

ZBYNĚK JANOŠKA, VÍT VOŽENÍLEK, PAVEL TUČEK

INTERVALOVÉ VYMEZENÍ PRŮBĚHU STŘEDNÍ LINIE GEOGRAFICKÝCH JEVŮ ZE SADY POLYGONŮ: PŘÍKLAD VYMEZENÍ REGIONU HANÁ Z MENTÁLNÍCH MAP

JANOŠKA, Z., VOŽENÍLEK, V., TUČEK, P. (2014): Interval Delimitation of Mean Line of Geographical Phenomena from a Set of Polygons: Case Study of Mental Maps for Haná Region. Geografie, 119, No. 1, pp. 91–104. – The paper proposes an innovative method for estimation of mean border from a set of polygons. This method is based on statistical estimation of mean line using distribution of distances at given directions. Confidence intervals for mean line are calculated, which allows to incorporate uncertainty into the analysis. Compared with currently used methods, the proposed procedure offers an objective statistical evaluation of data, it enables analyst to include his/her expert knowledge into the estimation process and enables statistical testing of hypotheses. Confidence intervals can be used to evaluate variability in different directions. Performance of method is demonstrated on both experimental and real data representing mental maps of ethnographic region Haná. The results with experimental data are very promising, since these data are generated by a well-known process. The results with real data are comparable with those, reported by previously published studies.

KEY WORDS: mentální mapy – střední linie – etnografický region Haná.

Autoři děkují za podporu Vnitřní grantové agentury Univerzity Palackého v Olomouci (grant PrF_2013_024) a Operačnímu programu Vzdělávání pro konkurenceschopnost – Evropského sociálního fondu (projekt CZ.1.07/2.3.00/20.0170 Ministerstva školství, mládeže a tělovýchovy České republiky).

1. Úvod

Mentální mapa je grafickým vyjádřením člověka o geografickém prostoru, a lze ji chápat jako subjektivní obraz atraktivity prostoru (Drbohlav 1991, Voženílek 2002, Pravda 2003, Tversky 2003). Mentální mapy mohou sloužit jako zdroj informací zejména pro sociologický či antropologický výzkum (Bláha, Hudeček 2010; Drbohlav 1991; Gould, White 2004; Pocock 1979) V prostředí české geografie byly mentální mapy lynchovského typu několikrát použity k vymezení etnografických regionů, například Slezska (Siwek, Kaňok 2000), Slovácka (Sečková 2007) či Hané (Kyncl 2011), kulturním a historickým regionům se věnují např. Chromý a Skála (2010), souhrnně pak věnují Chromý, Kučerová, Kučera (2009).

Běžně používaný přístup, který byl užit i ve výše uvedených pracích, je dotazníkové šetření, při němž respondenti zaznačí svou představu o vymezení regionu do mapy, a na základě těchto mentálních map je posléze odvozena hranice regionu. Algoritmy vymezení hranice regionu použité ve výše použitých

pracích byly založeny na jednoduchých překryvných funkcích běžně dostupných v GIS produktech. Ve všech uvedených pracích byl výsledný region vymezen jako území zaznačené určitým podílem respondentů. Tento způsob vymezení finálního průběhu hranice regionu je však značně subjektivní a není zcela jasné, jak stanovit hranici, která bude považována za limitní. Siwek a Kaňok (2000) i Sečková (2007) pouze vypočetli podíl zastoupení katastru dané obce mezi všemi dotazníky, zatímco Kyncl (2011) stanovil hranice regionu Haná jako 30%, 50% a 70% kvantil zastoupení katastru mezi zaznačenými mapami. Šerý a Šimáček (2012) pak vymezovali historický region Jesenicka a jeho hranici jako medián. V jejich práci byly regiony Moravy a Slezska definovány absolutní většinou respondentů. Autoři stanovují i hraniční linii obou regionů, včetně zahrnutí jisté nejistoty v jejím vymezení, nicméně jejich způsob vymezení je pouze heuristickou metodou bez opory ve statistické teorii. První dvě práce se nezabývaly výpočtem hranic regionu z mentálních map, protože autoři pouze řešili podíl zastoupení katastru dané obce mezi všemi dotazníky. Oproti tomu Kyncl (2011) tuto hranici stanovil jako n -tý kvantil zastoupení katastru mezi zaznačenými mapami, kde $n = \{30, 50, 70\}$. Bláha a Pastuchová Nováková (2013) volí jiný přístup a nevymezují hranice zaznačených prvků, ale jen jejich výčet, který je pak v mapě zobrazen v jejich „skutečné“ poloze. Zahraniční literatura v této problematice se spíše věnuje určování přesnosti liniových prvků nežli jejich vymezení (Cheung, Shi 2004; Heo a kol. 2012; Ariza-López, Mozas-Calvache 2012). Souvisejícím tématem je pak etnocentrismus, čili tendence umisťovat střed mapy do regionu, z něhož pochází autor (Blair, McCormack 2010; Saarinen 1987).

Cílem příspěvku je rozvinout postup vymezení hranice regionu z mentálních map použitý ve zmíněných pracích o statistický odhad polohy střední linie. Zatímco práce Kyncla (2011) stanovuje pouze bodový odhad střední linie, v této práci je navržen postup získání intervalového odhadu střední linie, tedy intervalu, v němž se střední linie s určitou pravděpodobností nachází. Vnesení jisté míry nejistoty do výsledků analýzy se může zdát kontraproduktivní, ale naopak poskytuje uživateli další informaci, která by jinak byla ztracena. Například fuzzy logika či entropie jsou přístupy, které s hodnocením nejistoty pracují a které jsou stále častěji používány při analýze geografických jevů (Tuček, Pászto, Voženílek 2009; Voženílek 2009). Vývoj postupů, které umožňují snazší vnímání a pochopení geografického prostoru je vítanou pomůckou pro uživatele geografických dat a přispívá ke zvýšení geografické gramotnosti, nezbytné pro správnou práci s prostorovými daty (Voženílek 2002). Použití metody je demonstrováno nejprve na počítačově generovaných datech a posléze na příkladu vymezení regionu Haná.

2. Metody a postup řešení

Pojem střední linie, kterým se tato práce zabývá, nebyl v geografické literatuře doposud exaktně definován. Lze však vycházet z pojmu střední hodnota rozdělení hodnot, což je parametr polohy definovaný jako vážený průměr tohoto rozdělení. Jsou-li X a Y náhodné veličiny, a , b reálné konstanty, pak střední hodnota (E) a rozptyl (D) mají následující základní vlastnosti (Anděl 1978):

$$E(a \cdot X + b) = a \cdot E(X) + b \quad (1)$$

$$D(a \cdot X + b) = a^2 \cdot D(X) \quad (2)$$

$$E(X - E(X)) = 0 \quad (3)$$

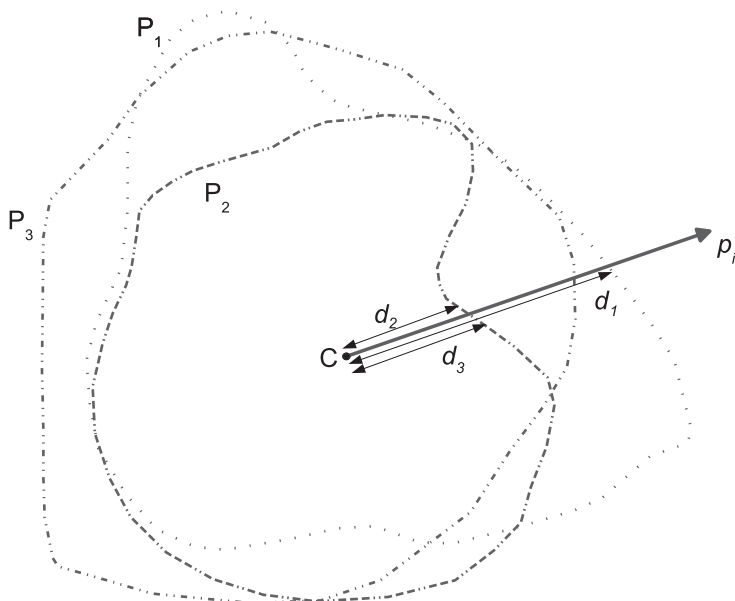
$$D \frac{X - E(X)}{\sqrt{D(X)}} = 1 \quad (4)$$

Střední linie sady polygonů tedy musí vycházet z výše uvedené definice a všechny vlastnosti (1) až (4) musí platit. Střední hodnotu lze vypočítat pouze u takového rozdělení hodnot, kde je rozdělení pravděpodobností známé, popř. jej lze dobře aproximovat. Protože pro data, která jsou výsledkem měření, tato podmínka neplatí, aproximuje se střední hodnota například aritmetickým průměrem nebo mediánem.

Střední linii lze aproximovat pomocí míry polohy rozdělení hranic sady polygonů. Obecně je nutné řídit se pravidly pro konstrukce měř polohy a brát v potaz vlastnosti těchto charakteristik s jejich klady i zápory. Proto nelze jednoznačně doporučit jednu míru polohy (např. aritmetický průměr či medián), která by byla vhodná za všech okolností.

Pro početní vymezení střední linie byl sestaven následující algoritmus:

- krok 1: určit vnitřní bod C tak, aby se nacházel uvnitř všech, nebo alespoň dostatečně velkého počtu studovaných polygonů
- krok 2: vést v libovolném počtu směrů polopřímky s počátkem v bodu C
- krok 3: vypočítat vzdálenosti d průsečíků polopřímky s hranicemi polygonů v každém směru



Obr. 1 – Grafické znázornění pojmů použitých v algoritmu výpočtu střední linie. Použité symboly – viz text.

- krok 4: vypočítat střední hodnotu rozdělení hodnot vzdáleností d v každém směru, jako odhad střední linie v daném směru
- krok 5: spojit vypočtené střední polohy ve všech směrech (odhady střední linie) a tím vytvořit průběh odhadu střední linie.

Při použití tohoto algoritmu splňuje výsledná střední linie všechny podmínky dané rovnicemi (1) až (4) a lze ji aplikovat pro výpočet průběhu střední linie geografických jevů, například hranice regionů z mentálních map, průměrného dosahu signálu z telekomunikačních věží, areály rozšíření organismů z telemetrických měření apod. Formální zápis algoritmu výpočtu střední linie má následující tvar:

- Je dána množina polygonů $P = \{P_1, \dots, P_n\}$ s předpokladem, že většina z těchto polygonů se překrývá, má podobný tvar i velikost.
- Je dán vnitřní bod $C = \{x, y\}$, kterým může být těžiště množiny všech polygonů, nebo předem známý bod, např. historické či správní centrum daného území.
- Pro množinu úhlů $\phi_i = [0; 360]$ je vedena polopřímka p_i z vnitřního bodu C ve směru ϕ_i . Polopřímka p_i protíná hranice polygonů P ve vzdálenostech d_i od středu C (obr. 1).
- Množina $D = \{d_1, d_2, \dots, d_k\}$ je množinou vzdáleností průsečíků polygonů s polopřímkou vedenou ze středu pod určitým úhlem a lze ji považovat za náhodný výběr. Pro toto rozdělení je možné vypočítat jeho průměr \bar{x} a směrodatnou odchylku σ .

Pokud se nahradí funkce, popisující průběh střední linie hodnotou \bar{x} , dojde k tzv. výběrové chybě (Everitt 2003). Úlohou intervalového odhadu je nalézt interval, v němž leží hledaná funkce (v tomto případě hranice regionu) s předem danou pravděpodobností α . Intervalový odhad střední hodnoty veličiny z normálního rozdělení se vypočte následovně:

Předpokládá se, že náhodný výběr pochází z populace, která je popsána pomocí rozdělení $N(\mu, \sigma^2)$. Po znormování platí následující vztah:

$$\frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0,1), \quad (5)$$

tudíž pro konstrukci intervalového odhadu je použito normovaného normálního rozdělení.

Potom je interval

$$\bar{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad (6)$$

100(1- α)-ním oboustranným intervalem spolehlivosti.

Při hodnotě $\alpha = 0,05$ lze intervalový odhad μ vypočítat pomocí vztahu

$$\bar{x} - 1,96 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1,96 \cdot \frac{\sigma}{\sqrt{n}}. \quad (7)$$

Ve skutečnosti není hodnota σ^2 známá a je nutné ji nahradit výběrovým rozptylem.

Při malém rozsahu výběru zřejmě není splněna podmínka normality rozdělení. V tomto případě lze využít centrální limitní věty, podle které se po vhodné normalizaci náhodný výběr blíží normálnímu rozdělení. Pro případ

jiných rozdělení, která se běžně vyskytují při dotazníkovém šetření (nebo obecně řešení kartografických úloh), lze použít pro intervalové odhady následující rovnice:

1. lognormální rozdělení

$$\exp\left[\hat{\mu} - t_{1-\frac{\alpha}{2}}\right] \cdot (n-1) \cdot \frac{\hat{\sigma}}{\sqrt{n}} < median < \exp\left[\hat{\mu} + t_{1-\frac{\alpha}{2}}\right] \cdot (n-1) \cdot \frac{\hat{\sigma}}{\sqrt{n}}, \quad (8)$$

kde $\hat{\mu} = \frac{1}{n} \cdot \sum_{i=1}^n \ln x_i$ je maximálně věrohodný odhad parametru polohy,

$\hat{\sigma} = \frac{1}{n} \cdot \sum_{i=1}^n \ln(x_i - \hat{\mu})^2$ je maximálně věrohodný odhad parametru rozptylu

a $t_1 = \frac{\alpha}{2}$ je příslušný kvantil Studentova t rozdělení.

Vzorec (7) slouží k výpočtu intervalů spolehlivosti pro medián.

2. exponenciální rozdělení

$$\frac{2 \cdot n \cdot \bar{x}}{X_{1-\frac{\alpha}{n}}^2} < \mu < \frac{2 \cdot n \cdot \bar{x}}{X_{\frac{\alpha}{n}}^2}, \quad (9)$$

kde $X_{1-\frac{\alpha}{n}}^2$ a $X_{\frac{\alpha}{n}}^2$ jsou příslušné kvantily rozdělení χ^2 s $2 \cdot n$ stupni volnosti.

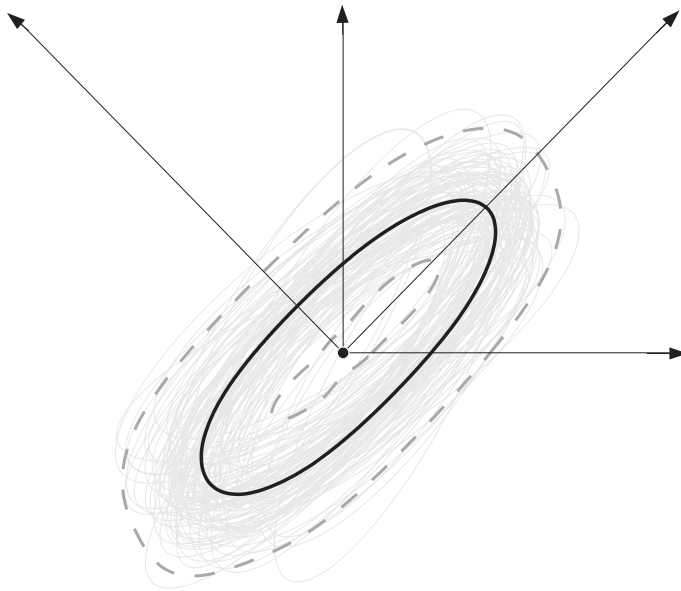
Tyto hodnoty se vypočtou pro všechny směry ϕ a jejich propojením se získá grafická reprezentace střední linie a jejích intervalů spolehlivosti.

Alternativně je možné použít i robustnější odhad konfidenčního intervalu pomocí mediánu a robustních měr variability, např. mezikvartilového rozptylu. Tento způsob může poskytnout uspokojující výsledky v případech, kdy je rozsah výběru malý, nebo existuje podezření, že se v datech vyskytují odlehlé hodnoty, které mohou výsledek zkreslit.

Výše uvedený algoritmus je použitelný jen v případě, kdy se polygony navzájem překrývají. Pokud polygony nesdílejí společný geografický prostor, je určování střední linie tímto způsobem zavádějící. Je-li střední linie definována jako střední hodnota vzdáleností hranice v daném směru, platí pro ni všechny vlastnosti střední hodnoty, jakož statistické míry polohy – viz rovnice (1) až (4). Algoritmus nabízí uživateli řadu parametrů, které může nastavit, např. volbu vnitřního bodu, volbu počtu směrů, v nichž je průměrná linie počítána, a jejich rozestupy. Volba těchto parametrů je podrobněji popsána v kapitole 6.

3. Numerická studie stanovení střední linie z experimentálních dat

Použití metody pro stanovení střední linie z experimentálních dat je demonstrováno na příkladu, kdy jsou data experimentálně generována ze známého rozdělení. Pomocí navrženého postupu byla stanovena střední linie a ověřeno, zda zjištěné parametry odpovídají vstupním datům.



Obr. 2 – Experimentálně generovaná data – 100 elips (šedě), odhad střední linie (černě) a 95% interval spolehlivosti pro střední linii (čárkovaně šedě), šipkou jsou vyznačeny směry, v nichž byla testována normalita

Pro studii bylo náhodně vygenerováno 100 elips se středem v bodě x,y , kde $x,y \sim (0,1)$, délka hlavní poloosa $a \sim (10, \frac{1}{2})$, délka vedlejší poloosa $b \sim (3, \frac{1}{5})$ a sklon hlavní poloosa $\beta \sim (45,7)$ (sklon je uváděn ve stupních; obr. 2).

Vnitřní bod C sady polygonů (elips) P byl v tomto případě známý, jednalo se o bod $[0;0]$. Nicméně jako alternativu lze použít i těžiště všech 100 generovaných elips. Navržený algoritmus byl použit pro výpočet střední vzdálenosti průsečíků polopřímky vedoucí ze středu a generovaných elips. V každém směru byla vypočtena směrodatná odchylka ve 360 směrech rovnoměrně rozložených na kruhu. Byla stanovena střední elipsa a její intervalový odhad na hladině významnosti $\alpha = 0,05$.

Ve čtyřech vybraných směrech (0° , 45° , 90° a 135°) byla testována normalita rozdělení vzdáleností průsečíků elips od středu pomocí Shapiro-Wilkova testu normality. V žádném ze směrů nebylo možné zamítnout nulovou hypotézu o normalitě rozdělení, což prokázalo správnost použití aritmetického průměru a směrodatné odchylky.

Tab. 1 – Srovnání očekávaných a odhadnutých parametrů střední elipsy

Parametr	Očekávaná hodnota	Odhadnutá hodnota
Střed x	0	-0,034
Střed y	0	-0,045
Hlavní poloosa a	10	9,109
Vedlejší poloosa b	3	3,034
Naklonění elipsy β	45	42,931

Ačkoli střední linie (obr. 2, černá tučná čára) vizuálně připomíná elipsu, intervaly spolehlivosti (čárkovaná) se od elipsy výrazně odlišují. Střední linie odhadnutá z dat byla proložena elipsou pomocí metody nejmenších čtverců a srovnána se zadáním (tab. 1). Očekávaná hodnota je parametr rozdělení, z něž byla data generována, Odhadnutá hodnota je odhad této hodnoty metodou nejmenších čtverců z vypočtené střední linie.

4. Vymezení etnografického regionu Haná z mentálních map: případová studie

Region Haná bývá zpravidla etnograficky vymezen poměrně vágně. Exaktní vymezení je obtížné, poněvadž hranice mezi Hanou a sousedními regiony (Horácko, Slovácko) není jednoznačně definovaná (Jeřábek a kol. 2004). Podle Etnografického atlasu Čech, Moravy a Slezska IV (Jeřábek a kol. 2004) Haná zasahuje jižně přes Vyškov až k Bučovicím, na východě se táhne k Otrokovicím a Holešovu, přes Moravskou Bránu kolem úbočí Nízkého Jeseníku a na severu zasahuje až k Mohelnici a Zábřehu. Jako centrum Hané je možné považovat území mezi Vyškovem a Litovlí zahrnující velká města Prostějov, Přerov a Olomouc.

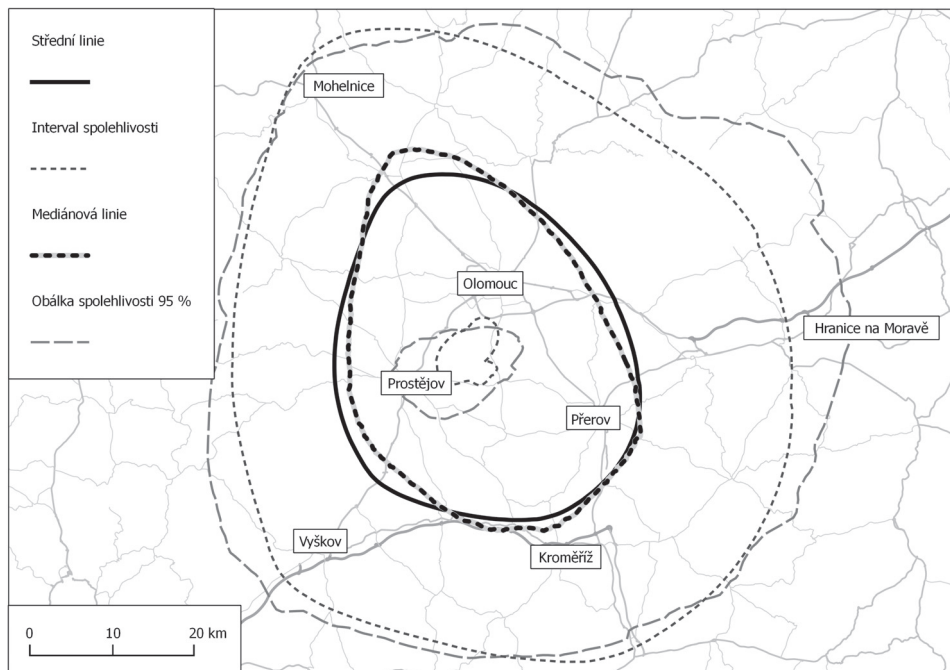
Ve studii Kyncla (2011) bylo pořízeno 317 mentálních map lynchovského typu. Tyto mapy byly použity v případové studii jako sada polygonů P . Dotazníkové šetření proběhlo ve 25 sídlech pokrývajících rovnoměrně celý region Hané vymezený dle Jeřábka a kol. (2004). Každý z respondentů do mapy zakreslil ručně čarou vlastní představu o vymezení regionu Haná. Následně byly shromážděny informace o všech respondentech ve struktuře: pohlaví, věk, bydliště, nejvyšší dosažené vzdělání. Podrobnější rozbor metodiky sběru dat a jejich základní popis provedl Kyncl (2011).

Protože se zakreslené polygony výrazně lišily rozsahem území (zakreslené území se pohybovalo mezi 57 a 12 069 km²), bylo odstraněno 5 % největších a 5 % nejmenších polygonů. Tím se rozsah rozlohy polygonů zmenšil na 296 až 3 795 km².

Protože Haná nemá historicky dané etnografické centrum, bylo za centrum C zvoleno těžiště všech polygonů mentálních map. Celkem 91 % polygonů překrývá toto těžiště, zbylých 9 % bylo z analýzy odstraněno.

Obdobně jako v případě experimentálních dat byly z centra C vedeny polopřímky p_i ve 360 směrech rozmístěných rovnoměrně na kruhu. V každém směru byly vypočteny vzdálenosti d_i průsečíků od centra C . V osmi hlavních směrech (0°, 45°, 90°, 135°, 180°, 225°, 270° a 315°) byl proveden test normality rozdělení. V 7 směrech byla zamítnuta normalita rozdělení na hladině $\alpha = 0,05$, což naznačuje, že aritmetický průměr a směrodatná odchylka nemusí být vhodnými popisnými statistikami. V každém směru proto byly vypočteny nejen aritmetický průměr a směrodatná odchylka, ale i robustní parametry rozdělení – medián a kvantily 2,5 %, 5 %, 25 %, 75 %, 95 % a 97,5 %.

Pomocí navrženého algoritmu byly vypočteny střední linie, linie intervalu spolehlivosti, mediánová linie a obálka spolehlivosti 95 % (obr. 3). Interval spolehlivosti je určen na základě statistické teorie, zatímco obálka spolehlivosti je definována jako interval obsahující určitý počet dat. Vymezení pomocí průměru



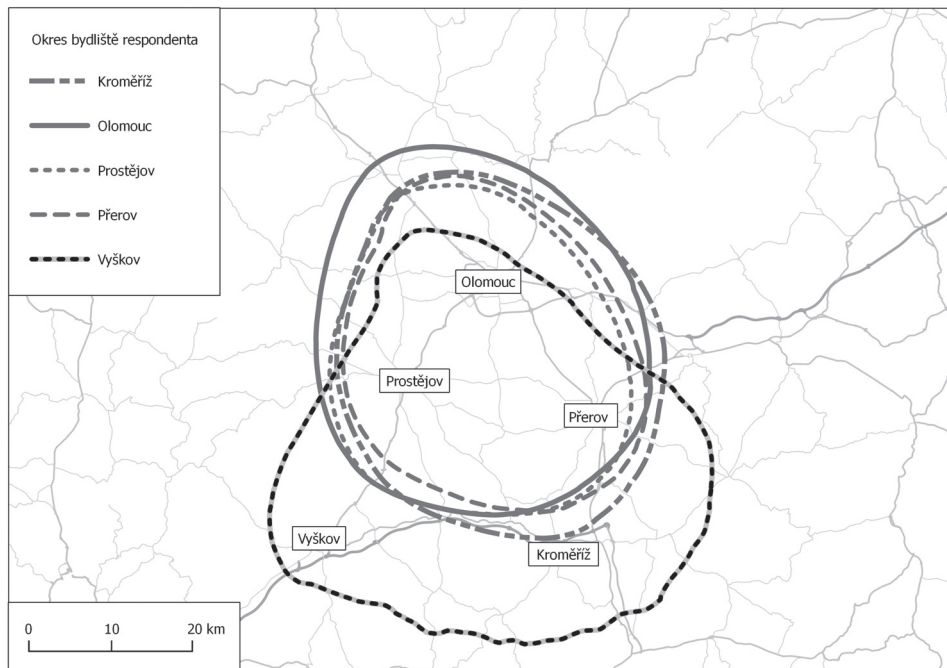
Obr. 3 – Srovnání způsobů vymezení regionu Haná. Zdroj podkladových dat: ČÚZK 2013.

poskytuje hladší průběh střední linie i jejího odhadu. Vymezení pomocí mediánu vykazuje větší lokální variabilitu. Odhad střední linie je u obou postupů velmi podobný. Odhad spodní hranice pomocí průměru je výrazně bližší centru než u kvantilové metody. Naproti tomu odhad horní hranice je u kvantilové metody vyšší, zejména ve východním směru, kde sahá hluboko do Nížkého Jeseníku.

Střední linie vymezená oběma metodami obsahuje města Prostějov, Přerov a Olomouc, kvantilová metoda i město Litovel. Vyškov a Kroměříž nejsou obsaženy v regionu vymezeném střední linií. Absolutní rozdíl mezi středními liniemi vymezenými pomocí průměru a mediánu jsou nevelké. Například na silnici E35 Olomouc–Mohelnice, je tento rozdíl 2,6 km, na téže silnici úsek Olomouc–Lipník nad Bečvou je rozdíl 2,5 km a na silnici R46 Prostějov–Vyškov je rozdíl 2,6 km. Větší rozdíly jsou pozorovatelné u obálek spolehlivosti. V tomto konkrétním případě je však použití intervalů spolehlivosti vzhledem k vysoké variabilitě dat méně vhodné.

5. Rozdíly ve vnímání regionu Haná podle charakteristik respondentů

Ve studii Kyncl (2011) uvedl každý respondent své pohlaví, věk, nejvyšší dosažené vzdělání a bydliště. Na základě těchto charakteristik byly dotazníky rozděleny do skupin a region Haná byl identifikován v každé z těchto skupin s následujícími závěry:



Obr. 4 – Průběh střední linie regionu Haná podle bydliště respondentů. Zdroj podkladových dat: ČÚZK 2013.

- rozdíly podle pohlaví – vymezení regionu respondenty odlišných pohlaví bylo téměř totožné; žádná významná odchylka obou hranic nebyla pozorována
- rozdíly podle věku – respondenti ve skupinách nezletilí (<18 let), v produktivním věku (18–60) a senioři (> 60 let) vymezily region Haná obdobně; neexistují významné rozdíly ve vymezení
- rozdíly podle nejvyššího dosaženého vzdělání – tři skupiny respondentů (základní, střední a vysokoškolské) vymezili region Haná velmi podobně; neexistují výrazné rozdíly
- rozdíly podle místa bydliště – dotazníky byly sbírány ve 25 sídlech, nicméně někteří z dotazovaných nežili v sídle, v němž sběr dotazníků probíhal, takže konečný počet sídel byl vyšší; ke každému bydlišti byl přiřazen okres, pod který sídlo náleželo, čímž vzniklo pět skupin podle okresů bydliště respondenta – Kroměříž, Prostějov, Přerov, Olomouc a Vyškov; hranice regionu Haná zaznačená obyvateli okresů Prostějov, Přerov, Olomouc a Kroměříž byly téměř totožné, zatímco region zaznačený obyvateli okresu Vyškov vykazuje velmi odlišnou hranici (obr. 4) – oproti střední linii skupin zbylých čtyř okresů probíhá vypočtená hranice regionu jižněji, rozloha regionu je výrazně větší a ve své jižní části zasahuje i více na východ i na západ.

Je zajímavé, že respondenti z okresu Kroměříž byli jediní, kteří nezaznačili své okresní město jako součást regionu Haná. Naopak města Prostějov, Přerov a Olomouc byla zaznačena jako součást regionu Haná obyvateli všech okresů, zatímco město Vyškov jen obyvateli okresu Vyškov. Město Litovel zaznačili

jako součást Hané pouze obyvatelé okresu Olomouc. Výrazný rozdíl mezi vymezením Hané obyvateli Vyškovska a ostatních okresů je zajímavým zjištěním a podnětem pro další směr výzkumu. Tyto výsledky se shodují se závěry Siwka a Bogdové (2007), jejichž výzkum ukázal, že rozdíly ve vymezení národopisných regionů Česka jsou závislé na historické zemi původu a též vzdělání, oproti tomu pohlaví nehraje při vymezení podstatnou roli.

6. Diskuse

Autory navržený postup vymezení střední linie průběhu geografických jevů je založena na odhadu střední vzdálenosti v libovolném počtu směrů a následném propojení takto odhadnutých hodnot. Oproti dříve používaným metodám tento přístup nabízí možnost statistického testování, čímž výrazně zvyšuje objektivitu interpretace výsledků. Existuje řada otázek, které lze pomocí této metody zodpovědět, například:

- Je vymezení regionu muži odlišné od vymezení téhož regionu ženami?
- Jsou mentální mapy zaznačené lidmi s vyšším vzděláním méně variabilní než mapy zaznačené lidmi s nižším vzděláním?
- Je vymezení regionu v určitém směru stabilnější než vymezení v jiném směru?
- Existují významné rozdíly mezi mapami zaznačenými lidmi žijícími v různých regionech?
- Je střední hranice daného regionu hladká, nebo existují výrazné lokální výkyvy?
- Které území lze označit za jádrové?
- Jak velký vliv na vymezení střední linie mají odlehlá pozorování? Dojde po jejich vyřazení k výrazným změnám ve tvaru/ploše vymezeného území?

Statistické testování je však možné pouze za předpokladu, že průsečíky polygonů v daném směru mají rozdělení hodnot blízké normálnímu či jinému teoretickému rozdělení. Tato podmínka u dat reprezentujících mentální mapy regionu Haná nebyla splněna, a proto nebylo statistické testování provedeno.

I bez možnosti statistického testování však navržená metoda přináší velkou výhodu v možnosti výpočtu statistických charakteristik rozdělení vzdáleností v daném směru (Voženílek, Kaňok a kol. 2011). Aritmetický průměr a rozptyl podávají odlišné informace než kvantily a jsou standardním způsobem popisu dat i při nesplnění podmínky normality. Dále je možné využít expertních znalostí při studiu geografického jevu, například je-li znám střed území, jehož střední hranice je určována. Další výhodou navrženého přístupu je skutečnost, že odhadnutá hranice je hladká a neobsahuje výrazné zlomy, čili je vizuálně příjemnější než hranice určená na základě n -tého kvantilu.

Naopak nevýhodou navržené metody je nutnost stanovení vnitřního bodu C , což nemusí být vždy jednoznačná úloha. Mohou nastat případy, kdy neexistuje bod, který by ležel uvnitř všech sledovaných polygonů – což byl i případ mentálních map regionu Haná. Polygony, které nepřekrývají zvolený vnitřní bod C , musejí být z analýzy vyřazeny, což mírně snižuje věrohodnost metody. Ovšem stanovení vnitřního bodu C záleží na úsudku uživatele. V případě, že je studovaný jev dobře teoreticky popsán, doporučujeme použít expertní znalosti

pro stanovení polohy bodu C . V případě, že jsou znalosti o jevu omezené, jeví se jako vhodné těžiště plochy všech polygonů, vstupujících do analýzy.

Důležitým parametrem je počet směrů, v nichž je vzdálenost střední linie počítána, a jejich rozestupy. Na menším počtu směrů (4 nebo 8 základních směrů) lze ověřit, zda data splňují podmínku normality, avšak vlastní určení střední linie musí být založeno na větším počtu směrů, nejlépe několika desítek. Při obou provedených analýzách (kap. 4 a 5) bylo použito 360 směrů při konstantním rozestupu 1° . Je však možné ve směrech, kde je žádaná přesnost nejvyšší, stanovit kratší rozestup a naopak ve směrech, kde je hrubší průběh střední linie, postačuje stanovit intervaly širší. Například může být žádoucí, aby hranice regionu Haná byla určena co nejpřesněji ve své jižní a západní části, kde sousedí s regiony Slovácka a Horácka a naopak severovýchodní hranice, která vede podhůřím Nízkého Jeseníku, může být stanovena hruběji.

Obecně lze říci, že navržená metoda je vhodná v případech, kdy jsou si vyhodnocované polygony množiny P velmi podobné, vznikly nějakým automatizovaným postupem, případně je-li znám střed území, jehož hranice má být odhadnuta. Pokud jsou zkoumané polygony odlišné, nepokrývají přibližně stejné území, nebo střed území není znám, nelze využít výhody metody, kterými je zejména zakomponování expertní znalosti a dále možnost statistického odhadu střední linie včetně intervalů spolehlivosti. V těchto případech je lepší použít metodu založenou na určení n -tého kvantilu obdobně jako v pracích Siwka a Kaňoka (2000), Sečková (2007) Kyncla (2011) nebo Šerého a Šimáčka (2012).

Možným zdrojem chyb je použití nekonvexních polygonů, protože by v jednom směru mohlo dojít k několikanásobnému křížení téhož polygonu (z téhož důvodu jsou vyřazeny polygony, které neobsahují vnitřní bod C). V takovém případě záleží na úsudku analytika, zda je smysluplné takovýto polygon z analýzy vyřadit, zda započítat všechny vzdálenosti průsečíků, zda započítat jen jednu z nich (nejbližší, nejvzdálenější, průměrnou apod.) nebo zda v daném směru určovat střední linii pro první, druhé atd. křížení.

7. Závěr

Autoři si kladli za cíl navrhnout metodu, která by rozšířila v současnosti používané metody stanovení střední linie průběhu geografických jevů na základě množiny polygonů vymezujících jeho hranice. Navržená metoda je založena na stanovení intervalového odhadu střední linie ve vybraných směrech a následného propojení těchto intervalů do spojitě křivky. Oproti předcházejícím pracím zabývajícím se podobným tématem, nabízí tato metoda zejména možnost objektivního stanovení variability v daném směru a dále umožňuje statistické testování hypotéz. Dále je možno rovněž studovat i vzniklou geometrii daných útvarů, jak autoři ukazují v práci Tuček a Janoška (2013).

Využití navržené metody při geografickém výzkumu je perspektivní. Příkladem může být odhad dosahu signálu z telekomunikační věže, či odhad střední hranice území, na němž je slyšet zvuk z varovné sirény, nebo odhad území zasaženého spadem z bodového zdroje znečištění, například komínu továrny. Další využití se nabízí při extrahování plošných objektů z leteckých a družicových snímků (Kudělka a kol. 2012). V případech, kdy se jedná o určení

oblasti kolem bodového centra, a existuje důvod domnívat se, že všechny plochy použité v analýze byly generovány stejným procesem, nabízí navržená metoda neoddiskutovatelné výhody zejména díky objektivnímu stanovení intervalů střední linie a možnostem statistického testování, jak bylo ukázáno na příkladu s počítačově generovanými daty. Pro vymezení etnografických regionů z dat polygonů mentálních map nepřináší navržená metoda výrazné výhody. Vzhledem k povaze dat nelze stanovit intervaly spolehlivosti na základě statistické teorie a metoda vymezení hranice na základě n -tého kvantilu tudíž zřejmě poskytuje srovnatelné výsledky.

Literatura:

- ANDĚL, J. (1974): *Matematická statistika*. SNTL, Praha, 346 s.
- ARIZA-LÓPEZ, F. J., MOZAS-CALVACHE, A. T. (2010): Comparison of four line-based positional assessment methods by means of synthetic data. *Geoinformatica*, 16, č. 2, s. 221–243.
- BLÁHA, J. D., HUDEČEK, T. (2010): Hodnocení kartografických děl mentálními mapami. *Kartografické listy*, 18, s. 21–28.
- BLÁHA, J. D., PASTUCHOVÁ NOVÁKOVÁ, T. (2013): Mentální mapa Česka v podání českých žáků základních a středních škol. *Geografie*, 118, č. 1, s. 59–76.
- BLAIR, J. G., McCORMACK, J. H. (2010): *Westers Civilization with Chinese Comparisons*, Fudan University Press, 322 s.
- DRBOHLAV, D. (1991): Mentální mapa ČSFR. Definice, aplikace, podmíněnost. *Geografie*, 96, č. 3, s. 163–176.
- EVERITT, B. S. (2003): *The Cambridge Dictionary of Statistics*, Cambridge University Press, 442 s.
- CHEUNG, C. K., SHI, W. Z. (2004) Estimation of the positional uncertainty in line simplification in GIS. *Cartographic Journal*, 41, č. 1, s. 37–45.
- GOULD, P. R., WHITE, R. R. (2004): *Mental Maps*, Routledge, 184 s.
- HEO, J., KIM, J. W., PARK, J S., SOHN, H.-G. (2012): New Line Accuracy Assesment Methodology Using Nonlinear Least-Squares Estimation. *Journal of Surveying Engineering*, 134, č. 1, s. 13–20.
- CHROMÝ, P., SKÁLA, P. (2010): Kulturně-geografické aspekty rozvoje příhraničních periferií: analýza vybraných složek územní identity obyvatelstva Sušicka. *Geografie*, 115, č. 2, s. 223–246.
- CHROMÝ, P., KUČEROVÁ, S., KUČERA, Z. (2009): Kulturní regiony a geografie kultury. *Kulturní reálie a kultura v regionech*. ASPI, Praha, s. 348.
- JERÁBEK, R. a kol. (2004): *Etnografický atlas Čech, Moravy a Slezska IV: etnografický a etnický obraz Čech, Moravy a Slezska (1500–1900): národopisné oblasti, kulturní areály, etnická a etnografické skupiny*. AV ČR, Etnologický ústav, Praha, 89 s.
- KUDĚLKA, M., HORÁK, Z., VOŽENÍLEK, V., SNÁŠEL, V. (2012): Orthophoto Feature Extraction and Clustering, *Neural Network World*, 22, č. 2, s. 103–121.
- KYNCL, M. (2011): Generování střední linie z datasetu mentálních map. *Univerzita Palackého, Olomouc*, 42 s.
- POCOCK, D. C. D. (1979): The Contribution of Mental Maps in Perception Studies. *Geography*, 64, s. 189–208.
- PRAVDA, J. (2003): *Stručný lexikon kartografie*. Veda, Bratislava, 325 s.
- SAARINEN, T. F. (1987): Centering of Mental Maps of the World. *Discussion Paper Series* 87–7. University of Arizona, 47 s.
- SEČKOVÁ, L. (2007): Slovácko a Moravští Slováci – pokus o vymezení pojmu. *Univerzita Palackého, Olomouc*, 83 s.
- SIWEK, T., BOGDOVÁ, K. (2007): České kulturně historické regiony ve vědomí svých obyvatel. *Sociologický časopis*, 43, č. 4, s. 1039–1053.

- SIWEK, T., KAŇOK, J. (2000): Vědomí slezské identity v mentální mapě. *Spisy FF OU*, 136, Ostrava, 98 s.
- ŠERY, M., ŠÍMÁČEK, P. (2012): Perception of the historical border between Moravia and Silesia by residents of the Jeseník area as a partial aspect of their regional identity. *Moravian Geographical Reports*, 20, č. 2, s. 36–46.
- TUČEK, P., PÁSZTO, V., VOŽENÍLEK, V. (2009): Regular use of entropy for studying dissimilar geographical phenomena. *Geografie*, 114, č. 2, s. 117–129.
- TUČEK, P., JANOŠKA, Z. (2013): Fractal Dimension As Descriptor of Urban Growth Dynamics. *Neural Network World*, 23, č. 2, s. 93–102.
- TVERSKY, B. (2003): Structures of Mental Spaces: How People Think About Space. *Environment and Behaviour*, 35, s. 66–80.
- VOŽENÍLEK, V. (2002): Geoinformatická gramotnost: nezbytnost nebo nesmysl? *Geografie*, 107, č. 4, s. 371–382.
- VOŽENÍLEK, V. (2009): Artificial intelligence and GIS: mutual meeting and passing. *International Conference On Intelligent Networking And Collaborative Systems (INCOS 2009)*, s. 279–284.
- VOŽENÍLEK, V., KAŇOK, J. a kol. (2011): *Metody tematické kartografie – vizualizace prostorových dat*. Univerzita Palackého, Olomouc, 216 s.

S u m m a r y

INTERVAL DELIMITATION OF MEAN LINE OF GEOGRAPHICAL PHENOMENA FROM A SET OF POLYGONS: CASE STUDY OF MENTAL MAPS FOR HANÁ REGION

Mental map is a graphical representation of geographical space in person's mind and can be understood as a subjective picture of attractiveness of a space (Drbohlav 1991, Voženilek 2002, Pravda 2003, Tversky 2003). Mental maps can be used as a source of information especially in social sciences (Bláha, Hudeček 2010; Drbohlav 1991; Gould, White 2004; Pocock 1979). In Czechia, the mental maps had been repeatedly used to delimit ethnographic regions (Siwek, Kaňok 2000; Sečková 2007; Kyncl 2011). Commonly used approach is to perform a questionnaire survey, which consists of (among other) drawing a map with individuals perception of a given geographic region. These maps are then used to delimit the border of region. The most commonly used criterion was to define the region as an area, which had been marked by at least $n\%$ of respondents. This method is rather subjective and it is unclear why area marked by i.e. 50% of respondents should be considered a part of ethnographic region, while area marked by 49% of respondents should not.

Aim of this paper is to propose an innovative method of estimation of mean line, based on statistical estimation theory. While most of the recent studies include only point estimation of boundary, this paper introduces the interval estimates in each direction from the centre, hence allows to measure variability of boundary. The method has been demonstrated on both experimental and real data from Kyncl's study (2011).

Following algorithm was used to calculate the mean line:

- Step 1: set inner point C, which lies inside all, or at least most of the polygons.
- Step 2: in arbitrary number of directions, draw a line originating in C.
- Step 3: calculate distances d of intersections between lines and polygons in every direction.
- Step 4: calculate mean of distribution of distances in each direction.
- Step 5: connect neighbouring mean distances in each very direction to form the mean line.

Alternatively, the robust statistical estimates (median) can be used instead of the mean. The confidence intervals in each direction can be calculated and used to capture variability of estimate.

The proposed method can be apply to delimit the mean border line of any set of polygons, for statistical testing, however, there is an assumption of normality of distribution of distances in each direction. It is also assumed, that there is a centre of a region, and that a vast majority

of polygons overlaps. This is a true for most of the ethnographic regions, which usually have a historical or cultural centre. However this may not be true for another phenomenon.

Even in case the distribution of distances is not normal, the method is still useful in providing information about variability in each direction. Another advantage of this method is that the calculated mean line is smooth, which is often not the case, if it is delimited as an area, marked by certain percentage of respondents.

Erroneous results can be expected if the polygons are concave. In this case there may be more than one intersection in some of the directions. In such cases, the analyst should decide, based on his/her expert knowledge, if all intersections will be used; or only the first one/last one/mean of all; or if such polygons should be entirely removed from the analysis.

The proposed method is perspective in geography, because it enables theoretically sound estimation of mean line, a feature, which is common in variety of topics. The calculations with experimental data provide very promising results, since the data are generated by a well-known process and all theoretical assumptions are met. The attempts with real data – boundary of ethnographic region Haná – give results, which are comparable with previous studies. There are several reasons, why the proposed method does not perform better than heuristic approaches used in previous studies: the variability of the data is greater, there is not one evident centre (people from different parts of the region tend to centre the region in their hometown) and some of the polygons do not overlap.

Obr. 1 – Grafické znázornění pojmů použitých v algoritmu výpočtu střední linie. P – množina polygonů, C – vnitřní bod, p_i – polopřímka z vnitřního bodu C ve směru ϕ_i , protíná hranice polygonů P ve vzdálenostech d_i od středu C .

Obr. 2 – Experimentálně generovaná data – 100 elips (šedě), odhad střední linie (černě) a 95% interval spolehlivosti pro střední linii (čárkovaně šedě), šipkou jsou vyznačeny směry, v nichž byla testována normalita.

Obr. 3 – Srovnání způsobů vymezení regionu Haná. V legendě: střední linie, interval spolehlivosti, mediánová linie, obálka spolehlivosti 95 %. Zdroj podkladových dat: ČÚZK 2013.

Obr. 4 – Průběh střední linie regionu Haná podle bydliště respondentů. V legendě: okres bydliště respondenta. Zdroj podkladových dat: ČÚZK 2013.

Pracoviště autorů: Univerzita Palackého v Olomouci, Přírodovědecká fakulta, katedra geoinformatiky, 17. listopadu 50, 771 46 Olomouc; e-mail: zbynek.janoska@centrum.cz, vit.vozenilek@upol.cz, pavel.tucek@upol.cz.

Do redakce došlo 15. 5. 2013; do tisku bylo přijato 8. 1. 2014.

Citační vzor:

JANOŠKA, Z., VOŽENÍLEK, V., TUČEK, P. (2014): Intervalové vymezení průběhu střední linie geografických jevů ze sady polygonů: příklad vymezení regionu Haná z mentálních map. *Geografie*, 119, č. 1, s. 91–104.