen	44836	44836	1	67633
612	Žarnovica	10841	27634	B29	Nová Baňa a Žarnovica	12044	12044	1	27634
613	Žiar nad Hronom	19706	48125	B29	Nová Baňa a Žarnovica	12044	12044	0,039937	1921,971734
613	Žiar nad Hronom	19706	48125	B65	Žiar nad Hronom	19150	19150	0,960063	46203,02827
701	Bardejov	24373	75793	B4	Bardejov	23501	23501	0,964223	73081,33151
701	Bardejov	24373	75793	B12	Giraltovce	3566	3566	0,035777	2711,668486
702	Humenné	22958	64845	B15	Humenné	23813	13813	1	64845
703	Kežmarok	19346	63231	B33	Poprad	55292	55292	0,902202	57047,13501
703	Kežmarok	19346	63231	B47	Spisská Stará Ves	1892	1892	0,097798	6183,864985

b) rok 2001 – pokračovanie

Kód okresu	ZDROJOVÉ GEOGRAFICKÉ JEDNOTKY		PREMENNÁ y_x		CIELOVÉ GEOGRAFICKÉ JEDNOTKY		VÝPOČET VÁHY w		ODHAD y_w w*PO
	Názov okresu	CD	PO	Kód FMR 91-B	Názov FMR 91-B	CD	CD v prieniku	Váha w	
704	Levoča	10861	31880	B35	Prešov	75757	150	0,013811	440,2909493
704	Levoča	10861	31880	B46	Spisská Nová Ves	45085	10711	0,986189	31439,70905
705	Medzilaborce	4944	12668	B15	Humenné	23813	855	0,172937	2190,764563
705	Medzilaborce	4944	12668	B24	Medzilaborce	4089	4089	0,827063	10477,23544
706	Poprad	37640	10434	B33	Poprad	55292	37640	1	104348
707	Prešov	57733	161782	B17	Košice	133996	284	0,004919	795,8375279
707	Prešov	57733	161782	B35	Prešov	75757	57449	0,995081	160986,1625
708	Sabinov	18283	54067	B35	Prešov	75757	18158	0,993163	53697,3465
708	Sabinov	18283	54067	B48	Stará Ľubovňa	16159	125	0,006837	369,6535033
709	Snina	14272	39633	B44	Snina	14272	14272	1	39633
710	Stará Ľubovňa	16034	50684	B48	Stará Ľubovňa	16159	16034	1	50684
711	Stropkov	6998	21027	B49	Stropkov	6930	6797	0,971278	20423,05216
711	Stropkov	6998	21027	B50	Svidník	8907	201	0,028722	603,9478422
712	Svidník	11400	33506	B12	Giraltovce	3566	2694	0,236316	7917,996842
712	Svidník	11400	33506	B50	Svidník	8907	8706	0,763684	25588,00316
713	Vranov nad Topľou	24572	76504	B49	Stropkov	6930	133	0,005413	414,0905095
713	Vranov nad Topľou	24572	76504	B62	Vranov nad Topľou	24439	24439	0,994587	76089,90949
800	Košice (I-IV)	94925	236093	B17	Košice	133996	94925	1	236093
801	Gelnica	10647	3084	B11	Gelnice a Prakovce	8266	8266	0,776369	23943,99418
801	Gelnica	10647	3084	B46	Spisská Nová Ves	45085	2381	0,223631	6897,005823
806	Košice-okolie	37923	106999	B17	Košice	133996	37923	1	106999
807	Michalovce	40711	109121	B25	Michalovce	36221	33774	0,829604	90527,19545
807	Michalovce	40711	109121	B60	Veľké Kapušany	6937	6937	0,170396	18593,80455
808	Rožňava	23475	61887	B40	Rožňava	23052	23052	0,981981	60771,84767
808	Rožňava	23475	61887	B55	Tornaľa	7769	423	0,018019	1115,152332
809	Sobrance	9434	23776	B25	Michalovce	36221	2447	0,259381	6167,041764
809	Sobrance	9434	23776	B45	Sobrance	6987	6987	0,740619	17608,95824
810	Spisská Nová Ves	32191	93516	B33	Poprad	55292	198	0,006151	575,1970427
810	Spisská Nová Ves	32191	93516	B46	Spisská Nová Ves	45085	31993	0,993849	92940,80296
811	Trebišov	37407	103779	B17	Košice	133996	864	0,023097	2397,012752
811	Trebišov	37407	103779	B18	Kráľovský Chlmec	14574	14574	0,389606	40432,94426
811	Trebišov	37407	103779	B56	Trebišov	21969	21969	0,587296	60949,04299

3. VÝSLEDKY A DISKUSIA

V nasledujúcich podkapitolách sa pokúsime zhodnotiť presnosť a spoľahlivosť jednotlivých metód areálovej transformácie vo vedomí, že transformácia geografických dát z jedného územného systému do druhého, odlišného systému územných jednotiek je proces, ktorý podlieha chybám a ani jedna metóda neposkytuje úplne presné výsledky, ale vždy ide len o odhad hodnôt danej premennej. Preto je dôležité celkové zhodnotenie výsledkov jednotlivých testovaných metód. Okrem „hrubého“ zhodnotenia celkovej úspešnosti jednotlivých metód a porovnania presnosti ich odhadov sa navyše pokúsime odpovedať na otázky:

- Dokážu metódy, ktoré využívajú pomocné údaje, významne redukovať chyby spojené s metódou priestorového váženía, tak ako tomu bolo v predchádzajúcich výskumoch?
- Od čoho závisí presnosť jednotlivých metód areálovej transformácie geografických dát?

3.1. Spôsob hodnotenia výsledkov metód areálovej transformácie

Vybrané metódy areálovej transformácie geografických dát a ich výsledky za rôzne premenné sa budú hodnotiť a porovnávať jednak za obe testované situácie navzájom, t. j. z okresov 1991 do FMR 91-B na jednej strane a z okresov 2001 do FMR 91-B na strane druhej a za každú situáciu osobitne, a to nielen vo všeobecnej rovine, ale aj posúdením jednotlivých individuálnych prípadov, t. j. odhadov za každú cieľovú geografickú jednotku zvlášť v rámci každej metódy a premennej. Na takého hodnotenie využijeme už spomínanú hierarchiu geografických jednotiek. Hodnoty jednotlivých premenných možno totiž za cieľové zóny (FMR 91-B) získať agregáciou dát za obce, čo nám umožní odhadnuté hodnoty transformovaných premenných porovnať so skutočnými hodnotami a určiť presnosť a spoľahlivosť odhadov jednotlivých metód.

Pod pojmom presnosť máme na mysli posúdenie a porovnanie odhadov konkrétnych premenných za jednotlivé metódy, termín spoľahlivosť naopak vzťahujeme na posúdenie úspešnosti jednotlivých metód celkovo za všetky premenné. Posudzovať presnosť a spoľah-

livosť odhadov testovaných metód len na základe porovnania vybraných deskriptívnych charakteristík (stredná hodnota, medián, smerodajná odchýlka, minimum, maximum, rozpätie, súčet) skutočných a odhadnutých hodnôt každej premennej, tak ako to navrhli a použili niektorí autori (Xie 1995, Mrozinski a Cromley 1999, Reibel a Buffalino 2005, Reibel a Agrawal 2007) nie je signifikantné a môže poskytnúť len hrubé zhodnotenie výsledkov jednotlivých použitých metód. Hlavnými numerickými ukazovateľmi súhrnných výsledkov (t. j. výsledkov za celý súbor geografických jednotiek) presnosti a spoľahlivosti každej metódy areálovej transformácie budú teda tzv. *miery globálnej vhodnosti*. V literatúre zaoberajúcej sa hodnotením presnosti jednotlivých metód používaných na odhad hodnôt premennej z jedného systému geografických jednotiek do druhého nekompatibilného systému geografických jednotiek sa zvyknú ako miery globálnej vhodnosti používať *stredná kvadratická chyba* a tzv. prispôbená *stredná kvadratická chyba* (Fisher a Langford 1995), vypočítané podľa rovníc:

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

Tieto miery však tvrdo postihujú rozdiely medzi skutočnou y_i a odhadnutou hodnotou \hat{y}_i vybranej premennej. Na začiatok v snahe zachytiť podstatné rozdiely v presnosti odhadov jednotlivých metód areálovej transformácie geografických dát sme sa rozhodli tieto miery nepoužiť a nahradiť ich podobnými, ale o niečo menej prísnyimi štatistikami – priemernou absolútnou, priemernou relatívnou a priemernou algebraickou chybou (S_1 , S_2 a S_3):

$$S_1 = \frac{1}{n} \sum_{c=1}^n |\hat{y}_c - y_c|$$

$$S_2 = \frac{1}{n} \sum_{c=1}^n \frac{|\hat{y}_c - y_c|}{y_c} \quad 100 \%$$

$$S_3 = \frac{1}{n} \sum_{c=1}^n \frac{\hat{y}_c - y_c}{y_c} \quad 100 \%$$

Zatiaľ čo S_1 meria chyby odhadu v jednotkách merania premennej y , S_2 je bezrozmerná štatistika, ktorá chyby uvažuje vo vzťahu k skutočnej hodnote premennej. Udáva relatívny počet nesprávne pridelených hodnôt premennej y . S_3 je mierou vychýlenia odhadu. Kladné hodnoty premennej naznačujú, že daná metóda poskytuje v priemere nadhodnotenú odhadu premennej y , záporné naopak signalizujú tendenciu k podhodnoteným odhadom (Bezák a Holická 1995).

Deskriptívne charakteristiky a miery S_1 , S_2 a S_3 sú teda štatistiky, ktoré posudzujú odhady premennej súhrnne v rámci celého súboru cieľových geografických jednotiek (n). Pri-

merné hodnoty za danú metódu však zastierajú rozdiely medzi chybami individuálnych predpovedí. Preto spomínané štatistiky rozšírime o ďalší indikátor hodnotenia úspešnosti testovaných metód, a to o relatívnu chybu vyjadrenú v percentách vypočítanú ako:

$$S_{4c} = \frac{\hat{y}_c - y_c}{y_c} \cdot 100 \%$$

ktorú použijeme na podrobnejšie zhodnotenie chýb prostredníctvom priestorovej vizualizácie. Hodnoty relatívnych chýb vyjadrených v percentách sa pre každú odhadovanú premennú a aplikovanú metódu znázornia v podobe kartogramov, ktoré zobrazujú priestorové rozmiestnenie chýb podľa jednotlivých cieľových geografických jednotiek, čo nám umožňuje ľahšie určiť, ako sú jednotlivé metódy hodnoverné.

Indikátor S_{4c} však okrem zhodnotenia chýb za každú konkrétnu cieľovú geografickú jednotku využijeme aj na znázornenie početnosti jednotlivých percentuálnych chýb v rámci každej premennej v jednotlivých metódach prostredníctvom histogramov. Tieto histogramy využijeme okrem iného na celkové zhodnotenie úspešnosti jednotlivých metód tak, že si podrobne všimneme podiel presných odhadov s relatívnou chybou od -5 do 5 % a podiel extrémnych odhadov s relatívnou chybou väčšou ako $\pm 50 \%$.

3.2. Porovnanie výsledkov vybraných metód

3.2.1. Miery globálnej vhodnosti a histogramy

Deskriptívne charakteristiky každej premennej podľa jednotlivých metód za každú testovanú situáciu zvlášť obsahujú tab. 11a,b. Porovnaním súčtov a stredných hodnôt pôvodnej skupiny dát a výsledných odhadov sa pri všetkých testovaných metódach potvrdilo zachovanie tzv. pyknofylaktickej vlastnosti, čo znamená, že v procese areálovej transformácie sa žiadne dáta nestratili, ale naopak, zachovali si svoj rozsah. Odchýlka hodnôt ostatných deskriptívnych štatistík odhadovaných premenných oproti skutočným hodnotám je pri premenných PO a POPRED v oboch testovaných situáciách väčšia pri metódach priestorového váženia, dĺžky a typov ciest a najmenšia pri metódach GKT. Naopak, pri premennej EAOPOL sa ako najmenej presné javia metódy GKT, obzvlášť metóda vážiaca podľa EAO, a ako najpresnejšie dazymetrická metóda vážiaca na základe poľnohospodárskych budov, metóda priestorového váženia a dĺžky ciest.

Na základe výsledkov mier charakterizujúcich rozpätie jednotlivých hodnôt premenných v danom súbore cieľových geografických jednotiek možno teda skonštatovať, že za obe testované situácie dosiahli najhoršie výsledky spomedzi všetkých aplikovaných metód v priemere za všetky premenné metódy priestorového váženia a dĺžky ciest. Ako najúspešnejšie sa naopak javia metódy GKT, ktoré vážia odhadované premenné prostredníctvom EAO a CD. Výsledky jednotlivých deskriptívnych štatistík zároveň napovedajú, že spomenuté zistenia úplne neplatia pre všetky odhadované dáta. Tomuto celkovému hodnoteniu sa čiastočne vymyká premenná EAOPOL. Okrem toho jednotlivé deskriptívne štatistiky napovedajú, že pri odhadoch relatívne veľkých hodnôt dáva pri všetkých premenných a v oboch testovaných situáciách najlepšie výsledky metóda GKT vážiaca podľa EAO,

najhoršie sa javia metódy priestorového váženia a dĺžky ciest. Pri premennej EAOPOL sa okrem toho radí k úspešným metódam aj metóda GKT vážiaca na základe rozlohy poľnohospodárskych budov. Pri odhadoch najmenších hodnôt premenných už nie je situácia taká jednoznačná. Medzi najmenej presné metódy areálovej transformácie geografických dát patrí dazymetrická metóda, v prvej testovanej situácii aj metóda typov ciest a v druhej testovanej situácii aj dazymetrická metóda vážiaca na základe rodinných a bytových domov a metóda GKT vážiaca podľa ZP. Výnimku predstavuje opäť premenná EAOPOL, pri ktorej metóda GKT vážiaca podľa ZP spolu s dazymetrickou metódou využívajúcou vrstvu poľnohospodárskych budov patrili medzi najúspešnejšie metódy a metóda GKT vážiaca podľa EAO spolu s metódou priestorového váženia medzi najmenej presné metódy pri odhadoch nízkych hodnôt premennej EAOPOL v roku 2001. GKT vážiace podľa EAO a CD patrili v prípade zvyšných premenných k najpresnejším metódam odhadu, až na premennú POPRED v roku 2001, kde úspešnejšie vystupovali metódy dĺžky a typov ciest.

Treba zdôrazniť, že v žiadnom prípade sa nemožno opierať len o takéto hrubé zhodnotenie výsledkov jednotlivých testovaných metód na základe deskriptívnych štatistík, ktoré nemôže obsiahnuť a toľko vysvetliť celkovú úspešnosť odhadov metód areálovej transformácie geografických dát. Na komplexnejšie posúdenia je preto potrebné vziať do úvahy ďalšie indikátory hodnotenia – miery globálnej vhodnosti: S_1 , S_2 a S_3 (tab.12 a,b).

Aj na základe hodnôt mier globálnej vhodnosti (čím nižšia je hodnota miery, tým menšia chyba spojená s určitou metódou) poskytovali v oboch testovaných situáciách v priemere za odhadované premenné najlepšie výsledky metódy GKT (hlavne tie, ktoré odhadovanú premennú vážili podľa EAO a CD). K najmenej presným patrili naopak metódy priestorového váženia a dĺžky ciest. Pri hodnotení úspešnosti zvyšných metód možno skonštatovať, že podľa týchto mier lepšie výsledky zaznamenala dazymetrická metóda vážiaca podľa vrstvy rodinných a bytových budov ako klasická dazymetrická metóda a metóda typov ciest. Priemerné hodnotenie všetkých testovaných metód však opäť neplatí pre všetky odhadované premenné. Uvádzaným tvrdeniam o najpresnejších a najmenej presných metódach sa v prvej testovanej situácii (rok 1991) vymyká premenná EAOPOL, keď sa ako najúspešnejšia javí metóda GKT vážiaca podľa ZP a klasická binárna dazymetrická metóda. Hodnotenie celkovej kvality testovaných metód na základe výsledkov, ktoré poskytli miery S_1 a S_2 však do istej miery sťažuje aj fakt, že nie pri každej premennej bolo poradie úspešnosti jednotlivých metód vytvorené na základe chýb S_1 a S_2 totožné. V podstate sa výsledky za jednotlivé metódy areálovej transformácie geografických dát pri jednotlivých premenných až na drobné výnimky zhodovali, len pri premennej EAOPOL v druhej testovanej situácii bolo poradie jednotlivých metód vytvorené na základe hodnôt mier S_1 a S_2 , odlišné (tab. 12b). Nejde však o podstatné rozdiely v kvalite jednotlivých metód areálovej transformácie dát, ide skôr len o drobné výmeny poradia úspešnosti testovaných metód.

Z tab. 12ab ďalej vyplýva, že zo všetkých premenných bola za obe testované situácie najmenej presná metóda priestorového váženia pre premennú POPRED, keď v priemere za všetky geografické jednotky nesprávne pridelila v roku 1991 2 384 obyvateľov, čo predstavovalo 21,3 % počtu všetkých POPRED, a za rok 2001 v priemere chybné odhadla 565 obyvateľov, resp. 8,7 %. Porovnaním jednotlivých metód sa preukázalo, že najmenšie chyby dosiahla zo všetkých premenných v oboch testovaných situáciách premenná PO,

a to prostredníctvom metódy GKT vážiacej podľa EAO: v roku 1991 nesprávne prideliла 584 osôb, t. j. 1,3 % celkového počtu obyvateľov, čo predstavovalo oproti najhoršej metóde – priestorovému váženiu – takmer 94-percentné zlepšenie v prípade absolútneho počtu a 93-percentné zlepšenie, ak ide o podiel nesprávne prideleného počtu obyvateľov (tab. 12a). V roku 2001 to bolo 171 nesprávne pridelených obyvateľov (0,46 %) a 93,7-percentné zlepšenie oproti metóde priestorového váženia v prípade absolútneho počtu, aj pokiaľ ide o podiel nesprávne prideleného počtu obyvateľov (tab. 12b). 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 premennej v jednotlivých geografických jednotkách, čím sú spôsobené aj odchýlky v presnosti jednotlivých metód areálovej transformácie.

Za povšimnutie stojí aj fakt, že v rámci jednej premennej možno v istých prípadoch pozorovať pomerne značné rozdiely medzi veľkosťou chýb jednotlivých mier globálnej vhodnosti v závislosti od použitej metódy. Ide hlavne o premenné PO a POPRED. V prípade premennej EAOPOL sa naopak výsledky jednotlivých metód javia ako pomerne vyvážené a táto premenná v priemere nevykazuje až také výrazné zlepšenie oproti metóde priestorového váženia ako zvyšné premenné.

Tab. 11. Deskriptívne štatistiky za jednotlivé premenné a metódy:

a) pre rok 1991

Štatistiky	Počet obyvateľov							
	skutočné hodnoty	PV	GKT-EAO	GKT-CD	GKT-ZP	DM	DC	TC
Stredná hodota	79914	79914	79914	79914	79914	79914	79914	79914
Medián	59336	54008	58817	58909	54112	56393	54544	54955
Smer.odchýlka	91272	92534	91358	91178	91677	91557	91545	91048
Minimum	6013	9041	5699	6036	9508	11959	11555	12807
Maximum	652590	686402	652087	651939	665371	667471	672214	664032
Rozpätie	646577	677361	646388	645903	655863	655513	660659	651225
Súčet	5274335	5274335	5274335	5274335	5274335	5274335	5274335	5274335

Štatistiky	Počet obyvateľov v predproduktívnom veku							
	skutočné hodnoty	PV	GKT-EAO	GKT-CD	GKT-ZP	DM	DC	TC
Stredná hodota	19909	19909	19909	19909	19909	19909	19909	19909
Medián	14474	14000	14336	14371	14139	14521	13917	14131
Smer.odchýlka	22130	22333	22100	225056	22166	22112	22108	21988
Minimum	1590	2536	1663	1761	2774	2913	3241	3593
Maximum	153503	161705	153341	153307	156576	157068	158217	156230
Rozpätie	151913	159169	151678	151545	153802	154155	154975	152637
Súčet	1313961	1313961	1313961	1313961	1313961	1313961	1313961	1313961

Štatistiky	Počet EAO pracujúcich v poľnohospodárstve							
	skutočné hodnoty	PV	GKT-EAO	GKT-CD	GKT-ZP	DM	DC	TC
Stredná hodota	5527	5527	5527	5527	5527	5527	5527	5527
Medián	4352	4438	3987	4066	4122	4380	4373	4182
Smer.odchýlka	3992	4008	4099	4076	4056	4038	3980	3982
Minimum	629	820	366	388	611	980	896	1021
Maximum	24073	26657	23726	23697	24805	25201	25482	24728
Rozpätie	23444	25837	23360	23309	24194	24220	24586	23707
Súčet	364762	364762	364762	364762	364762	364762	364762	364762

b) pre rok 2001

Štatistiky	Počet obyvateľov								
	skutočné hodnoty	PV	GKT-EAO	GKT-CD	GKT-ZP	DM	DM-RDBD	DC	TC
Stredná hodota	81507	81507	81507	81507	81507	81507	81507	81507	81507
Medián	60755	59076	60753	60705	59485	60256	60386	58987	59032
Smer.odchýlka	91766	92525	91678	91767	92142	92284	92373	92349	91786
Minimum	5982	7851	5801	6184	9539	9386	9542	8705	8881
Maximum	646724	660888	645901	647179	653862	655086	654887	657893	651807
Rozpätie	640742	653037	640100	640995	644324	645699	645344	649189	642926
Súčet	5379455	5379455	5379455	5379455	5379455	5379455	5379455	5379455	5379455

Štatistiky	Počet obyvateľov predproduktívnom veku								
	skutočné hodnoty	PV	GKT-EAO	GKT-CD	GKT-ZP	DM	DM-RDBD	DC	TC
Stredná hodota	15386	15386	15386	15386	15386	15386	15386	15386	15386
Medián	11763	11825	11697	11714	11647	11813	11829	11843	11821
Smer.odchýlka	15558	15571	15484	15497	15524	15534	15577	15534	15439
Minimum	1367	1625	1579	1683	1687	1660	1687	1539	1570
Maximum	98088	100288	97574	97804	99020	99221	99198	99739	98643
Rozpätie	96721	98663	95995	96121	97333	97561	97511	98200	97072
Súčet	1015493	1015493	1015493	1015493	1015493	1015493	1015493	1015493	1015493

(PV=priestorové váženie, GKT-EAO=metóda geografických konverzných tabuliek vážiaca podľa počtu ekonomicky aktívnych obyvateľov, GKT-CD=metóda geografických konverzných tabuliek vážiaca podľa počtu cenových domácností, GKT-ZP=metóda geografických konverzných tabuliek vážiaca podľa rozlohy zastavanej plochy, DC=metóda dĺžky ciest, TC=metóda typov ciest, DM=dazymetrická metóda, DM-RDBD=dazymetrická metóda pri použití rodinných a bytových domov ako váh, DM-PD=dazymetrická metóda pri použití poľnohospodárskych budov ako váh)

Štatistiky	Počet EAO pracujúcich v poľnohospodárstve									
	skutočné hodnoty	PV	GKT-EAO	GKT-CD	GKT-ZP	DM	DM-RDBD	DM-PB	DC	TC
Stredná hodota	2240	2240	2240	2240	2240	2240	2240	2240	2240	2240
Medián	1816	1847	1781	1766	1798	1745	1780	1777	1816	1750
Smer.odchýlka	1410	1407	1450	1447	1425	1436	1439	1427	1413	1416
Minimum	315	232	233	249	304	383	357	297	353	371
Maximum	6521	7454	6924	6974	7201	7290	7260	6477	7356	7125
Rozpätie	6206	7222	6691	6725	6898	6906	6903	6180	7004	6754
Súčet	147855	147855	147855	147855	147855	147855	147855	147855	147855	147855

Tab. 12. Miery globálnej vhodnosti za jednotlivé premenné a metódy

a) pre rok 1991

	Počet obyvateľov						
	PV	GKT-EAO	GKT-CD	GKT-ZP	DN	DC	TC
S ₁	8992.57	583.69	746.47	3998.24	5397.60	6131.71	4564.65
S ₂	18.98	1.31	1.68	8.49	13.43	14.21	11.48
S ₃	8.62	-0.39	0.21	3.04	5.68	6.91	6.02
zlepšenie oproti PV (abs.)	-	93.51	91.70	55.54	39.98	31.81	49.24
zlepšenie oproti PV (rel.)	-	93.09	91.14	55.27	29.25	25.12	39.52

	Počet obyvateľov v predproduktívnom veku						
	PV	GKT-EAO	GKT-CD	GKT-ZP	DN	DC	TC
S ₁	2384.01	290.08	423.85	1130.90	1486.83	1629.29	1322.15
S ₂	21.28	2.59	3.98	10.17	15.38	16.48	13.76
S ₃	10.29	0.47	1.16	4.19	7.09	8.40	7.43
zlepšenie oproti PV (abs.)	-	87.83	82.22	52.56	37.63	28.85	44.54
zlepšenie oproti PV (rel.)	-	87.81	81.31	52.22	27.73	22.57	35.35

	Počet EAO pracujúcich v poľnohospodárstve						
	PV	GKT-EAO	GKT-CD	GKT-ZP	DN	DC	TC
S ₁	541.71	378.47	376.73	250.99	276.56	387.11	318.83
S ₂	15.46	10.80	10.77	6.61	8.30	11.96	10.35
S ₃	5.69	-1.10	-0.60	0.43	2.27	4.32	3.26
zlepšenie oproti PV (abs.)	-	30.13	30.46	53.67	48.95	28.54	41.14
zlepšenie oproti PV (rel.)	-	30.13	30.35	57.24	46.34	22.62	33.05

b) pre rok 2001

	Počet obyvateľov							
	PV	GKT-EAO	GKT-CD	GKT-ZP	DM-RDBD	DN	DC	TC
S ₁	2696.25	171.08	239.18	1765.31	1825.31	1324.56	2671.62	1763.71
S ₂	7.36	0.46	0.69	5.32	6.52	3.51	7.30	6.02
S ₃	2.78	0.02	0.08	1.92	2.45	1.04	2.96	3.09
zlepšenie oproti PV (abs.)	-	93.65	91.13	34.53	32.30	50.87	0.91	34.59
zlepšenie oproti PV (rel.)	-	93.70	90.64	27.75	11.34	52.29	0.75	18.18

	Počet obyvateľov v predproduktívnom veku							
	PV	GKT-EAO	GKT-CD	GKT-ZP	DM-RDBD	DN	DC	TC
S ₁	564.79	112.35	129.93	398.97	410.78	303.08	551.83	415.49
S ₂	8.69	1.55	2.02	6.60	7.75	4.61	8.45	7.47
S ₃	3.71	0.51	0.62	2.78	3.40	1.74	3.93	4.13
zlepšenie oproti PV (abs.)	-	80.11	76.99	29.36	27.27	46.34	2.29	26.43
zlepšenie oproti PV (rel.)	-	82.13	76.74	24.00	10.80	46.94	2.73	14.04

	Počet EAO pracujúcich v poľnohospodárstve								
	PV	GKT-EAO	GKT-CD	GKT-ZP	DM	DM-RDBD	DN	DC	TC
86.46S ₁	111.61	76.71	75.81	81.33	88.06	82.95	112.74	102.94	86.46
S ₂	6.87	4.63	4.45	4.49	5.12	4.32	7.42	6.08	5.65
S ₃	1.78	-0.66	-0.64	0.74	1.01	0.22	0.30	1.82	1.70
zlepšenie oproti PV (abs.)	-	31.27	32.08	27.14	21.10	25.68	1.01	7.77	22.53
zlepšenie oproti PV (rel.)	-	32.62	35.25	34.68	25.47	37.15	7.95	11.54	17.79

Podľa miery S₃ z oboch testovaných situácií vidieť, že metódy areálovej transformácie geografických dát majú tendenciu nadodhadovať hodnoty daných premenných. Len metóda GKT vážiaca podľa EAO pri premennej PO v prvej testovanej situácii a premennej EAOPOL aj v druhej testovanej situácii spolu s metódou GKT vážiacej podľa CD mierne podhodnotili odhady oboch premenných. Presnejšie však budeme vedieť zhodnotiť odhady jednotlivých premenných prostredníctvom posúdenia individuálnych prípadov (za každú cieľovú jednotku) pri priestorovej vizualizácii – porovnaním jednotlivých máp relatívnych percentuálnych chýb.

Zhodnotenie úspešnosti jednotlivých testovaných metód prostredníctvom výsledkov, ktoré poskytli miery charakterizujúce percentuálny podiel odhadov s relatívnou chybou od -5 do 5 % (t. j. odhadov považovaných za presné) a percentuálneho podielu odhadov

s relatívnou chybou väčšou ako $\pm 50\%$ (t. j. odhadov považovaných už za neprijateľné) sumarizujú za obe testované situácie tab. 13 – 15. Početnosť jednotlivých tried charakterizujúcich relatívne percentuálne chyby odhadu danej premennej v každej cieľovej jednotke zase znázorňujú histogramy (obr. 13 – 15 za rok 1991 a obr. 16 – 19 za rok 2001) vytvorené pre každú premennú zvlášť a za každú použitú metódu. Spomínané tabuľky, ako aj histogramy vychádzajú z podrobných tabuliek odhadnutých a skutočných hodnôt jednotlivých premenných podľa použitej metódy za každú cieľovú geografickú jednotku, ktoré sú v dôsledku veľkého rozsahu uvedené v CD prílohe v dizertačnej práci Madajová (2011).

Výsledky získané analýzou jednotlivých histogramov sa v podstate zhodujú s predchádzajúcimi tvrdeniami. Vo všeobecnosti teda možno považovať metódy GKT vážiace podľa EAO a CD za najúspešnejšie a metódy priestorového váženía a dĺžky ciest za najmenej presné metódy areálovej transformácie dát z okresov do FMR 91-B. V oboch testovaných situáciách sa metóda typov ciest celkovo javila lepšie ako klasická dazymetrická metóda. V druhej testovanej situácii, v ktorej sme testovali dazymetrickú metódu používajúcu ako váhu budovy typu rodinných a bytových domov, však pozorovať, že táto metóda sa v úspešnosti svojich odhadov zaraďuje hneď za metódy GKT. Premenná EAOPOL opäť predstavuje výnimku oproti výsledkom ostatných premenných, predovšetkým v prvej testovanej situácii. V prípade tejto premennej patrili v roku 1991 k najpresnejším metódam metóda GKT vážiaca podľa ZP a klasická binárna dazymetrická metóda, v priemere k najmenej presným sa k metóde priestorového váženía zaradili aj inak úspešné metódy GKT vážiace podľa CD a dazymetrická metóda pracujúca s vrstvou RD a BD. V druhej situácii sa naopak posledná menovaná metóda spolu s klasickou binárnou dazymetrickou metódou radila k najúspešnejším metódam odhadu premennej EAOPOL. Na rozdiel od počiatočného predpokladu a deskriptívnych štatistík sa nepotvrdilo, že dazymetrická metóda pracujúca s vrstvou poľnohospodárskych budov dokáže vylepšiť odhady danej premennej – táto metóda totiž patrila spolu s metódou priestorového váženía a dĺžky ciest k najmenej presným metódam odhadu premennej EAOPOL v roku 2001. Je však potrebné podotknúť, že v prípade tejto premennej, a obzvlášť v roku 2001, dosahovali všetky testované metódy pomerne vyrovnané výsledky (tab. 15).

V oboch testovaných situáciách možno pozorovať odlišné (lepšie) výsledky pri premenných PO a POPRED oproti premennej EAOPOL. Metódy GKT vážiace podľa EAO a CD odhadli hodnoty premennej PO takmer (a v mnohých prípadoch rovno) so 100-percentnou úspešnosťou, pokiaľ ide o odhady do $\pm 5\%$ a extrémnych (nad $\pm 50\%$) odhadov (tab. 13).

Tab. 13. Poradie jednotlivých metód areálovej transformácie pre počet obyvateľov podľa relatívnych percentuálnych chýb

1991					
Odhadovaná premenná	Metóda	Relatívna chyba od -5 do 5 %	Poradie	Relatívna chyba nad \pm 50 %	Poradie
Počet obyvateľov	GKT-EAO	100.00	1	0	1
	GKT-CD	92.31	2	0	1
	GKT-ZP	63.08	3	18.46	2
	TC	63.08	3	24.62	3
	DM	52.31	4	27.69	4
	DC	44.62	5	27.69	4
	PV	43.08	6	32.31	5

2001					
Odhadovaná premenná	Metóda	Relatívna chyba od -5 do 5 %	Poradie	Relatívna chyba nad \pm 50 %	Poradie
Počet obyvateľov	GKT-CD	100.00	1	0.00	1
	GKT-EAO	100.00	1	0.00	1
	DM-RDBD	85.71	2	5.36	2
	GKT-ZP	82.14	3	10.71	4
	TC	78.57	4	8.93	3
	DC	76.79	5	16.07	6
	DM	75.00	6	8.93	3
	PV	71.43	7	12.50	5

Tab. 14. Poradie jednotlivých metód areálovej transformácie pre počet obyvateľov v predproduktívnom veku podľa relatívnych percentuálnych chýb

1991					
Odhadovaná premenná	Metóda	Relatívna chyba od -5 do 5 %	Poradie	Relatívna chyba nad \pm 50 %	Poradie
Počet obyvateľov v predproduktívnom veku	GKT-EAO	89.23	1	1.47	1
	GKT-CD	76.92	2	10.77	2
	GKT-ZP	63.08	3	21.54	3
	TC	55.38	4	27.69	5
	DM	52.31	5	26.15	4
	DC	47.69	6	30.77	6
	PV	40.00	7	30.77	6

2001					
Odhadovaná premenná	Metóda	Relatívna chyba od -5 do 5 %	Poradie	Relatívna chyba nad \pm 50 %	Poradie
Počet obyvateľov v predproduktívnom veku	GKT-EAO	89.29	1	3.57	2
	GKT-CD	85.71	2	1.79	1
	DM-RDBD	83.93	3	7.14	3
	GKT-ZP	76.79	4	12.50	6
	TC	75.00	5	10.71	5
	DM	75.00	5	8.93	4
	PV	71.43	6	14.29	7
	DC	69.64	7	14.29	7

Tab. 15. Poradie jednotlivých metód areálovej transformácie pre počet EAO pracujúcich v poľnohospodárstve podľa relatívnych percentuálnych chýb

1991					
Odhadovaná premenná	Metóda	Relatívna chyba od -5 do 5 %	Poradie	Relatívna chyba nad \pm 50 %	Poradie
Počet EAO pracujúcich v poľnohospodárstve	GKT-ZP	69.23	1	7.69	1
	DM	64.62	2	13.85	2
	TC	56.92	3	20.00	4
	GKT-EAO	53.85	4	13.85	2
	DC	52.31	5	23.08	5
	GKT-CD	50.77	5	16.92	3
	PV	44.62	7	26.15	6

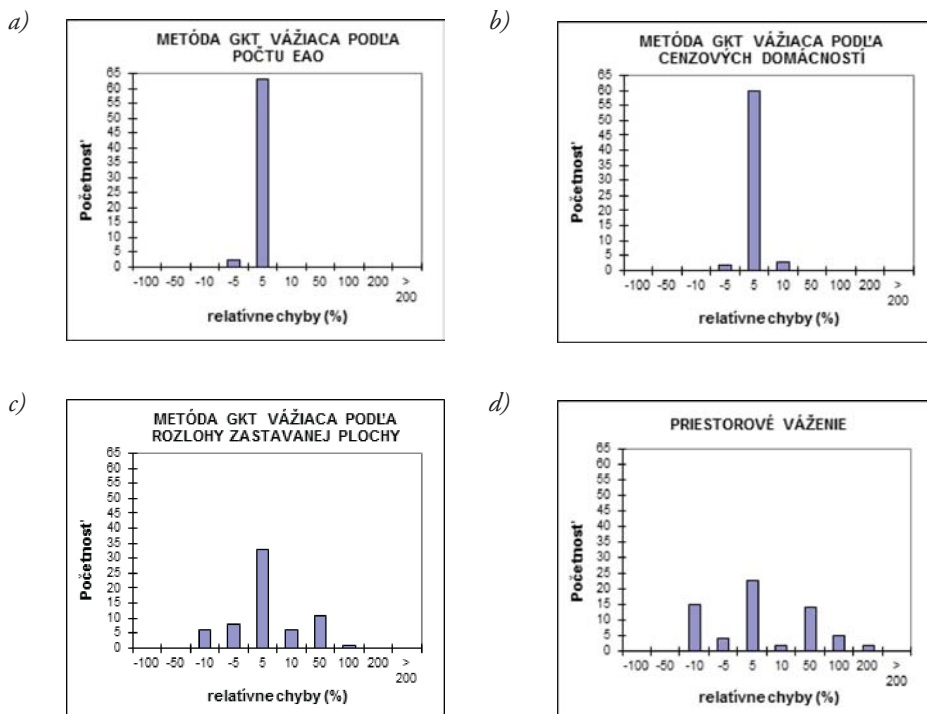
2001					
Odhadovaná premenná	Metóda	Relatívna chyba od -5 do 5 %	Poradie	Relatívna chyba nad \pm 50 %	Poradie
Počet EAO pracujúcich v poľnohospodárstve	DM-RDBD	78.57	1	10.71	3
	DM	79.57	1	10.71	3
	GKT-EAO	76.79	2	7.14	2
	GKT-ZP	76.79	2	10.71	4
	TC	75.00	3	7.14	2
	GKT-CD	73.21	4	5.36	1
	DM-PB	71.43	5	14.29	4
	DC	71.43	5	10.71	3
	PV	64.29	6	14.29	4

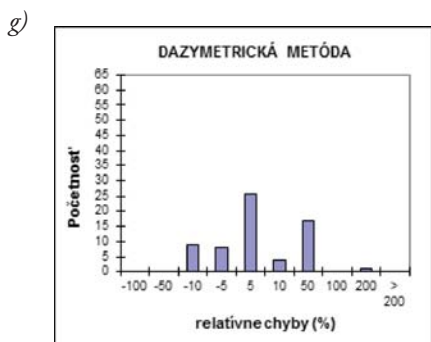
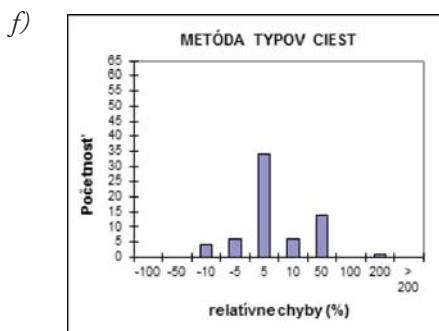
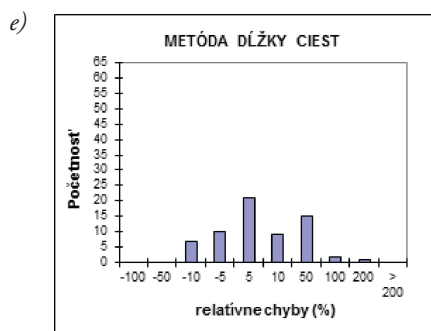
Pokiaľ hodnotíme úspešnosť odhadov metód GKT, ktoré vážili vybrané premenné podľa špecifických váh EAO a ZP, tak na základe výsledkov, ktoré poskytli jednotlivé histogramy, ako aj miery globálnej vhodnosti, sa potvrdilo, že metóda GKT vážiaca podľa EAO bola podľa očakávaní presnejšia pri odhadoch premenných PO a POPRED ako metódy

GKT, ktoré ako váhu použili všeobecnejšie váhy CD a ZP. Pri premennej EAOPOL sa predpoklad úspešnejšieho vystupovania metódy GKT vážiacej podľa ZP ako metód využívajúcich EAO a CD potvrdil len v prvej testovanej situácii, keď táto metóda patrila dokonca k najpresnejším spomedzi všetkých aplikovaných metód. V druhej testovanej situácii nie je rozdiel v presnosti metód GKT vážiacich podľa EAO a ZP viditeľný, metódy dosiahli v priemere v podstate porovnateľné výsledky, s tým rozdielom, že metóda GKT vážiaca podľa EAO dosahuje menej extrémnych odhadov (Tab. 15).

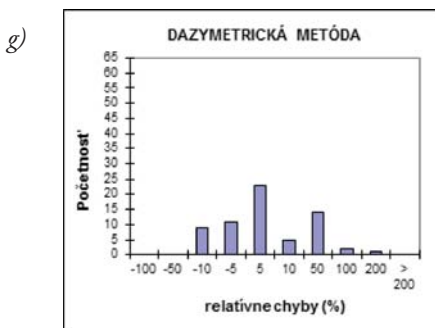
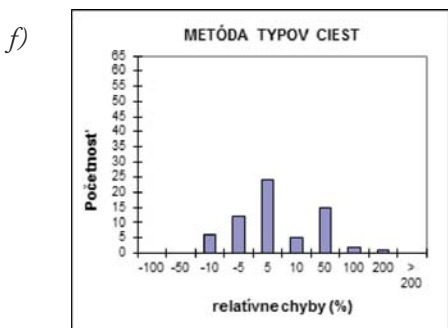
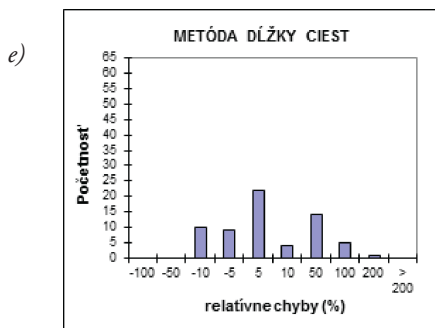
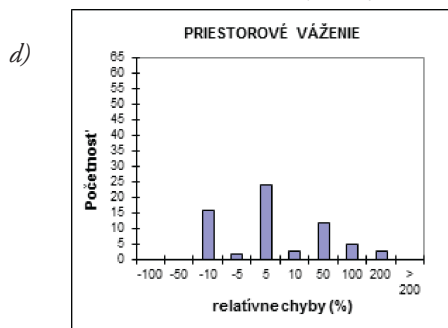
Celkovo možno skonštatovať, že transformácia hodnôt premenných z okresov 1991 do FMR 91-B prebehla v porovnaní s odhadom hodnôt z okresov 2001 do tých istých FMR s omnoho horšími výsledkami v prípade všetkých premenných. Zatiaľ čo sa v prvej testovanej situácii pohybovalo percento akceptovateľných chýb (chyby do $\pm 5\%$ oproti skutočným hodnotám) v globále (t. j. pri zohľadnení všetkých premenných) pri najlepších metódach okolo 70 – 80 % (čím vyššie, tým lepšie) a pri najhorších okolo 45 % zo všetkých odhadov za cieľové geografické jednotky, v druhej testovanej situácii to bolo okolo 90 % a 75 % (tab. 16). Najmenej presné metódy v roku 2001 teda odhadli hodnoty daných premenných s takou percentuálnou úspešnosťou ako najpresnejšie metódy v prvej testovanej situácii. Aj v prípade percentuálnej úspešnosti metód pri zohľadnení chýb nad $\pm 50\%$ je situácia podobná – zatiaľ čo sa úspešnosť najlepších metód v roku 1991 v tomto prípade pohybuje okolo 20 % (čím nižšie, tým lepšie) a najhorších 50 – 80 %, v roku 2001 je tento pomer 5 a 12 %.

Obr. 13a,b,c,d,e,f,g. Histogramy pre počet obyvateľov podľa jednotlivých metód areálovej transformácie (rok 1991)

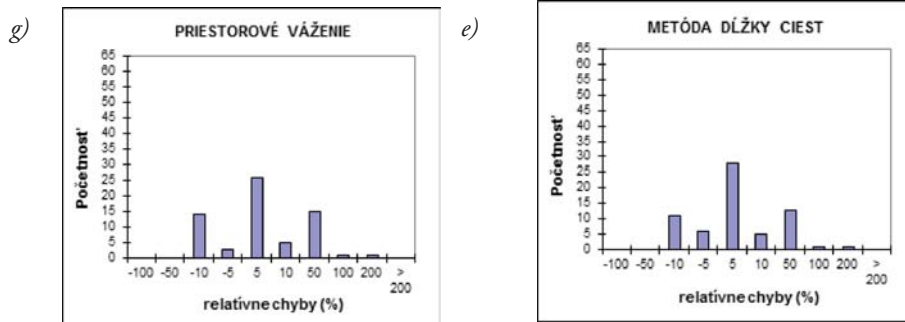




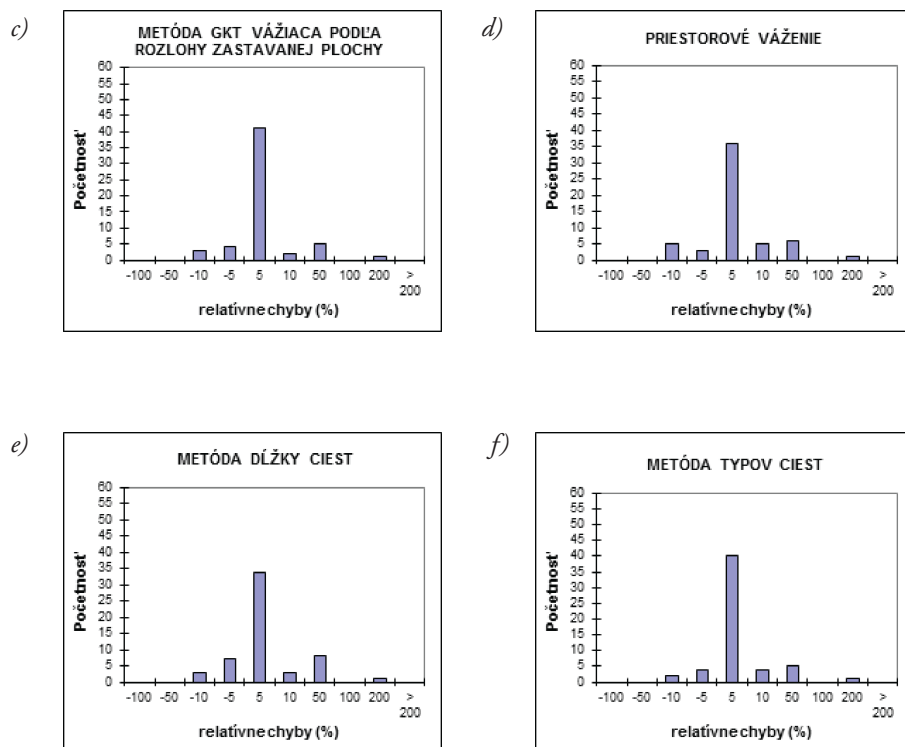
Obr. 14d,e,f,g. Vybrané histogramy pre počet obyvateľov v predproduktívnom veku podľa jednotlivých metód areálovej transformácie (rok 1991)



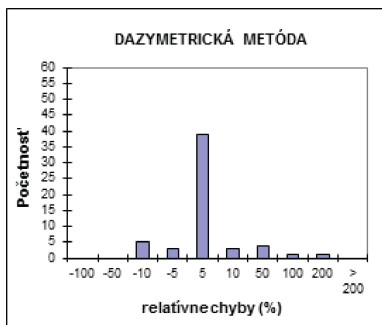
Obr. 15d,e. Vybrané histogramy pre počet EAO pracujúcich v poľnohospodárstve podľa jednotlivých metód areálovej transformácie (rok 1991)



Obr. 16c,d,e,f,g. Vybrané histogramy pre počet obyvateľov podľa jednotlivých metód areálovej transformácie (rok 2001)

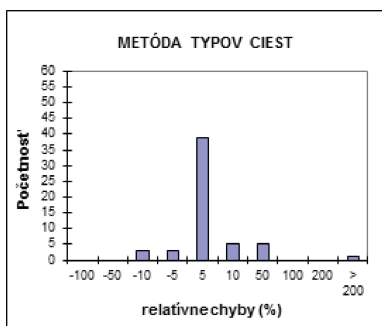


g)



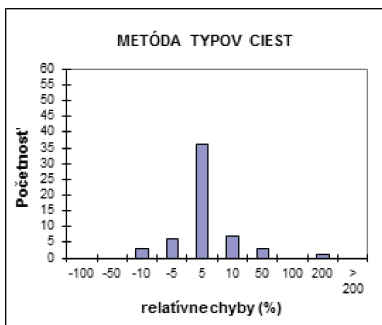
Obr. 17f. Histogram pre počet obyvateľov v predproduktívnom veku podľa metódy typov ciest (rok 2001)

f)



Obr. 18f. Histogram pre počet EAO pracujúcich v poľnohospodárstve podľa metódy typov ciest (rok 2001)

f)



Tab. 16: Spôľahlivosť odhadov a poradie jednotlivých metód areálovej transformácie geografických dát v roku 1991 a 2001 podľa relatívnych chýb v %

1991				
Metóda	Relatívna chyba od -5 do 5 %	Poradie	Relatívna chyba nad ± 50 %	Poradie
GKT-EAO	81.31	1	43.94	5
GKT-CD	73.74	2	51.01	6
GKT-ZP	65.66	3	84.85	7
TC	59.09	4	23.74	2
DM	57.07	5	22.22	1
DC	48.99	6	26.77	3
PV	43.43	7	29.29	4

2001				
Metóda	Relatívna chyba od -5 do 5 %	Poradie	Relatívna chyba nad ± 50 %	Poradie
GKT-EAO	90.40	1	3.03	2
GKT-CD	88.38	2	2.02	1
DM-RDBD	85.35	3	6.57	3
GKT-ZP	81.82	4	9.60	6
TC	79.80	5	7.58	4
DM	79.80	6	8.08	5
DC	76.77	7	11.62	7
PV	73.74	8	11.62	8

Aj pri porovnaní početnosti extrémnych odhadov jednotlivých premenných za obe testované situácie jednoznačne vidieť, že rok 1991 bol menej presný ako rok 2001. Z jednotlivých histogramov máme možnosť pozorovať, že extrémne odhady hodnôt premenných sa skôr nadodhadujú ako pododhadujú. Ani jedna metóda nepodhodnotila v oboch testovaných situáciách odhady premenných o viac ako 50 % oproti ich skutočnej hodnote. V roku 1991 vykazovali všetky metódy početnejšie odhady nad ± 50 % oproti skutočnosti ako v roku 2001 (až na premennú PO, ktorej percentuálna chyba odhadu pri metódach GKT vážiach podľa EAO a CD v oboch testovaných situáciách nepresiahla hodnotu nad 50 %). Chyby nad 100 % pozorovať v prvej testovanej situácii pri premenných PO a POPRED pri dazymetrickej metóde, oboch sieťových metódach a metóde priestorového váženia, pri premennej EAOPOL len pri metóde priestorového váženia a dĺžky ciest (obr. 13 a 14d – g a 15de). V druhej testovanej situácii vykázala chyby nad 100 % pri premennej EAOPOL len metóda typov ciest (obr. 18f) a pri premenných PO a POPRED okrem metód spomenutých v roku 1991 aj metóda GKT vážiaca podľa ZP (obr. 16c-g), tieto však neboli až tak početné ako v prvom prípade. Celkovo dosiahla spomedzi všetkých premenných najhoršie výsledky premenná POPRED, ktorá sa v jednom prípade odhadla extrémnejšie ako nad 200 % oproti jej skutočnej hodnote, a to v druhej testovanej situácii pri metóde typov ciest (obr. 17f).

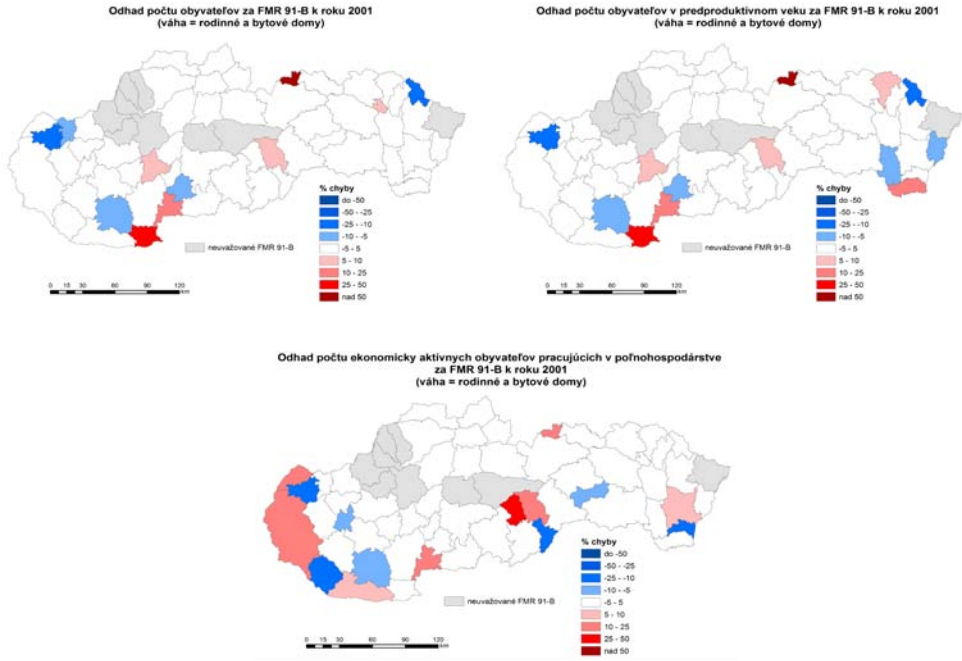
3.2.2. Priestorová vizualizácia chýb

K zaujímavejším výsledkom však dospejeme najmä na základe analýzy máp relatívnych chýb vyjadrených v percentách zostrojených pre všetky premenné a pre každú z metód, z ktorých máme možnosť posúdiť odhady za jednotlivé regióny. To nám umožňuje lepšie preskúmať úspešnosť odhadov hodnôt každej premennej v rámci tej-ktorej metódy, čo nám neskôr pomôže pri konečnom objasňovaní presnosti a spoľahlivosti jednotlivých metód areálovej transformácie geografických dát. Jednotlivé mapy sú k dispozícii v dizertačnej práci Madajová (2011). V predloženej monografii uvádzame ako príklad máp percentuálnych chýb obr. 22 a 23, ktoré dokumentujú odhad testovaných premenných za dazymetrickú metódu používajúcu ako váhu vrstvu budov (rodinných a bytových domov, poľnohospodárskych budov). Vo všetkých mapách bola na porovnanie zvolená rovnaká klasifikácia chýb. Každá mapa relatívnej percentuálnej chyby používa pre negatívne percentuálne chyby (pododhadnuté hodnoty premennej oproti jej skutočným hodnotám) modrú farbu rôznej intenzity, pozitívne percentuálne chyby (nadodhadnuté hodnoty premennej oproti skutočným hodnotám) sú naopak vyjadrené rôznymi odtieňmi červenej farby. Stredná trieda rozmiestnenia je biela a reprezentuje cieľové zóny s relatívnymi percentuálnymi chybami odhadu menšími ako $\pm 5\%$ oproti skutočným hodnotám. Nasledujúca trieda v kladnom i zápornom smere reprezentuje cieľové zóny s absolútnymi chybami medzi 5 – 10 %, ďalšie od 10 – 25 %, 25 – 50 % a 50 a viac %.

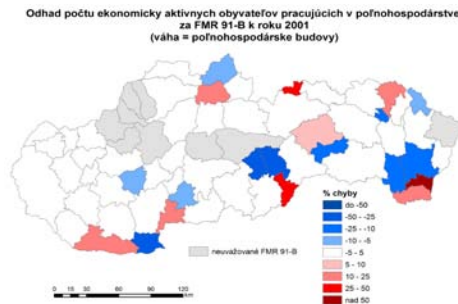
Medzi regióny, ktoré zaznamenali v priemere za všetky odhadované premenné najväčšie kladné chyby, patrili FMR Spišská Stará Ves, Gelnica a Prakovce, Hnúšťa, Tornaľa, Veľké Kapušany, Šahy, Štúrovo v roku 2001 a FMR Spišská Stará Ves, Sobrance, Brezno, Gelnica a Prakovce, Krupina, Námestovo, Nová Baňa a Žarnovica, Zlaté Moravce a Medzilaborce v roku 1991. Zo všetkých premenných a za všetky FMR sa celkovo s najväčšími zápornými chybami odhadli premenné za FMR Spišská Stará Ves, Krupina, Zlaté Moravce a Hnúšťa v roku 1991 a FMR Spišská Stará Ves, Hnúšťa, Tornaľa, Veľké Kapušany, Štúrovo a Medzilaborce v roku 2001. Zatiaľ čo v prvej testovanej situácii ide o regióny omnoho menšie a zároveň plne obsiahnuté v rámci veľkej zdrojovej zóny, v druhej testovanej situácii to až tak neplatí. Z máp teda pozorovať, že za jednotlivé FMR sa môžu hodnoty premenných oproti svojim skutočným hodnotám odhadovať raz lepšie, inokedy horšie a zároveň tak nadodhadovať, ako aj pododhadovať, a to v závislosti od použitej metódy.

Pokiaľ ide o jednotlivé premenné, je nutné rozlišovať medzi premennými PO a POPRED na jednej strane (dosiahli porovnateľné výsledky) a premennou EAOPOĽ na strane druhej. V prvej testovanej situácii sa ukázalo, že pri premenných PO a POPRED boli celkovo za všetky cieľové geografické jednotky najpresnejšie metódy GKT vážiace podľa počtu EAO a CD, keď sa takmer všetky FMR odhadli v rámci akceptovateľných chýb (chýb do $\pm 5\%$). Najmenej presne dané premenné odhadovali metódy priestorového váženía a dĺžky ciest, treba však zdôrazniť, že v niekoľkých prípadoch nepatrili tieto metódy medzi najhoršie – existujú územia, v ktorých sa menej presne prejavili aj iné metódy, predovšetkým však metóda typov ciest a dazymetrická metóda (napr. vo FMR Giraltove, Myjava, Spišská Stará Ves a i.). Pri týchto premenných sa ukázala ako najmenej presná metóda typov ciest, ktorá najviac zo všetkých aplikovaných metód nadodhadla hodnoty spomínaných premen-

Obr. 22. Odhad testovaných premenných za FMR 91-B k roku 2001 binárnou dazymetrickou metódou pracujúcou s vrstvou rodinných a bytových domov



Obr. 23. Odhad počtu ekonomicky aktívnych obyvateľov pracujúcich v poľnohospodárstve za FMR 91-B k roku 2001 binárnou dazymetrickou metódou pracujúcou s vrstvou poľnohospodárskych budov



ných, a to konkrétne vo FMR Spišská Stará Ves (192 % pri premennej POPRED). Najväčší pododhad dosiahla metóda priestorového váženía metóda vo FMR Banská Bystrica (-39 % pri premennej POPRED).

Odlíšne výsledky pozorovať pri premennej EAOPOL, pri ktorej patrili medzi najúspešnejšie metódy tie, ktoré sa v prípade predchádzajúcich premenných radili skôr k priemerným až najmenej presným. Ide o metódu GKT vážiacu podľa ZP a dazymetrickú metódu. Ako najmenej presná sa naopak ukázala metóda GKT vážiacu podľa EAO spolu s metódou priestorového váženía. Metóda GKT vážiacu podľa EAO najviac zo všetkých aplikovaných metód pododhadla premennú EAOPOL vo FMR Krupina (- 48 %). Najväčšie kladné chyby dosiahla metóda priestorového váženía vo FMR Hnúšťa (145 %).

Pokiaľ ide o vystupovanie ostatných metód, tak v prvej testovanej situácii pri metóde priestorového váženía nadobudla premenná EAOPOL najnižšie hodnoty vo FMR Giraltovcie a Prakovce (-34 %). Dazymetrická metóda a metóda GKT vážiacu podľa ZP najviac pododhadli hodnoty premennej EAOPOL vo FMR Krupina (-23 a -33 %) a premenných PO a POPRED vo FMR Hnúšťa (-25 – -30 %). Pri sieťových metódach sa zo všetkých FMR 91-B najviac pododhadovali hodnoty všetkých premenných vo FMR Skalica a Holíč (okolo -30 %) a pri premennej EAOPOL pri metóde typov ciest aj vo FMR Krupina (-22 %). Metóda GKT vážiacu podľa EAO najviac podhodnotila všetky premenné vo FMR Krupina (-6 – -50 %). Vo FMR Krupina poskytla najnižšie odhady premennej EAOPOL aj metóda GKT vážiacu podľa cenových domácností (- 47 %). Pri premenných PO a POPRED v tomto prípade najhoršie vystupovala vo FMR Námestovo (okolo -10 %). Pokiaľ ide o najväčšie kladné chyby, pri metóde GKT vážiacej podľa EAO sa premenná POPRED spomedzi všetkých FMR najviac nadhodnotila vo FMR Gelnica a Prakovce (13 %) a premenná EAOPOL vo FMR Hnúšťa (96 %). V prípade premennej PO táto metóda poskytla vo všetkých FMR pomerne presné výsledky (všetky v rozmedzí od -5 do 5 %). Pri metóde GKT vážiacej podľa CD patrili k regiónom s najviac odhadnutými kladnými hodnotami FMR v prípade premenných PO a POPRED Medzilaborce (9 – 24 %), pri premennej EAOPOL FMR Hnúšťa (95 %). Vo FMR Hnúšťa sa premenná EAOPOL odhadla s najvyššími kladnými chybami aj pri metóde GKT vážiacej podľa ZP (39 %). V prípade premenných PO a POPRED to bol FMR Spišská Stará Ves (58 – 74 %). Pri metóde priestorového váženía sa premenné PO a POPRED nadpriemerne odhadli vo FMR Spišská Stará Ves (112 –134 %), Sobrance (cca 100 %), Medzilaborce (70 – 90 %), Gelnica a Prakovce (80 – 100 %), Krupina a Zlaté Moravce (70 %), Brezno (60 – 70 %) a pri premennej POPRED aj Štúrovo (59 %). V prípade premennej EAOPOL táto metóda poskytla výrazne nadhodnotené odhady len vo FMR Hnúšťa (145 %) a FMR Gelnica a Prakovce (86 %). Pri metóde dĺžky ciest sa hodnoty premennej PO najviac nadhodnotili vo FMR Spišská Stará Ves, Gelnica a Prakovce a Sobrance (56 – 132 %), pri premennej POPRED aj vo FMR Myjava, Krupina, Štúrovo a Medzilaborce (51 – 156 %) a pri EAOPOL vo FMR Gelnica a Prakovce (59 %) a Hnúšťa (120 %). Pri metóde typov ciest a dazymetrickej metóde sa premenná PO najhoršie odhadla vo FMR Spišská Stará Ves (164, resp. 154 %). FMR Spišská Stará Ves patril k najhoršie odhadnutým regiónom v prípade oboch metód aj pri premennej POPRED (192 %. resp. 180 %) spolu s FMR Medzilaborce pri dazymetrickej metóde (69 %) a aj FMR Myjava pri metóde typov ciest (54 a 62 %).

V prípade metód cestnej siete sa premenná EAOPOL najviac nadodhadla vo FMR Hnúšťa (120, resp. 94 %), Gelnica a Prakovce (cca 60 %), a pri metóde typov ciest aj vo FMR Sobrance (61 %). Vo FMR Gelnica a Prakovce sa premenná EAOPOL najviac nadodhadla pri dazymetrickej metóde (56 %).

V prvej testovanej situácii teda patril k najhoršie vystupujúcim regiónom FMR Spišská Stará Ves. Pri metóde priestorového váženía, dĺžky a typov ciest, ako aj pri dazymetrickej metóde sa v ňom odhadnuté hodnoty premenných PO a POPRED oproti ich skutočným hodnotám líšili o viac ako 100 %. V tomto regióne sa celkovo zo všetkých premenných a použitých metód najmenej presne odhadla premenná POPRED (až o takmer 200 %), a to pri metóde typov ciest. Zaujímavým zistením je, že v prípade premennej EAOPOL nepatrilo tento región k najhoršie vystupujúcim. Bol ním FMR Hnúšťa, ktorý sa pri všetkých aplikovaných metódach s výnimkou metódy GKT vážiacej podľa ZP a klasickej binárnej dazymetrickej metódy odhadol najmenej presne v porovnaní s ostatnými regiónmi (chyby v rozmedzí hodnôt 91 – 145 %).

V druhej testovanej situácii (rok 2001) patrili pri premenných PO a POPRED celkovo za všetky cieľové geografické jednotky k najpresnejším metódam metódy GKT vážiace podľa počtu EAO a CD spolu s dazymetrickou metódou využívajúcou vrstvu rodinných a bytových domov (obr. 22 – 23). V priemere najmenej presné výsledky vykázali metóda priestorového váženía a metóda dĺžky ciest. Vo FMR Giraltovce sa za obe premenné a všetky metódy zaznamenal najmenej presný záporný odhad premennej PO pri metóde priestorového váženía (-31 %). Oproti svojim skutočným hodnotám sa premenné PO a POPRED najviac nadodhadli, rovnako ako v roku 1991, vo FMR Spišská Stará Ves (227 % za premennú POPRED) pri metóde typov ciest. Aj v prípade premennej EAOPOL došlo k najväčšej chybe odhadu práve v spomínanom FMR a za metódu typov ciest (110 %). V priemere za všetky cieľové geografické jednotky patrili medzi najmenej presné metódy odhadu tejto premennej, popri metóde priestorového váženía, aj inak celkovo najúspešnejšia metóda GKT vážiaca podľa EAO a dokonca aj dazymetrická metóda využívajúca vrstvu poľnohospodárskych budov, pri ktorej sa očakávalo, že poskytne pomerne presné odhady danej premennej (obr. 23). Pri tejto metóde sa zároveň vyskytla najmenej presná záporná chyba odhadu, a to vo FMR Hnúšťa (-42 %). Medzi najpresnejšie metódy odhadu premennej EAOPOL v roku 2001 možno zaradiť metódu GKT vážiacu podľa ZP, CD a dazymetrickú metódu využívajúcu vrstvu rodinných a bytových domov (obr. 22c).

Pokiaľ hodnotíme presnosť jednotlivých metód areálovej transformácie v druhej testovanej situácii, metóda GKT vážiaca podľa EAO a CD odhadli premenné PO a POPRED vo väčšine prípadov v rámci akceptovateľných chýb -5 – 5 % (pri PO za metódu vážiacu podľa EAO dokonca vo všetkých cieľových geografických jednotkách). Premenná POPRED a pri metóde GKT vážiacej podľa CD aj premenná PO nadobudli najnižšie hodnoty vo FMR Giraltovce a Prakovce (-7 – -15 %), premenná EAOPOL pri oboch metódach vo FMR Tornaľa (-24, resp. -27 %). Pokiaľ ide o najväčšie kladné chyby, premenná POPRED sa spomedzi všetkých FMR najviac nadhodnotila vo FMR Spišská Stará Ves (16, resp. 23 %) a premenná EAOPOL vo FMR Hnúšťa (cca 30 % pri oboch metódach). Vo FMR Hnúšťa sa premenná EAOPOL odhadla s najvyššími kladnými chybami aj pri dazymetrickej metóde využívajúcej vrstvu rodinných a bytových domov (26 %, obr. 22c).

V prípade premenných PO a POPRED to bol taktiež FMR Spišská Stará Ves (63, resp. 95 %). Najmenej presné záporné odhady sa pri danej metóde vyskytli v prípade premenných PO a POPRED vo FMR Senica (-12 a -14 %) a pri premennej EAOPOL vo FMR Veľké Kapušany (-25 %). Hodnoty premennej EAOPOL vo FMR Veľké Kapušany naopak najviac nadodhadla dazymetrická metóda pracujúca s vrstvou poľnohospodárskych budov (71 %, obr. 23). Táto metóda, ako sme už spomínali, zároveň poskytla najmenej presné záporné odhady premennej EAOPOL v roku 2001, a to vo FMR Hnúšťa (-42 %). FMR Hnúšťa patril k regiónom s najviac odhadnutými zápornými chybami aj v prípade premenných PO a POPRED pri klasickej binárnej dazymetrickej metóde (okolo -25 %) Premennú EAOPOL táto metóda najviac pododhadla vo FMR Senica (-21 %). Všetky tri testované premenné nadobudli najvyššie kladné percentuálne chyby v prípade tejto metódy a zároveň aj metódy GKT vážiaca podľa ZP vo FMR Spišská Stará Ves (53 –196 %). Pri metóde GKT vážiacej podľa ZP patrili k regiónom s najmenej presne odhadnutými zápornými hodnotami premenných PO a EAOPOL FMR Senica (-13 a -19 %), pri premennej POPRED to bol FMR Giraltovce (-18 %). Pri sieťových metódach sa zo všetkých FMR 91-B najviac pododhadovali hodnoty premenných PO a POPRED vo FMR Medzilaborce (okolo -20 %) a pri premennej EAOPOL vo FMR Tornaľa (-25, resp. -20 %). V prípade všetkých premenných táto metóda poskytla najmenej presné kladné odhady vo FMR Spišská Stará Ves (81 – 227 %). Pri metóde priestorového váženía aj sa hodnoty všetkých troch premenných najviac pododhadli vo FMR Giraltovce (cca 35 %), výrazne nadhodnotené odhady premenných PO a POPRED dosiahla táto metóda vo FMR Spišská Stará Ves (105, resp. 144 %) a v prípade premennej EAOPOL vo FMR Hnúšťa (71 %).

V druhej testovanej situácii patril k najhoršie vystupujúcim regiónom FMR Spišská Stará Ves. Odhadnuté hodnoty premenných PO a POPRED sa za tento región oproti ich skutočným hodnotám líšili o viac ako 100 % pri metóde typov ciest, dazymetrickej metóde, metóde dĺžky ciest, priestorového váženía a GKT vážiacej podľa ZP. Práve premenná POPRED sa v tomto regióne pri metóde typov ciest odhadla najmenej presne, pokiaľ ide o obe testované situácie, všetky premenné, metódy a ciele geografické jednotky (o 227 %). Za tento región sa spomedzi všetkých prípadov, na rozdiel od roku 1991, najmenej presne a nad 100% oproti svojej skutočnej hodnote odhadla aj premenná EAOPOL, a to opäť aplikovaním metódy typov ciest. Celkovo najmenej presná záporná chyba odhadu bola zaznamenaná v prvej testovanej situácii pri metóde GKT vážiaca podľa EAO a premennej EAOPOL vo FMR Krupina (- 48 %).

Ako vidieť, v rámci jedného FMR sa pri použití tej istej metódy môžu premenné odhadovať s rôznymi výsledkami. Napríklad, zatiaľ čo sa v roku 1991 vo FMR Spišská Stará Ves pri metóde GKT vážiacej podľa ZP odhadli premenné PO a POPRED výrazne nepresne (nad 50 % oproti ich skutočnej hodnote), pri premennej EAOPOL tomu tak nie je (-3 %). Podobne je to aj s nadhodnocovaním a podhodnocovaním hodnôt premenných – premenné PO a POPRED sa nadodhadovali, POPRED sa naopak pododhadla. Existujú teda premenné, ktoré sa v rámci jednej metódy raz nadhodnocujú a inokedy podhodnocujú a zároveň také, ktoré sa odhadujú relatívne presne (myslí sa v rámci akceptovateľných chýb), ale aj také, ktoré vykazujú zlé výsledky. Niekedy možno dokonca pozorovať až extrémne rozdiely v odhadoch hodnôt viacerých premenných v rámci jednej metódy. Na základe

toho sa potom hodnotí nielen presnosť, ale aj spoľahlivosť každej metódy pri odhadoch viacerých premenných. Čím viac premenných daná metóda odhadne najpresnejšie, a čím menšie sú rozdiely v odhadoch jednotlivých premenných za tú istú geografickú jednotku, tým viac možno danú metódu považovať za spoľahlivú. Možno teda s určitou konštatovať, že medzi odhadmi premenných rôzneho charakteru za tú istú geografickú jednotku existujú veľké rozdiely, a to, čo platí pre jednu premennú v rámci jednej metódy, nemusí vôbec platiť pre inú premennú. Podobné výsledky bolo možné v oboch testovaných situáciách pri všetkých metódach a vo väčšine cieľových geografických jednotiek bez ohľadu na ich tvar pozorovať pri premenných PO a POPRED.

3.3. Celkové zhodnotenie výsledkov vybraných metód areálovej transformácie geografických dát

Na základe výsledkov získaných porovnaním jednotlivých štatistických mier použitých na hodnotenie úspešnosti vybraných metód areálovej transformácie dát, ako aj analýzy jednotlivých histogramov a máp znázorňujúcich percentuálne odchýlky odhadov jednotlivých premenných oproti ich skutočným hodnotám, došlo k potvrdeniu začiatočného predpokladu, že areálová transformácia a následný odhad hodnôt premenných z okresov 1991 do FMR 91-B prebehol v porovnaní s odhadom hodnôt z okresov 2001 do tých istých FMR s menej presnými výsledkami, a to v prípade všetkých metód a za všetky premenné. Zároveň možno skonštatovať, že v oboch testovaných situáciách poskytli zo všetkých aplikovaných metód značne konzistentné výsledky metódy založené na princípe geografických konverzných tabuliek. Z nich sa ako najpresnejšie ukázali hlavne tie, ktoré na odhad hodnôt vybraných premenných používali počet ekonomicky aktívnych obyvateľov. Na základe výsledkov odhadu všetkých premenných v druhej testovanej situácii sa ako pomerne sľubná javí binárna dazymetrická metóda používajúca ako váhu vrstvu rodinných a bytových domov. Naopak, v priemere ako najmenej presná sa ukázala metóda priestorového váženia. Pri hodnotení poradia zvyšných metód sa metóda typov ciest javí lepšie ako klasická binárna dazymetrická metóda a tiež ako metóda dĺžky ciest. Metódu dĺžky ciest možno po priestorovom vážení zaradiť medzi najmenej presné metódy areálovej transformácie dát z navzájom nekompatibilných zonálnych systémov.

Uvedené skutočnosti však platia v priemere za všetky odhadované premenné. V procese testovania sa potvrdilo, že to, čo platí pre jednu premennú, nemusí rovnako fungovať aj v prípade druhej odhadovanej premennej. V oboch situáciách sa priemerným výsledkom jednotlivých testovaných metód vymykala premenná EAOPOL, pri ktorej patrili metódy, ktoré zvyšné dve premenné odhadovali najpresnejšie, skôr k tým menej presným. Pri premennej EAOPOL patrili k najúspešnejším metóda GKT vážiaca podľa ZP a klasická dazymetrická metóda. Dazymetrická metóda pracujúca s vrstvou poľnohospodárskych budov v druhej testovanej situácii nedokázala poskytnúť také presné odhady premennej EAOPOL, ako sa od nej očakávalo. V priemere najpresnejšie odhady tejto premennej sa v roku 2001 zaznamenali pri binárnej dazymetrickej metóde s váhou rodinných a bytových domov.

Odlišnú úspešnosť jednotlivých metód areálovej transformácie pri odhade rôznych premenných možno čiastočne vysvetliť na základe korelácie medzi odhadovanou premennou a informáciou používanou na jej odhad. Tu ale nastáva problém, ako tento vzťah korektne zmerať, pretože dáta, ktoré treba odhadnúť, nie sú zvyčajne dostupné za prieniky zdrojových a cieľových jednotiek. Preto sa použili korelačné koeficienty len za zdrojové geografické jednotky, teda za okresy (tab. 17a,b). Takto zostavené poradie korelačných koeficientov medzi jednotlivými premennými a pomocnými údajmi použitými na ich odhad sa ale nie vždy úplne presne zhodovalo s poradím presnosti jednotlivých metód plynúcich z testovania. Len o samotné korelačné koeficienty sa teda nemožno relevantne opierať. V podstate len do istej miery napovedajú o vhodnosti danej metódy pri odhade tej-ktorej premennej. Ak je totiž korelačný koeficient medzi odhadovanou premennou a pomocnou informáciou, ktorú daná metóda využíva, relatívne vysoký (nízky), vieme povedať len to, že daná metóda by mohla byť pomerne úspešná (neúspešná), resp. presná (nepresná), ale napríklad nie to, že ak je tento koeficient najvyšší (najnižší), metóda bude spomedzi ostatných tá najpresnejšia (najmenej presná). Napríklad, zatiaľ čo sa pri odhade premennej PO v roku 1991 poradie korelačných koeficientov medzi počtom obyvateľov a pomocnými premennými (EAO 1; CD 0,99; ZP 0,55; typ ciest 0,54; clc 0,42; dĺžka ciest 0,24 a rozloha 0,18) zhodovalo s poradím presnosti jednotlivých metód vyplývajúcich z testovania, t. z., že metódy GKT vážiace podľa EAO, CD a ZP boli pri odhade premennej PO najpresnejšie a metódy dĺžky ciest a priestorového váženia najmenej presné, tak v druhej testovanej situácii a pri premennej POPRED to až tak jednoznačne neplatí – hoci ide len o drobné zmeny poradia, na základe korelačných koeficientov by mala metóda typov ciest poskytnúť presnejšie odhady ako metóda GKT vážiaca podľa ZP a klasická binárna dazymetrická metóda ako metóda dĺžky ciest. To sa však nepotvrdilo (Tab. 13 a 17). Výraznejšie sa táto skutočnosť prejavila pri premennej EAOPOL. Korelácia medzi touto premennou a pomocnými informáciami bola nasledovná: dĺžka ciest 0,87; poľnohospodárske budovy 0,86; zastavaná plocha 0,83; rozloha a clc 0,82; typ ciest 0,79; rodinné a bytové domy 0,76; cenové domácnosti a ekonomicky aktívny obyvateľia 0,27. Zo samotného testovania však vyplynulo, že pri odhade premennej EAOPOL sa ako najlepšia ukázala dazymetrická metóda s váhou rodinných a bytových domov a klasická binárna dazymetrická metóda. Najmenej presné neboli metódy GKT vážiace podľa CD a EAO, ale metódy dĺžky ciest a priestorového váženia.

K preniknutiu do podstaty procesu odhadu dát a rozpoznanie pravidelnosti v úspešnosti jednotlivých metód pri odhadoch rôznych premenných popri histogramoch významnou mierou prispela priestorová vizualizácia relatívnych chýb vyjadrených v percentách za každú cieľovú geografickú jednotku, ktorá pomohla lepšie objasniť to, čo naznačili výsledky jednotlivých mier globálnej vhodnosti. Porovnaním jednotlivých máp za obe testované situácie a všetky použité metódy jasne vidieť, že :

1. ak sú zdrojové a cieľové geografické jednotky geometricky (rozlohou) podobné, premená sa odhadne s minimálnymi chybami;
2. areálová transformácia hodnôt premenných dosiahla v prvej testovanej situácii v porovnaní s druhou omnoho horšie výsledky.

2.1. Na vyššej úrovni sa teda potvrdilo, že pokiaľ sú zdrojové geografické jednotky oveľa väčšie ako cieľové (okresy 1991 vs. FMR 91-B, t. z. 38 vs. 66 jednotiek), záujmová premenná sa odhadne nepresnejšie, ako tomu bolo v opačnom prípade (okresy 2001 vs. FMR 91-B / 72 vs. 66 jednotiek).

2.2. Táto pravidelnosť sa prejavila aj na nižšej úrovni – myslí sa pri posudzovaní odhadov za každú testovanú situáciu zvlášť, ale len pri metóde priestorového váženía. V prípade ostatných metód sa dá tento fakt výrazným spôsobom zmierniť v závislosti od toho, ako pomocná premenná odzrkadľuje rozmiestnenie odhadovanej premennej.

2.3. Porovnaním odhadov z roku 1991 a 2001 však vidieť, že aj pri metóde priestorového váženía nie vždy platí, že čím je cieľová jednotka v rámci zdrojovej geografickej jednotky menšia, tým je odhad danej premennej nepresnejší. V tomto prípade totiž rovnako platí, že odhad premennej závisí nielen od tvaru (resp. veľkosti) zóny, ale hlavne od toho ako rovnomerne je odhadovaná premenná v rámci zdrojovej zóny rozmiestnená. (Např. v roku 1991 bol FMR Spišská Stará Ves súčasťou okresu Poprad. Ten istý FMR bol ale v roku 2001 začlenený do približne o polovicu menšieho okresu Kežmarok. Ak by platilo, že čím menšia je cieľová geografická jednotka oproti zdrojovej geografickej jednotke, tým horší je odhad premennej pri metóde priestorového váženía, tak potom by to znamenalo, že odhady premennej budú v roku 2001 pre tento FMR omnoho lepšie ako v roku 1991. To sa však potvrdilo len pri premennej počet obyvateľov. Tu sa ukázala dôležitosť testovania metód areálovej transformácie geografických dát z navzájom odlišných zonálnych systémov pri použití viacerých premenných odlišného charakteru, keďže pri premenných POPRED a EAOPOL sa preukázal opak, t. z., že hodnoty týchto premenných sa v roku 2001 odhadli menej presne ako v roku 1991).

Z predchádzajúcich štúdií vieme, že metódy, ktoré využívajú pomocné informácie, sú v priemere schopné redukovať chyby odhadov vo viacerých cieľových geografických jednotkách. Taktiež sa ukázalo, že v rámci jednej metódy sa jednotlivé premenné môžu odhadnúť s rozdielnou presnosťou. Zistilo sa, že premenné PO a POPRED sa odhadovali presnejšie ako premenná EAOPOL, ktorá predstavovala výnimku oproti ostatným premenným v tom zmysle, že pri odhadoch jej hodnôt poskytovali jednotlivé metódy areálovej transformácie odlišné výsledky ako pri zvyšných premenných. Z analýzy máp relatívnych chýb vyjadrených v percentách však vyplynuli ďalšie skutočnosti, a síce, že aj medzi odhadmi premenných rôzneho charakteru za tú istú geografickú jednotku existujú veľké rozdiely a to, čo platí pre jednu premennú v rámci jednej metódy, nemusí vôbec platiť pre inú premennú. Pri všetkých metódach a vo väčšine cieľových geografických jednotiek bez ohľadu na ich tvar poskytovali v oboch testovaných situáciách podobné výsledky premenné PO a POPRED. Spoľahlivosť odhadov jednotlivých metód areálovej transformácie geografických dát je preto, na rozdiel od ich presnosti, nesmierne ťažké zhodnotiť. Ako pomerne spoľahlivé, t. z., že sú vhodné na odhad premenných odlišného charakteru, resp. poskytujú pomerne konzistentné výsledky pri odhadoch rôznych premenných, sa ukázali dazymetrické metódy, obzvlášť metóda pracujúca s vrstvou rodinných a bytových domov, ktorá v priemere za všetky cieľové geografické jednotky patrila k najpresnejším metódam odhadu všetkých testovaných premenných.

Tab. 17. Koefficienty korelácie medzi jednotlivými premennými a pomocnými informáciami:
a) za okresy 1991

	A	B	C	D	E	F	G	H	I	J
A Rozloha okresu (ha)	1									
B Počet obyvateľov	-0.18	1								
C PO v predproduktívnom veku	-0.16	0.99	1							
D Počet EAO	-0.20	1.00	0.98	1						
E EAO prac. v poľnohospodárstve	0.46	0.12	0.09	0.08	1					
F Zastavaná plocha (ha)	0.24	0.55	0.50	0.52	0.79	1				
G Cenzové domácnosti	-0.22	0.99	0.96	0.99	0.09	0.54	1			
H Trieda 1.11 z CLC (ha)	0.27	0.42	0.39	0.38	0.80	0.90	0.38	1		
I Dĺžka ciest (m)	0.68	0.24	0.21	0.21	0.82	0.82	0.21	0.78	1	
J Typ ciest (m)	0.34	0.54	0.51	0.51	0.73	0.90	0.51	0.90	0.85	1

b) za okresy 2001

	A	B	C	D	E	F	G	H	I	J	K	L
A Rozloha okresu (ha)	1											
B Počet obyvateľov	0.26	1										
C PO v predproduktívnom veku	0.34	0.97	1									
D Počet EAO	0.23	1.00	0.95	1								
E EAO prac. v poľnohospodárstve	0.82	0.30	0.37	0.27	1							
F Zastavaná plocha (ha)	0.68	0.69	0.72	0.67	0.83	1						
G Cenzové domácnosti	0.21	0.99	0.94	1.00	0.27	0.67	1					
H Trieda 1.11 z CLC (ha)	0.65	0.66	0.70	0.63	0.82	0.95	0.62	1				
I Rodinné a bytové domy (ha)	0.60	0.77	0.77	0.75	0.74	0.93	0.74	0.94	1			
J Poľnohospodárske budovy a skleníky	0.57	0.34	0.34	0.32	0.86	0.80	0.33	0.8	0.76	1		
K Dĺžka ciest (m)	0.83	0.60	0.64	0.57	0.87	0.93	0.56	0.91	0.87	0.78	1	
L Typ ciest (m)	0.69	0.75	0.78	0.73	0.79	0.95	0.72	0.95	0.94	0.74	0.95	1

Zo všetkých odhadovaných premenných a za všetky aplikované metódy sa najmenej presne odhadla premenná POPRED. Samozrejme, toto tvrdenie neplatí pre všetky cieľové geografické jednotky. Kým vo väčšine regiónov sa jej odhady pohybovali v rámci akceptovateľných chýb, vyskytli sa aj prípady, v ktorých dochádzalo až k extrémnemu nadhodnoteniu tejto premennej. Na základe toho (a to nielen pri tejto, ale pri všetkých premenných) možno skonštatovať, že presnosť odhadu pre každú cieľovú jednotku závisí nielen od toho, ako vážené kritérium koreluje s odhadovanou premennou na úrovni zdrojových geografických jednotiek, ale hlavne od toho ako pomocná informácia používaná na váženie a následný odhad premennej odzrkadľuje jej rozmiestnenie v danej územnej jednotke. Ak je rozmiestnenie (distribúcia) javu vyjadreného vybranou premennou v rámci zdrojovej zóny značne nerovnomerné, prerozdelenie medzi viac než jednu cieľovú jednotku bude mať za následok významné chyby pre každý takýto región. Vážené kritérium by teda malo byť nielen optimálne k skupine dát, ktoré majú byť odhadnuté, ale zároveň aj k ich rôznemu

rozmiestneniu v zdrojových jednotkách. Na základe toho možno potom vysvetliť viaceré skutočnosti, ktoré sa ukázali pri priestorovej vizualizácii chýb, a to, že:

- a) metóda, ktorá sa v priemere zo všetkých metód ukazovala ako najmenej presná, neposkytovala vždy tie najhoršie výsledky, ba dokonca dokázala v niektorých prípadoch danú premennú odhadnúť presnejšie ako zvyšné metódy alebo naopak. Napríklad metóda priestorového váženía patrila v roku 2001 pri odhade premennej EAOPOL vo FMR Veľké Kapušany medzi tie najpresnejšie metódy. Celkovo sa pri metóde priestorového váženía v oboch testovaných situáciách a takmer pri všetkých FMR odhadovala premenná EAOPOL v porovnaní s ostatnými testovanými premennými presnejšie.
- b) metóda, ktorá pri väčšine regiónov danú premennú odhadovala pomerne presne, v iných územiach dosahovala veľmi zlé výsledky alebo naopak. Napríklad odhad premennej POPRED prostredníctvom dazymetrickej metódy vážiacej podľa RD a BD v roku 2001 – pomerne presné výsledky vo všetkých regiónoch okrem FMR Medzilaborce, Senica, Kráľovský Chlmec, Šahy, Štúrovo a Spišská Stará Ves (obr. 22b v prílohe).
- c) niektoré územia vykazovali konzistentné výsledky pri všetkých metódach a za všetky odhadované premenné, iné meniace sa vzory :

* od metódy k metóde, teda už spomínané odlišné odhady jednej premennej v rámci jedného FMR v závislosti od aplikovanej metódy. Napríklad odhad premennej POPRED vo FMR Spišská Stará Ves v roku 2001 – všetky metódy hodnoty danej premennej odhadli s najväčšími chybami spomedzi všetkých geografických jednotiek a odhad premennej PO vo FMR Spišská Stará Ves v roku 2001 – metódy GKT vážiace podľa EAO a CD odhadli hodnoty danej premennej v rámci akceptovateľných chýb do $\pm 5\%$ (-3, resp. 3 %), dazymetrická metóda pracujúca s vrstvou rodinných a bytových domov s chybou 63 % a zvyšné metódy nad 100 %;

* a/alebo v rámci jednej metódy od premennej k premennej, teda taktiež už spomínané rozdielne odhadovanie premenných v jednom FMR pri rovnakej metóde (jedna premenná sa môže odhadovať presne a druhá extrémne nepresne). Napríklad vo FMR Hnúšťa sa premenné PO a POPRED odhadli v rámci akceptovateľných chýb do $\pm 5\%$ (-1, resp. 4,5 %) a premenná EAOPOL s najväčšími chybami spomedzi všetkých regiónov (26,5 %); (obr. 22).

Okrem týchto skutočností je dôležité porozumieť aj nadodhadovaniu a pododhadovaniu hodnôt jednotlivých premenných oproti ich skutočným hodnotám. Ako už vyplynulo z predchádzajúcich tvrdení, v cieľových geografických jednotkách sa môžu v rámci jednej metódy rôzne premenné nadodhadovať aj pododhadovať, pričom pri tej istej premennej si nadhodnotenú a podhodnotenú územia tento fakt nie vždy zachovávajú pri všetkých použitých metódach. Z testovania vyplynulo len to, že extrémne odhady hodnôt premenných sa skôr nadodhadujú ako pododhadujú. Hodnoty ani jednej premennej sa neodhadli so zápornými chybami väčšími ako -50 % oproti ich skutočnej hodnote. Nadhodnocovanie a podhodnocovanie príslušných premenných nemožno vysvetliť prostredníctvom korelácie medzi danou premennou a informáciou použitou na jej odhad. Predpoklad, že záporná korelácia by mohla vysvetliť pododhadovanie a naopak, kladná korelácia nadodhadovanie

hodnôt príslušných premenných sa nepotvrdil a zatiaľ nie je jasné, čo všetko a do akej miery tento proces ovplyvňuje, ale zdá sa, že rozloha cieľovej zóny v rámci zdrojovej zóny na to nepôsobí. Z analýzy v rámci prvej testovanej situácie však došlo k zaujímavému zisteniu, a to, že v miestach, kde zdrojová zóna úplne obsiahne dve alebo viac cieľových zón, máme možnosť sledovať na všetkých mapách nasledovný jav. Ak jeden okres (napr. Banská Bystrica) obsahuje dve alebo viac FMR 91-B (FMR Banská Bystrica a Brezno), potom je hodnota premennej jednej z cieľových zón nadodhadnutá (FMR Banská Bystrica) a druhej pododhadnutá (FMR Brezno). Tento fakt je obzvlášť zreteľný pri metóde priestorového váženia, ale zachováva sa pri všetkých použitých metódach a mohol by naznačovať negatívnu priestorovú autokoreláciu v odhadnutých hodnotách. Aj v tomto prípade sa ale potvrdilo, že proces nadodhadovania a pododhadovania hodnôt premenných neovplyvňuje veľkosť cieľových zón, keďže nie vždy sa menší FMR pododhadol a väčší nadodhadol.

ZÁVER

Získavanie a spracovanie dát je neodmysliteľnou súčasťou práce každého geografa. V regionálnej analýze a vôbec pri posudzovaní základných trendov vývoja spoločnosti však geografi častokrát narážajú na množstvo problémov, ktoré im znemožňujú priame porovnanie, prípadne kombináciu dát z rôznych zdrojov. V tejto súvislosti tak vyvstáva predovšetkým otázka konzistentnosti údajov, a to nielen v čase, ale aj v priestore. Riešením týchto problémov sa zaoberá práve harmonizácia geografických dát, a to prostredníctvom rozličných prístupov. Pod týmto pojmom rozumieme proces zosúladenia navzájom rozdielnych entít, ktorými môžu byť jednak atribútové informácie – premenné (napríklad údaje zo sčítania obyvateľstva), ale aj geografické jednotky, na ktoré sa tieto dáta vzťahujú. V práci sa kládol dôraz na riešenie z nášho pohľadu zložitejšej problematiky harmonizovania dát dostupných za rôzne geografické jednotky, a teda na riešenie problému nekompatibility zonálnych systémov. Samotný proces odvodenia dát z jedného systému geografických jednotiek do druhého, odlišného zonálneho systému, v literatúre označovaný skôr ako problém areálovej interpolácie a nami v rámci procesu harmonizácie geografických dát v užšom zmysle slova chápaný ako problém areálovej transformácie dát z navzájom odlišných zonálnych systémov, však v slovenskej geografickej literatúre patrí k doposiaľ nerozvinutým, hoci veľmi potrebným témam. Práca preto predstavila základný pojmový aparát, zosumarizovala, charakterizovala a v konečnom dôsledku klasifikovala množstvo metód používaných na riešenie tohto problému.

Metódam odhadu geografických dát z navzájom nekompatibilných územných systémov sa venuje veľké množstvo zahraničných autorov, predovšetkým z oblasti GIS, DPZ a štatistického modelovania. Na základe stanovených cieľov, dostupných a požadovaných informácií možno vo všeobecnosti metódy areálovej transformácie geografických dát rozdeliť na metódy priestorovej interpolácie a štatistického modelovania, pričom štatistické metódy bývajú často inkorporované do samotných priestorovo-interpoláčnych metód. Oveľa praktickejšie je však ich delenie podľa využitia pomocných informácií a zachovania objemu odhadovanej premennej. Jednotlivé metódy areálovej transformácie geografických dát možno využiť v rôznych oblastiach, predovšetkým však na porovnanie priestorovej štruktúry za určité časové obdobie a na odhad vybraných premenných za určitú skupinu zón, pre ktorú nám tieto informácie chýbajú, ale sú zároveň dostupné za inú skupinu geo-

grafických jednotiek, z ktorých ich možno odvodiť prostredníctvom rôznych prístupov. Takto možno napríklad určiť, koľko obyvateľov žije v určitom záujmovom území, prípadne odhadnúť hranice dosahu územia zasiahnutého prírodnou katastrofou a v nich určiť počet postihnutých obyvateľov, domov, obchodov atď. Samotný výber metódy použitej na odhad premennej závisí od rôznych faktorov, predovšetkým však od požadovaných údajov, územných jednotiek a od prijatých predpokladov o týchto entitách. Nemalou mierou však k rozhodnutiu, akú metódu použiť, prispievajú aj také skutočnosti, akými sú dostupnosť dát, softvérových možnosti, ale aj obtiažnosť, časová náročnosť implementácie, či odhadovaná presnosť danej metódy.

Práve otázkam presnosti jednotlivých metód areálovej transformácie geografických dát sa venovala druhá časť predkladanej práce, pretože všetky metódy poskytujú „iba“ odhady premennej a ani jedna metóda nie je úplne spoľahlivá. Keďže nebolo možné otestovať všetky metódy, práca aplikuje aspoň tie najzákladnejšie, ktoré sa líšia v predpokladoch o rozmiestnení odhadovanej premennej v zdrojových geografických jednotkách. V slovenských podmienkach a v dvoch rôznych modelových situáciách (z okresov v roku 1991 na jednej strane a okresov v roku 2001 na strane druhej, do systému funkčných mestských regiónov) sa porovnali výsledky najjednoduchšej metódy, priestorového váženia, so zložitejšími metódami založenými na využití pomocných informácií, o ktorých sa predpokladalo, že dokážu skvalitniť samotný proces odhadu dát a poskytnúť lepšie výsledky. Boli nimi binárna dazy-metrická metóda s použitím rozlohy urbanizovanej zástavby z údajov o krajinnej pokrývke Slovenska na jednej strane a rodinných a bytových domov z databázy ZBGIS na strane druhej, metódy cestnej siete založené na využití dĺžky a typov ciest a metódy založené na princípe geografických konverzných tabuliek, v ktorých sa odhadovaná premenná vážila na základe počtu ekonomicky aktívnych obyvateľov, cenových domácností a rozlohy zastavanej plochy. Všetko sú to informácie, o ktorých sa predpokladalo, že sú v nejakom vzťahu s odhadovanými premennými. Okrem v literatúre najčastejšie a často aj jedinej odhadovanej premennej, počtu obyvateľov, sa odhadovali ešte dve ďalšie premenné rôzneho typu s cieľom vzájomného porovnania spoľahlivosti každej metódy. Porovnanie výsledkov jednotlivých metód za rôzne premenné bolo možné, pretože sme pracovali s dátami, ktorých hodnoty sme poznali (za FMR ich bolo možné získať z údajov za obce). Odhadnutú hodnotu premennej za cieľové geografické jednotky bolo teda možné porovnať s jej skutočnou hodnotou a zhodnotiť tak presnosť jednotlivých metód areálovej transformácie geografických dát.

Pre obe testované situácie sa na základe jednotlivých štatistických mier doplnených o histogramy a priestorovú vizualizáciu percentuálnych chýb preukázali veľmi podobné výsledky jednotlivých metód areálovej transformácie geografických dát, len s tým rozdielom, že v prvej testovanej situácii sa premenné odhadli v priemere nepresnejšie, ako v druhej. Všetky aplikované metódy zachovali rozsah hodnôt odhadovaných premenných, čo je dôležitá vlastnosť v procese odhadu dát z navzájom odlišných zonálnych systémov. Potvrdilo sa, že metódy, ktoré využívajú pomocné informácie, sú schopné redukovat' chyby odhadu vo všeobecnosti najmenej presnej metódy priestorového váženia a poskytnúť lepšie výsledky. V oboch testovaných situáciách poskytlí zo všetkých aplikovaných metód značne konzistentné výsledky metódy založené na princípe geografických konverzných tabuliek. Z nich sa ako najpresnejšie ukázali hlavne tie, ktoré na odhad hodnôt vybraných premenných

používali počet ekonomicky aktívnych obyvateľov. Na základe výsledkov odhadu všetkých premenných v druhej testovanej situácii sa ako pomerne sľubná javí binárna dazymetrická metóda používajúca ako váhu vrstvu rodinných a bytových domov. Pri hodnotení poradia zvyšných metód sa metóda typov ciest javí lepšie ako klasická binárna dazymetrická metóda a tiež ako metóda dĺžky ciest. Metódu dĺžky ciest možno po priestorovom vážení zaradiť medzi najmenej presné metódy areálovej transformácie dát z navzájom nekompatibilných zonálnych systémov.

Rozhodnutie testovať úspešnosť odhadov viacerých premenných za dve modelové situácie založené na rôznom priestorovom rozlíšení (z väčšieho počtu geografických jednotiek do menšieho a naopak) sa ukázalo ako mimoriadne užitočné, a to z viacerých dôvodov. Predovšetkým prispelo k potvrdeniu, že presnosť jednotlivých metód areálovej transformácie geografických dát závisí od tvaru, resp. veľkosti zdrojových a cieľových zón. Areálová transformácia a následný odhad hodnôt premenných sa totiž uskutočnili v prvej testovanej situácii (z väčších geografických jednotiek do menších) omnoho nepresnejšie, ako tomu bolo v opačnom prípade. Na druhej strane sa však na príklade metód založených na princípe geografických konverzných tabuliek ukázalo, že tento fakt možno výrazným spôsobom zmierniť až potlačiť tak, že na odhad hodnoty premennej použijeme metódu, ktorá využíva také pomocné údaje, ktoré úzko súvisia s odhadovanou premennou. Celkovo možno teda konštatovať, že existuje spojenie medzi geometriou územia, jeho atribútovými charakteristikami a pomocnými informáciami použitými na ich odhad na jednej strane a chybami na strane druhej. Tento vzťah však vzhľadom na mierku testovaného územia a množstvo odhadovaných premenných zatiaľ nebol kvantifikovaný.

V procese testovania sa ďalej ukázalo, že to, čo platí pre jednu premennú, nemusí rovnako fungovať aj v prípade inej odhadovanej premennej (či už je to v rámci jednej metódy alebo jednej územnej jednotky). Existujú totiž premenné, ktorých odhady vykazovali pomerne veľké rozdiely medzi jednotlivými metódami, ale aj také, pri ktorých boli rozdiely medzi chybami jednotlivých metód pomerne vyvážené. Aj medzi odhadmi premenných rôzneho charakteru za tú istú geografickú jednotku sa niekedy zistili až extrémne rozdiely, a teda to, čo platí pre jednu premennú v rámci jednej metódy, nemusí vôbec platiť pre inú premennú. O podobnosti v presnosti odhadov rôznych premenných v rámci jednej metódy prípadne cieľovej geografickej jednotky možno teda len ťažko hovoriť. Spôľahlivosť odhadov jednotlivých metód areálovej transformácie geografických dát sa preto, na rozdiel od ich presnosti, nepodarilo s určitosťou zhodnotiť. Pomerne spoľahlivé, t. z. vhodné na odhad premenných odlišného charakteru, resp. poskytujúce pomerne konzistentné výsledky pri odhadoch rôznych premenných, sa ukázali dazymetrické metódy, obzvlášť metóda pracujúca s vrstvou rodinných a bytových domov, ktorá v priemere za všetky cieľové geografické jednotky patrila k najpresnejším metódam odhadu všetkých testovaných premenných.

Celkovo sa teda preukázalo, že niektoré územia vykazujú konzistentné výsledky pri všetkých metódach a za všetky odhadované premenné, v iných územiach sa naopak vyskytujú meniace sa vzory od metódy k metóde a rovnako aj v závislosti od odhadovanej premennej. Podobne, ako sa pri jednom regióne môžu odlišne presne odhadovať hodnoty rôznych premenných, rovnako sa v nich môže líšiť aj úspešnosť jednotlivých metód. To znamená, že najhoršia metóda priestorového vážení sa nie vždy prejavila ako tá najmenej

presná a naopak, najpresnejšie metódy v niektorých územiach a pri niektorých premenných dosahovali najmenej presné výsledky.

Hoci sú teda metódy, ktoré využívajú pomocné informácie, v priemere za všetky geografické jednotky schopné vylepšiť výsledky procesu odhadu dát z rôznych zonálnych systémov, presnosť odhadu každej premennej a za každú cieľovú jednotku závisí nielen od toho, ako informácia, používaná na odhad danej premennej koreluje s odhadovanými dátami, ale predovšetkým od toho, ako odzrkadľuje rozmiestnenie odhadovanej premennej v zdrojovej geografickej jednotke. Vážené kritérium by teda malo byť optimálne vzhľadom na skupinu dát, ktoré sa majú transformovať, a zároveň aj rôzne rozmiestnenie v zdrojových jednotkách. Preto mohla napríklad metóda priestorového váženía v niektorých prípadoch vykazovať lepšie výsledky ako zvyšné metódy – bolo to v územiach, kde bola premenná rozmiestnená relatívne rovnomerne.

Preukázalo sa teda, že kvalita pomocných informácií výrazným spôsobom ovplyvňuje výsledky areálovo-transformačného procesu. Z tohto pohľadu sa do budúcnosti javí spomedzi testovaných metód ako mimoriadne sľubná dazymetrická metóda, ktorej presnosť možno významnou mierou zlepšiť, a to v súvislosti s objektívnejším určením váh prísúdených jednotlivým kategóriám urbanizovanej zástavby (triedam krajinnej pokrývky), či pridelením zodpovedajúcich váh jednotlivým obytným budovám (napr. na základe ich veľkosti). Okrem toho existuje ešte množstvo ďalších metód, ktoré by bolo v budúcnosti potrebné navzájom testovať. Určite by bolo mimoriadne zaujímavé sledovať úspešnosť moderných metód založených na neurónových sieťach, prípadne iný spôsob zapojenia viacerých pomocných kategórií údajov do procesu odhadu dát, čím by sa zvýšila nielen presnosť, ale aj komplexnosť danej metódy. Ďalší výskum v tejto oblasti by sa mal pohybovať okrem už aplikovaného rozdielneho priestorového rozlíšenia (regióny rôznej veľkosti) aj v územiach väčšej mierky s rôznou diferenciaciou povrchu (zastúpenie rôznych areálov – obývaných, poľnohospodárskych, priemyselných, zalesnených...), čo by nám malo umožniť hlbšie preniknúť do podstaty procesu odhadovania dát, pomôcť k objasneniu doposiaľ nevyriešeného problému nadhodnocovania a podhodnocovania odhadov jednotlivých premenných, lepšie zhodnotiť spoľahlivosť metód areálovej transformácie pri odhadoch rôznych premenných a, ako dúfame, v konečnom dôsledku aj dospieť k akémusi pravidlu, na základe ktorého by bolo možné zovšeobecniť úspešnosť jednotlivých metód transformácie geografických dát z navzájom odlišných zonálnych systémov v tom zmysle, že by neplatili len pre danú situáciu tak ako to bolo v našom prípade.

LITERATÚRA

- BHADURI, B., BRIGHT, E., COLEMAN, P., DOBSON, J. (2002). LandScan: Locating people is what matters. *Geoinformatics*, 5, 2, 34-37.
- BENTLEY, G. C., CROMLEY, R. G., ATKINSON-PALOMBO, C. (2013). The Network Interpolation of Population for Flow Modeling Using Dasymetric Mapping. *Geographical Analysis*, 45, 3, 307-323.
- BEZÁK, A. (2000). *Funkčné mestské regióny na Slovensku*. Geographia Slovaca, 15. Bratislava (Geografický ústav SAV).
- BEZÁK, A., HOLICKÁ, A. (1995). Komparatívna analýza extrapoláčnych modelov pre regionálne populačné projekcie. *Geografický časopis*, 47, 4, 233-246.
- BIELECKA, E. (2005). A dasymetric population density map of Poland. In *Proceedings of the International Cartographic Conference*. A Coruña, July 9 – 15, 2005. CD-ROM.
- BRACKEN, I. (1991). A surface model approach to small area population estimation. *Town Planning Review*, 62, 2, 225–237.
- CIESIN, IFPRI, WRI (2000). *Gridded Population of the World, Version 4. (GPWv4): Population Count Grid*. Palisades (Center For International Earth Science Information Network, International Food Policy Research Institute, World Resources Institute).
- COCKINGS, S., FISHER, P., LANGFORD, M. (1997). Parameterization and visualization of the errors in areal interpolation. *Geographical Analysis*, 29, 4, 314-328.
- DEICHMANN, U. (1996). *A review of spatial population database design and modelling*. NCGIA Technical Reports, 96-3. Santa Barbara (University of California).
- DEMPSTER, A. P., LAIRD, N.M., RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 39,1-38.
- 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.
- DURHAM, H. (2004). *The development and application of geographically harmonised population databases*. Paper presented at the session of the PopFest (A Population Studies Conference for Postgraduate Students), Leeds, June 22nd - 24th, 2004.
- EICHER, C. L., BREWER, C. A. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, 28, 2, 25-138.
- FERANEC, J., OŤAHEL, J. (2001). *Krajinná pokrývka Slovenska*. Bratislava (Veda).
- FERANEC, J., OŤAHEL, J. (2003). Land cover and landscape changes mapping by application of remote sensing data. *Životné prostredie*, 37, 1. [Online]. Dostupné na: <http://uke.sav.sk/zp/2003/zp1/feranec.htm> [cit: 21-05-2010].
- FISHER, P., LANGFORD, M. (1995). Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation. *Environment and Planning A*, 27, 211-24.
- FISHER, P., LANGFORD, M. (1996). Modeling sensitivity to accuracy in classified imagery: A study of areal interpolation by dasymetric mapping. *Professional Geographer*, 48, 3, 299-309.
- FLOWERDEW, R. (1988). *Statistical methods for areal interpolation: predicting count data from a binary variable*. Lancaster (Northern Regional Research Laboratory).
- 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 and Francis), pp. 239-248.

- FLOWERDEW, R., GREEN, M. (1991). Data integration: statistical methods for transferring data between zonal systems. In Masser, E., Blakemore, M., eds. *Handling geographic information: Methodology and potential applications*. New York (Longman), pp. 38-54.
- FLOWERDEW, R., GREEN, M. (1992). Developments in areal interpolation methods and GIS. *The Annals of Regional Science*, 26, 67-78.
- FLOWERDEW, R., GREEN, M. (1994). Areal interpolation and types of data. In Fotheringham, S., Rogerson, P., eds. *Spatial analysis a GIS*. London (Taylor and Francis), pp. 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*, 70, 303-15.
- FLOWERDEW, R., OPENSHAW, S. (1987). *A review of the problems of transferring data from one set of areal units to another incompatible set*. Lancaster (Northern Regional Research Laboratory).
- FOTHERINGHAM, S., ROGERSON, P. (1994). *Spatial analysis and GIS*. London (Taylor & Francis).
- GALLEGO J., PEEDELL S. (2001). Using CORINE Land Cover to map population density. *Towards Agri-environmental indicators*, Topic report 6/2001. Copenhagen (European Environment Agency), pp. 92-103.
- GALLEGO, J. (2010). A population density grid of the European Union. *Population and Environment*, 31, 6, 460-473.
- GEHLKE, C. E., BIEHL, K. (1934). *Certain effects of grouping upon the size of the correlation coefficient in census tract material*. *Journal of the American Statistical Association*, 29, (185A), 169-170.
- GKÚ (1991). 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).
- GKÚ (2001). Vrstva Hranice obce. Bratislava (Geodetický a kartografický ústav).
- GOODCHILD, M. F. (1978). Statistical aspects of the polygon overlay problem. In Dutton, G. eds. *Harvard Papers on Geographic Information Systems*, 6. Reading (Addison Wesley).
- GOODCHILD, M. F., ANSELIN, L., DEICHMANN, U. (1993). A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, 199, 25, 383-97.
- GOODCHILD, M. F., GOPAL, S. (1989). *The accuracy of spatial databases*. London (Taylor and Francis).
- GOODCHILD, M. F., LAM, N. (1980). Areal interpolation: A variant of traditional spatial problem. *Geoprocessing*, 1, 297-312.
- GOODCHILD, M. F., KEMP, K. (1990). *NCGIA Core Curriculum in GIS*. Santa Barbara (National Center for Geographic Information and Analysis, University of California).
- GREEN, N. E. (1956). Aerial photographic analysis of residential neighborhoods: an evaluation of data accuracy. *Social Forces*, 35, 142-147.
- GREEN, M. (1989). *Statistical methods for areal interpolation: the EM algorithm for count data*. Lancaster (North West Regional Research Laboratory).
- GREEN, M. (1990). Statistical models for areal interpolation. In Harts, J., Owens, H. F. L., and Scholten, H. J., eds. *Proceedings of the First European Conference on Geographical Information Systems*. Utrecht (EGIS Foundation), pp. 392-399.
- 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, 4, 293-314.
- GRIFFITH, D. A. (2013). Estimating missing data values for georeferenced poisson counts. *Geographical Analysis*, 45, 3, 259-284.

- HARVEY, J. T. (2000). Small area population estimation using satellite imagery. *Statistics in Transition*, 4, 4, 611–633.
- HARVEY, J. T. (2002a). Estimating census district populations from satellite imagery: Some approaches and limitations. *International Journal of Remote Sensing*, 23, 10, 2071–2095.
- HARVEY, J. T. (2002b). Population estimation models based on individual TM pixels. *Photogrammetric Engineering and Remote Sensing*, 68, 11, 1181–1192.
- HAWLEY, K., MOELLERING, H. (2005). A comparative analysis of areal interpolation methods. *Cartography and Geographic Information Science*, 32, 4, 411–423.
- HLÁSNY, T. (2007). *Geografické informačné systémy – Priestorové analýzy*. Banská Bystrica (Zephyros a Národné lesnícke centrum – Lesnícky výskumný ústav Zvolen).
- HOLLOWAY, R. S., SCHUMACHER, J., REDMOND, R. (1997). *People & place: Dasymetric mapping using Arc/Info*. Missoula (University of Montana, Wildlife Spatial Analysis Lab).
- 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, 2, 103–121.
- HSU, S.Y. (1973). *Population estimation from ERTS imagery: methodology and evaluation*. Paper presented at the session of the Proceedings of the American Society of Photogrammetry 39th Annual meeting, Washington, DC, March 11–16. 1973.
- IISAKA, J., HEGEDUS, E. (1982). Population estimation from Landsat imagery. *Remote Sensing of Environment*, 12, 259–272.
- KIM, H. (2006). Population estimation using land use land cover data from Landsat TM images - implementation and limitations. *The Geographical Journal of Korea*, 40, 4, 489–496.
- KIM, H. (2007). Comparison of three dasymetric methods for population density mapping. *The Geographical Journal of Korea*, 41, 4, 411–419.
- KIM, H., YAO, X. (2010). Pycnophylactic interpolation revisited: Integration with the dasymetric-mapping method. *International Journal of Remote Sensing*, 31, 21, 5657–5671.
- KYRIAKIDIS, P. C. (2004). A geostatistical framework for area-to-point spatial interpolation. *Geographical Analysis*, 36, 1–3.
- 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. *Handling geographic information*. Essex (Longman Scientific & Technical), pp. 55–77.
- LANGFORD, M., UNWIN, D. J. (1994). Generating a mapping population density surfaces within a geographical information system. *The Cartographic Journal*, 31, 21–26.
- LIN, J., CROMLEY, R. G., CIVCO, D. L., HANINK, D. M., ZHANG, CH. (2013). Evaluating the use of publicly available remotely sensed land cover data for areal interpolation. *GIScience & Remote Sensing*, 50, 2, 212–230.
- LIU, X., CLARKE, K. C. (2002). Estimation of residential population using high resolution satellite imagery. In Maktav, D., Juergens, C., Sunar-Erbek, F., eds., *Proceedings of the 3rd Symposium in Remote Sensing of Urban Area*, June 11–13. 2002. Istanbul (Istanbul Technical University Press), pp. 153–160.
- LO, C. P. (1986). *Applied Remote Sensing*. New York (Longman), pp. 40–70.
- LO, C. P. (1989). A raster approach to population estimation using high-altitude aerial and space photographs. *Remote Sensing of Environment*, 27, 1, 59–71.
- LO, C. P., CHAN, H. F. (1980). Rural population estimation from aerial photographs. *Photogrammetric Engineering and Remote Sensing*, 46, 3, 337–345.
- LO, C. P., WELCH, R. (1977). Chinese urban population estimates. *Annals of the Association of American Geographers*, 67, 2, 246–253.

- MADAJOVÁ, M. (2010). Metódy transformácie priestorových dát: prehľad, klasifikácia a hodnotenie. *Acta Geographica Universitatis Comenianae*, 54, 1, 119-135.
- MADAJOVÁ, M. (2011). *Harmonizácia a areálová transformácia geografických dát: princípy, metódy a aplikácia na území Slovenska*. Doktorská dizertačná práca, Prírodovedecká fakulta UK, Bratislava.
- MAANTAY, J. A., MAROKO, A. R., HERRMANN, C. (2007). Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDs). *Cartography and Geographic Information Science*, 34, 2, 77-102.
- MARKOFF, J., SHAPIRO, G. (1973). The linkage of data describing overlapping geographical units. *Historical Methods Newsletter*, 7, 34-46.
- MARTIN, D. (1989). Mapping population data from zone centroid locations. *Transactions of the Institute of British Geographers*, 14, 90-97.
- MARTIN, D., BRACKEN, I. (1991). Techniques for modeling population-related raster databases. *Environment and Planning A*, 23, 1065-1079.
- MARTIN, D. (1992). Postcodes and the 1991 Census of population: issues, problems and prospects. *Transactions of the Institute of British Geographers*, 17, 350-357.
- MARTIN, D. (1996). *Geographic Information Systems: Socioeconomic applications*. London (Routledge).
- MARTIN, D., DORLING, D., MITCHELL, R. (2002). Linking censuses through time: problems and solutions. *Area*, 34, 1, 82-91.
- MENNIS, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55, 31-42.
- MENNIS, J., HULTGREN, T. (2005). Dasymetric mapping for disaggregating coarse resolution population data. In *Proceedings of the International Cartographic Conference*. A Coruña, July 9 – 15, 2005. CD-ROM.
- MENNIS, J., HULTGREN, T. (2006a). 'Intelligent' dasymetric mapping and its comparison with other areal interpolation techniques. Paper presented at the session of the *Proceedings of AutoCarto*, Vancouver, June 26-28, 2006.
- MENNIS, J., HULTGREN, T. (2006b). Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science*, 33, 3, 179-194.
- MONMONIER, M., SCHNELL, G. A. (1984). Land use and land cover data and the mapping of population density. *International Yearbook of Cartography*, 24, 24-29.
- MROZINSKI, R. D., CROMLEY, R. G. (1999). Singly and doubly constrained methods of areal interpolation for vector-based GIS. *Transactions in GIS*, 3, 285-301.
- NORDBECK, S. (1965). The law of allometric growth. *Michigan Inter-University Community of Mathematical Geographers, Discussion Paper 7*. Ann Arbor (Institute of Mathematical Geography).
- NORMAN, P., REES, P., BOYLE, P. (2001). *Achieving data compatibility over space and time: creating consistent geographical zones*. Working Paper. [Online]. Dostupné na: <http://eprints.whiterose.ac.uk/5016/1/01-6.pdf> [cit: 24-8-2010].
- NORMAN, P., REES, P., BOYLE, P. (2003). Achieving data compatibility over space and time: creating consistent geographical zones, *International Journal of Population geography*, 9, 5, 365-386.
- NORMAN, P. (2004). *Constructing a sociodemographic data time-series: computational issues and solutions*. ESRC Research Methods Programme, [Online]. Dostupné na: www.ccsr.ac.uk/methods/publications. [cit: 21-2-2008].
- NORMAN, P. (2006a). *Sociodemographic spatial change in UK: Data and computational issues and solutions*, [Online]. Dostupné na: http://www.gisdevelopment.net/magazine/years/2006/dec/30_1.htm. [cit: 21-2-2008].

- NORMAN, P. (2006b). *Sociodemographic spatial change in UK: Data and computational issues and solutions*. [Online]. Dostupné na: http://www.gisdevelopment.net/magazine/years/2006/dec/30_2.htm. [cit: 21-2-2008].
- NORMAN, P. (2006c). *Sociodemographic spatial change in UK: Data and computational issues and solutions*. [Online]. Dostupné na: http://www.gisdevelopment.net/magazine/years/2006/dec/30_3.htm. [cit: 21-2-2008].
- OKABE, A., SADAHIRO, Y. (1997). Variation in count data transferred from a set of irregular zones to a set of regular zones through the point-in-polygon method. *International Journal of Geographical Information Science*, 11, 1, 93-106.
- OPENSHAW, S. (1997). Optimal zoning systems for spatial interaction models. *Environment and Planning A*, 9, 169-184.
- OPENSHAW, S. (1983). *The modifiable areal unit problem*. Norwich (Geobooks).
- OWENS, P. M., TITUS-ERNSTOFF, L., GIBSON, L., BEACH, M. L., BEAUREGARD, S., DALTON, M. A. (2010). Smart density: A more accurate method of measuring rural residential density for health-related research. *International Journal of Health Geographics*, 9:8.
- PORTER, P. W. (1956). *Population distribution and land use in Liberia*, Ph.D. dissertation, London School of Economics and Political Science, London.
- PROSPERIE, L., EYTON, R. (2000). The relationship between brightness values from a nighttime satellite image and Texas county population. *The Southwestern Geographer*, 4, 16-29.
- RASE, W. (2001). Volume-preserving interpolation of a smooth surface from polygon-related data. *Journal of Geographical Systems*, 3, 2, 199-213.
- REIBEL, M., BUFALINO, M. E. (2005). Street weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A*, 37, 1, 127-129.
- REIBEL, M., AGRAWAL, A. (2007). Areal interpolation of population counts using pre-classified land cover data. *Population Research and Policy Review*, 26, 619-633.
- ROBINSON, A., MORRISON, J., MUEHRCKE, P., KIMERLING, A., GUPTILL, S. (1984). *Elements of cartography*. New York (John Wiley & Sons).
- ROSINA, K. (2015). *Spatial disaggregation of population density using land cover and remote sensing*. Doktorská dizertačná práca. Prírodovedecká fakulta Univerzity Komenského, Geografický ústav SAV, Bratislava.
- ROSINA, K., HURBÁNEK, P., ATKINSON, P. M. (2012). Priestorová dezagregácia populačných dát s využitím máp krajiny pokrývky a nepriepustnosti povrchu. In Růžička, J. ed. *Symposium GIS Ostrava 2012. Současné výzvy geoinformatiky: proceedings, Ostrava, 23-25.1.2012*. Ostrava (VŠB – Technická univerzita Ostrava), [Online]. Dostupné na: http://gisak.vsb.cz/GIS_Ostrava/GIS_Ova_2012/sbornik/papers/rosina.pdf [cit: 04-8-2016].
- ROSINA, K., HURBÁNEK, P. (2013). Improved method for population disaggregation based on European land monitoring services. In *European Forum for Geostatistics 2013 conference, Sofia 23.-25.10.2013*. Sofia (National Statistical Institute), pp. 1-40.
- ROSINA, K., HURBÁNEK, P. (2016). *Spatial Disaggregation Of Population Density Using Land Cover And Remote Sensing Data*. Geographia Slovaca, 31. Bratislava (Geografický ústav SAV).
- SADAHIRO, Y. (2000a). Accuracy of count data estimated by the point-in-poly method. *Geographical Analysis*, 32, 1, 64-89.
- SADAHIRO, Y. (2000b). Accuracy of count data transferred through the areal weighting interpolation method. *International Journal of Geographical Information Science*, 14, 25-50.
- SAŽP (1996). *GIS vrstva CORINE landcover 1990*, Banská Bystrica (Slovenská agentúra životného prostredia).

SAŽP (2004). *GIS vrstva CORINE landcover 2000*, Banská Bystrica (Slovenská agentúra životného prostredia).

SRIDHARAN, H., QIU, F. (2013). A Spatially disaggregated areal interpolation model using light detection and ranging-derived building volumes. *Geographical Analysis*, 45, 3, 238-258.

RUTHER, M., LEYK, S., BUTTENFIELD, B. (2015). Comparing the effects of an NLCD-derived dasymetric refinement on estimation accuracies for multiple areal interpolation methods, *GIScience and Remote Sensing*, 52, 2, 158-178.

SUTTON, P., ROBERTS, D., ELVIDGE, C., BAUGH, K. (2001). Census from heaven: an estimate of the global human population using night-time satellite imagery. *International Journal of Remote Sensing*, 22, 16, 3061-3076.

SIMPSON, L. (2002). Geography conversion tables: a framework for conversion of data between geographical units. *International Journal of Population Geography*, 8, 69-82.

SIMPSON, L., YU, A. (2003). Public access to conversion of data between geographies with multiple look up tables derived from a postal directory. *Computers, Environment and Urban Systems*, 27, 283-307.

ŠÚ SR (Štatistický úrad Slovenskej republiky). 1991. *Ščítanie obyvateľov, domov a bytov 1991*. Bratislava (ŠÚSR).

ŠÚ SR (Štatistický úrad Slovenskej republiky). 2001. *Ščítanie obyvateľov, domov a bytov 2001*. Bratislava (ŠÚSR).

TAPP, A. F. (2010). Areal interpolation and dasymetric mapping methods using local ancillary data sources. *Cartography and Geographic Information Science*, 37, 3, 215-228.

THIEKEN, A. H., MÜLLER, M., KLEIST, L., SEIFERT, I., BORST, D., WERNER, U. (2006). Regionalisation of asset values for risk analyses. *Natural Hazards and Earth System Sciences*, 6, 167-178.

TOBLER, W. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74, 367, 519-530.

TURNER, A., OPENSHAW, S. (2001). *Disaggregative spatial interpolation*. Paper presented at the GISUK 2001 Conference. Glamorgan, Wales, April 2001.

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*. Working Paper. Madison (Center for Demography and Ecology, University of Wisconsin-Madison).

WEBER, C. (1994). Per-zone classification of urban land use cover for urban population estimation. In Foody, G. M., Curran, P. J., eds. *Environmental remote sensing from regional to global scales*. New York (Wiley), pp.142-148.

WILSON, T., REES, P. (1998). Lookup tables to link 1991 population statistics to the 1998 local government areas, Working Paper 98/5, School of Geography, University of Leeds. [Online]. Dostupné na: <http://www.geog.leeds.ac.uk/wpaper/wp98-5.pdf>. [cit: 7-10- 2001].

WILSON, T., REES, P. (1999). Linking 1991 population statistics to the 1998 local government geography of the United Kingdom. *Population Trends*, 97, 37-45.

WRIGHT, J. K. (1936). A method of mapping densities of population with Cape Code as an example. *The Geographical Review*, 26, 103-110.

WU, S.S., QIU, X., WANG, L. (2005). Population estimation methods in GIS and remote sensing: A review. *GIScience and Remote Sensing*, 42, 80-96.

www.geog.ubc.ca/courses/klink/gis.../u40.html [cit: 7-10- 2009].

www.ccsr.ac.uk/research/afpd/summaryreport.doc [cit: 7-10- 2009].

http://www.geog.leeds.ac.uk/courses/postgrad/geog5105/unit8/Unit8_2004/GEOG5105Unit8_2004.pdf [cit: 7-10- 2009].

- WU, S. S., WANG, L., QIU, X. (2008). Incorporating GIS building data and census housing statistics for sub-block-level population estimation. *The Professional Geographer*, 60, 1, 121-135.
- XIE, Y. (1995). The overlaid network algorithms for areal interpolation problem. *Computers, Environment and Urban Systems*, 19, 287-306.
- XIE, Z. (2006). A Framework for interpolating the population surface at the residential-housing-unit level. *GIScience and Remote Sensing*, 43, 3, 1-19.
- 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.
- Úrad geodézie, kartografie a katastra Slovenskej republiky (2002). *Databáza údajov ÚHDP (Úhrnné hodnoty druhov pozemkov). Stav k roku 1987 a 2001*. Bratislava (Úrad geodézie, kartografie a katastra Slovenskej republiky).
- ZANDBERGEN, P. A. (2011). Dasymeric mapping using high resolution address point datasets. *Transactions in GIS*, 15, 5-27.
- ZANDBERGEN, P. A., IGNIZIO, D. A. (2010). Comparison of dasymetric mapping techniques for small-area population estimates. *Cartography and Geographic Information Science*, 37, 3, 199-214.
- ZB GIS (Základná báza údajov pre geografický informačný systém), verzia 2008, Banská Bystrica (Úrad geodézie, kartografie a katastra SR, Topografický ústav).

SUMMARY

Population data and other human geography data are usually collected for individual persons or objects, however, for various reasons, especially data volume reduction and personal data confidentiality, they are typically published at certain levels of spatial aggregation for several sets of administrative or statistical units, for example communes or districts. On many occasions, there is a demand for these data aggregated for a different set of areal units, for example healthcare, school or electoral districts, natural units such as watersheds or flood zones, or analytical units such as buffer zones derived in a geographic information system. These situations are frequent especially in interdisciplinary research, where quality integration of data from various sources (demographic, economic, social, environmental, etc.) using different spatial divisions of geographical space is a common task. Additional issues related to the inaccessibility of other than the predefined pre-aggregated data include the modifiable areal unit problem (MAUP), which implies that statistics, relationships and patterns obtained from aggregated data are valid only for the particular mode of aggregation and cannot be generalised (Martin 1996). For some applications, the aggregated data are simply too coarse and they cannot adequately represent the underlying geographical distribution with fine spatial resolution variation. Also, territorial units change over time – there are unions, divisions, and small shifts of the boundaries between them. Boundary inconsistencies between two or more sets of areal units that are supposed to be identical, but – as they come from different sources – do not match (for various reasons), represent another frequently occurring problem.

The situation when aggregate data are available for a system of areal units that does not correspond with another system of areal units, for which the data are needed, is called a spatial mismatch problem of zonal systems. The transfer of data from the former to the latter is a process in literature usually referred to as areal interpolation (Goodchild & Lam 1980, Flowerdew & Openshaw 1989, Mrozinski & Cromley 1999, Langford et al. 2001, Lin et al. 2013). Since the spatial mismatch problem of zonal systems can be addressed not only by areal-interpolation techniques, but also by other spatial-interpolation techniques and also other approaches such as statistical modelling, hereinafter the term areal transformation is preferred.

The first part of the study provides an overview of theoretical and methodological base of areal transformation, a classification of methods for areal transformation, and finally a brief description and general comparison of basic methods for areal transformation of geographical data. After an extensive literature review, the following classifications of areal transformation methods are suggested:

- According to content:
 - o Volume-preserving (pycnophylactic) methods
 - o Non-volume-preserving methods (do not complying with the pycnophylactic property/condition)
- According to form:

- o Spatial interpolation methods
 - Point based methods (e.g. inverse distance weighting method)
 - Line based methods (e.g. road network method)
 - Areal/polygon based methods (e.g. areal weighting method, binary dasymetric method)
- o Statistical modelling methods
- o Combined methods
- According to use of ancillary data:
 - o Simple methods using no ancillary data (e.g. inverse distance weighting method, areal weighting method, Tobler's pycnophylactic method)
 - o Intelligent methods using ancillary data
 - Specific methods using ancillary data for source zones (e.g. road network method, binary dasymetric method)
 - Flexible methods using more (complex) ancillary data or ancillary data for target zones (e.g. combined Tobler's pycnophylactic and dasymetric method (Kim & Yao 2010), expectation-maximization algorithm methods, neural network based methods)
 - Hybrid methods combining specific with flexible methods or simple with intelligent methods

The objective of the second part of the study is to assess the accuracy of *selected methods for areal transformation* of geographical data between incompatible zonal systems:

- areal weighting method – using data on surface area of intersecting zones
- binary dasymetric method – using data on surface area of
 - o CORINE Land Cover (CLC) class 1.1 “urban fabric”
 - o ZBGIS classes 16 “family house” and 17 “apartment building” footprints
 - o ZBGIS classes 19 “agricultural building” and 20 “greenhouse” footprints
- road network methods – using SVM50 data on highways (type 1), 1st, 2nd and 3rd class roads (type 2), and main and side roads in built-up areas (type 3), more specifically
 - o simple total length of all type roads
 - o weighted total length of all type roads (types 1-2-3 with weights 5-10-85)
- geography conversion tables – using data on
 - o census household counts
 - o economically active population counts
 - o built-up surface area

Geographical data chosen for this assessment comprise three commonly used population geography variables: total population counts (PO), pre-productive age (0-14) population counts (POPRED), and employed in agriculture population counts (EAOPOL). While these are the variables in focus, whose true values are known only for source zones and whose values for target zones are being estimated (interpolated, transformed), the data described above within the list of selected areal transformation methods are called ancillary or weighting variables, whose true values are known not only for source zones, but also for target zones as well as for the intersections of these two types of zones.

Two pairs of *incompatible zonal systems* were chosen in the study area of Slovakia so that performance of the selected methods could be assessed in two possible transformation settings (based on the relation between the average size of source zones and the average size of target zones): The first transformation from less numerous larger source zones (38 districts from 1991) to more numerous smaller target zones (66 FMR), and the second transformation from more numerous smaller source zones (72 districts from 2001) to less numerous larger target zones (66 FMR). The choice of FMR – functional urban regions based on 1991 census daily travel-to-work data (Bezák 2000) – as target zones in both cases was intentional: While in real-world situations target zones are usually those for which true values of the variables in focus are unknown and therefore need to be estimated by areal transformation, the target zones used in this study are such that the true values of the variables in focus for these zones can be easily derived. This is possible through the fact that all three zonal systems (two sets of districts and a set of FMR) are just three different modes of aggregation of the same set of elementary building blocks, i.e. communes (more-or-less, anyway, as some small commune boundary changes between 1991 and 2001 had to be resolved). Finally, the comparison of the FMR values estimated by areal transformation and the FMR true values aggregated from commune data allows for the accuracy assessment of selected areal transformation methods.

When Y is the estimated variable in focus, X is the ancillary or weighting variable, z denotes a source zone, c denotes a target zone, zc denotes their intersection, and m denotes the number of source zones intersecting the given target zone c , then the estimation process can be formalized by the following formula:

$$\hat{y}_c = \sum_{z=1}^m \hat{y}_{zc} = \sum_{z=1}^m w_{zc} y_z = \sum_{z=1}^m \frac{x_{zc}}{\bar{x}_z} y_z$$

The accuracy assessment of selected areal transformation methods was implemented through comparison of their results – i.e. the estimated values (\hat{y}_c) – with the true values (y_c) using descriptive statistics, three different overall (for all target zones) error measures S_1 , S_2 and S_3 , and one individual-zone-level (for a single target zone) error measure S_{4c} :

$$S_1 = \frac{1}{n} \sum_{c=1}^n |\hat{y}_c - y_c|$$

$$S_2 = \frac{1}{n} \sum_{c=1}^n \frac{|\hat{y}_c - y_c|}{y_i} 100 \%$$

$$S_3 = \frac{1}{n} \sum_{c=1}^n \frac{\hat{y}_c - y_c}{y_c} 100 \%$$

$$S_{4c} = \frac{\hat{y}_c - y_c}{y_c} 100 \%$$

Values of S_1 , S_2 and S_3 associated with each examined combination of variable in focus with ancillary variable (linked with respective areal transformation method) were tabulated

for both of the tested settings (years 1991 and 2001). Frequency distribution and spatial distribution of relative errors in target zones (values S_{4c}) were visualised using histograms and maps, which helped to identify and interpret specific cases substantially diverging from general patterns due to a distinctive interplay of involved factors.

As it was expected, the accuracy assessment confirmed that each of the tested areal transformation methods produced on average larger errors in 1991 setting (when transforming from 38 to 66 zones) than in 2001 setting (from 72 to 66 zones) for a given variable in focus. However, there were several target zone values (e.g. FMR Spišská Stará Ves), which were estimated – rather unexpectedly – more accurately in 1991 (from a larger source zone of Poprad) than in 2001 (from a smaller source zone of Kežmarok). However, this might have been so for some variables in focus (POPRED and EAOPOL), but not for the other (PO), as the example of FMR Spišská Stará Ves (described in parentheses) illustrates on the results of the areal weighting method.

The testing also confirmed that intelligent methods using ancillary data generally yield more accurate estimates than simple methods. However, this finding holds true only to the extent, to which the spatial distribution of variable in focus (Y) correlates with the spatial distribution of ancillary variable (X). Since the values of the two variables are only known for source zones (before areal transformation is approached), Pearson correlation coefficient was calculated at the level of source zones for each of the 46 studied pairs of variables X and Y (21 pairs in year 1991 and 25 pairs in year 2001). The findings show that these 46 values are strongly correlated with 46 respective values of error measure S_i normalized (i.e. divided by sum of y) and also with 46 values of small relative error frequency (S_{4c} between -5% and $+5\%$) – the former yielding $r = (-0,76)$ and the latter $r = 0,76$. If X and Y values were known not only for source zones but also for the intersecting blocks of source and target zones, one could hypothesise that the two r values at the level of these blocks would be even higher (in absolute terms). The difference between the r values for source zones and the r values for these blocks could be then attributed to MAUP, more specifically to the variation of the relationship between X and Y that is intra-zonal (within target zones) and at the same time inter-block (between blocks of the same target zone but of different source zone origin) in nature. However, testing of this hypothesis was beyond the scope of this study and remains to be explored.

In conclusion, it was shown that some methods generate on average better results than others. However, transformation accuracy of certain method may vary considerably between variables in focus. Even an appropriate combination of a specific method with a specific variable may produce very inaccurate estimates for some of the target zones.

Before approaching areal transformation of geographical data, one should examine the correlation between the variable in focus and the ancillary variable (linked with respective areal transformation method) at the level of source zones. Because of MAUP, however, one should also try to evaluate and consider (to the greatest possible extent) the intra-zonal variation of relationship between the two variables. For when space is not homogeneous (which is usually not) the juxtaposition of source and target zone boundaries may bring about unexpected results.

GEOGRAPHIA SLOVACA 32

Posledné vydané čísla / Last issues

24. Cebecauerová, M. (2007). ANALÝZA A HODNOTENIE ZMIEN ŠTRUKTÚRY KRAJINY (na príklade časti Borskej nížiny a Malých Karpát)
Analysis and assessment of changes of landscape structure (case study of a selected part of lowland Borská nížina and mountains Malé Karpaty)
25. Ira, V. ed. (2008). ĽUDIA, GEOGRAFICKÉ PROSTREDIE A KVALITA ŽIVOTA
People, geographical environment and quality of life
26. Ira, V., Lacika, J. eds. (2009). SLOVAK GEOGRAPHY AT THE BEGINNING OF THE 21st CENTURY
27. Ira, V., Podolák, P. eds. (2010). SÍDELNÁ ŠTRUKTÚRA SLOVENSKA (diferenciácie v čase a priestore)
Settlement structure of Slovakia (differentiations in time and space)
28. Michálek, A., Podolák, P. eds. (2014). REGIONÁLNE A PRIESTOROVÉ DISPARITY NA SLOVENSKU, ICH VÝVOJ V OSTATNOM DESAŤROČÍ, SÚČASNÝ STAV A KONZEKVENCIE
Regional and spatial disparities in Slovakia: development in the last decade, the present status and consequences
29. Šuška, P. (2014). AKTÍVNE OBČIANSTVO A POLITIKA PREMIEN MESTSKÉHO PROSTREDIA V POSTSOCIALISTICKEJ BRATISLAVE
Active Citizenship and the Politics of the Urban Environment Transformation in Post-socialist Bratislava
30. Kopecká, M., Rosina, K., Ořáhel, J., Feranec, J., Pazúr, R., Nováček, J. (2015). MONITORING DYNAMIKY ZASTAVANÝCH AREÁLOV
Monitoring the Dynamis of Built-up Areas
31. Rosina, K., Hurbánek, P. (2016). SPATIAL DISAGGREGATION OF POPULATION DENSITY USING LAND COVER AND REMOTE SENSING DATA
Priestorová dezagregácia hustoty zaľudnenia s využitím máp krajiny pokrývky a údajov diaľkového prieskumu Zeme



GEOGRAFICKÝ ÚSTAV SAV
INSTITUTE OF GEOGRAPHY SAS

Bratislava 2016

ISBN 978-80-89548-03-3

ISSN 1210-3519