

PAVEL TUČEK, VÍT PÁSZTO, VÍT VOŽENÍLEK

POUŽITÍ ENTROPIE PŘI STUDIU NESTEJNORODOSTI GEOGRAFICKÝCH JEVŮ

P. Tuček, V. Pászto, V. Voženílek: *Regular use of entropy for studying dissimilar geographical phenomena*. – Geografie–Sborník ČGS, 114, 2, pp. 117–129 (2009). – The concept of entropy connected with GIS is relatively new. Its mathematical background was defined after World War II by Claude E. Shannon, well-known mathematician, electronic engineer and founder of information theory. Information theory deals with entropy as measure of information which every single message has. And thus entropy quantifies the amount of information in a message. The paper is based on entropy applications in cartography and demonstrates its usage as a measure of information in GIS. The authors provide an algorithm for setting number of intervals in thematic maps with using entropy calculations. Finally, the obtained knowledge is applied to sample datasets for creating climatic maps within GIS environment.

KEY WORDS: entropy – GIS – thematic maps – information rate.

Príspevek byl zpracován v rámci řešení projektu GA ČR 205/06/0965 „Vizualizace, interpretace a percepce prostorových informací v tematických mapách“.

1. Úvod

Pojem informační entropie prezentoval po druhé světové válce Claude E. Shannon, významný matematik, elektronický inženýr a zakladatel teorie informace. S entropií, jako mírou informace, kterou má jakékoliv sdělení, se pracuje v teorii informace. Míra entropie kvantifikuje množství informace ve zprávě. Takovou zprávou je i mapa, dokument sdělující velké množství prostorových informací rychlým a přesným způsobem (Voženílek 2005). Mapa ve smyslu předmětu studia kartografické informatiky (Koláčný 1969) je součástí výzkumných metod prostorových jevů, a tím i geografického výzkumu. Prostorové informace reprezentující geografické poznatky jsou do mapy kódovány v procesu kartografické tvorby použitím grafických proměnných kartografických znaků (MacEachren, Kraak 2001) a poté dekódovány uživatelem s cílem poznání prostorových objektů a procesů.

Cílem článku je jednak seznámit odbornou veřejnost s entropií v geografii, tedy i v kartografii a GIS, ale také apelovat na možnosti použití entropie v těchto oborech. Díky vývoji geoinformatiky je k dispozici mnoho analytických nástrojů, které dokáží funkci entropie zapracovat do studia geografických jevů, resp. jejich znázornění v mapě. Na několika ukázkových příkladech je naznačena aplikace entropie, která díky širokému využití nemůže rozhodně být kompletní. Nicméně je důležité podotknout, že problematika entropie ve spojení s geografii, kartografií a GIS je kontinuálně řešena na katedře geoinformatiky Přírodovědecké fakulty Univerzity Palackého v Olomouci. V textu

jsou uvedeny základní pojmy teorie informace, dále pak výsledky výzkumných projektů.

2. Teorie informace

Vznik teorie informace je přisuzován C. E. Shannonovi, jenž po druhé světové válce poprvé zformuloval pojem matematizace informace (Shannon 1948). Matematická teorie informace popisuje informaci pomocí matematického aparátu a podle Kučerové (2006) zkoumá matematickou reprezentaci podmínek a parametrů ovlivňujících přenos a zpracování informací. S tím souvisejí pojmy jako zpráva, symboly, abeceda, signál a kódování (viz tab. 1).

Existují také další vymezení pojmu informace, a to z různých hledisek – filozofického, komunikačního, kybernetického, biologického, laického aj. Obecně lze konstatovat, že informace je jakékoliv sdělení či údaj o okolním reálném světě, procesech a jevech v něm. Informace omezuje nebo odstraňuje nejistotu či nevědomost o daném jevu z dané množiny jevů (upraveno podle Kučerové 2006).

Informace je tedy sdělení, které jedinci objasní určitý problém a sníží jeho nevědomost. Nicméně příjemcem informace nemusí být pouze člověk. Tím je myšleno, že odesílatel (stejně tak i příjemcem) zprávy nesoucí informaci může být kniha, mapa, počítač či jiné médium. S rozvojem počítačů se začalo uvažovat o kvantifikaci informace, tedy míře informace. Je tedy důležité zjistit relevantnost informace v požadovaném kontextu. Jinými slovy určit zda, a jak moc, je pro příjemce informace užitečná a pomůže mu odstranit neurčitost či nevědomost o daném jevu. Míra informace je funkcí pravděpodobnosti určitého sdělení pro příjemce (Brillouin 2004). Jako příklad nechť poslouží vyhledávání slov ve webových vyhledávacích. Ne vždy je totiž navrácen výsledek, jaký je v daném smyslu očekáván (Jaká bude tedy informační hodnota vrácené odpovědi?).

Dalším příkladem může být použití abecedy informace a posloupnost symbolů. Například slovo „prígl“ je pro obyvatele Prahy neznámé a nese pro něj minimum informace. Pro obyvatele Brna však toto slovo význam má. Znamená v brněnském slangu Brněnskou přehradu. A právě slangy, nářečí či cizí jazyky, ale i odborné terminologie, jsou typickým použitím odlišné abecedy. Jakou míru informace tedy získá obyvatel Prahy a Brna?

To samé platí s posloupností symbolů. Desetinné číslo 4,31 samo o sobě podstatnou informaci nenese. Avšak číslo z těch samých symbolů, ale v určitém

Tab. 1 – Základní pojmy teorie informace. Upraveno podle Pěchouček, Železný (2004)

Pojem	Význam	Příklad
Zpráva	Posloupnost rozlišitelných znaků	Petr je doma
Symbol	Rozlišitelný prvek ve zprávě (graficky – znak)	Podstatné jméno, sloveso, příslovce
Abeceda	Množina všech symbolů	Věta = {Petr; je; doma}
Signál	Nositel zprávy	Zvuk/řeč/text/...
Kódování	Transformace zprávy (z abecedy v jinou abecedu)	Peter is home
Informace	Vztah mezi symboly zprávy a okolním světem	Odpověď: „Petr je doma“

pořadí, informaci nese – 3,14; tedy hodnota π na dvě desetinná místa. Jak byla vydatná tato informace?

V kartografii se teorie informace nejvíce uplatňuje v kartografické sémiologii, a to zejména při sestavování kartografických znaků při respektování pravidel jazyka mapy (Pravda 1990, 2003; Voženílek 2005). Ve zjednodušené formě lze stejný význam informace demonstrovat na příkladech použití základních kompozičních prvků mapy (Kaňok 1990, Voženílek 2001).

Tyto otázky lze obrátit – jak moc byla odstraněna neurčitost? Tato neurčitost se obecně nazývá entropie. Informatický a matematický aparát umožňuje informaci měřit. Jak uvádí Komenda (1991), výše uvedené požadavky na míru neurčitosti situace respektuje veličina zvaná entropie.

3. Entropie

Pojem entropie od svého vzniku pronikl do mnoha vědních oborů a podobně jako informace má mnoho definic. Jak již bylo dříve uvedeno, entropie představuje určitou neurčitost systému před přijetím informace. Podle Komendy (1991) je množství neurčitosti spojené s danou situací zároveň množstvím informace, kterou je potřeba dodat k odstranění této neurčitosti. Jinými slovy entropie určuje, kolik informace musí být dodáno k pochopení problému.

Jak uvádí Pechanec (2006) na příkladu člověka (jako příjemce informace), množství informace (a tedy i odstraněné entropie) může být malé, pokud je příjemcem člověk, který buď nemá zkušenost s danou problematikou, nebo naopak disponuje velikými zkušenostmi s danou problematikou. V prvním případě je pro tohoto příjemce daná informace zbytečná (není s to využít informaci). Ve druhém případě je pro příjemce také ta samá informace zbytečná (nepřináší pro něj nic nového). Ani v jednom případě není odstraněna entropie. Podle Pechance (2006) je ideálním příjemcem středně zkušený jedinec.

Entropii lze obecně definovat následovně (upraveno podle Shannon 1948: Entropie je střední hodnota míry informace k odstranění neurčitosti, která je dána konečným počtem vzájemně vylučujících se jevů. Entropie je míra informační vydatnosti pokusu. Zakladatel pojmu entropie C. E. Shannon definoval entropii matematicky následovně (Shannon 1948):

Pro systém s konečným počtem možných stavů $S \in \{s_1, s_2, \dots, s_n\}$, $n \leq \infty$ a pravděpodobnosti jejich výskytu $P(s_i)$ je informační entropie definována:

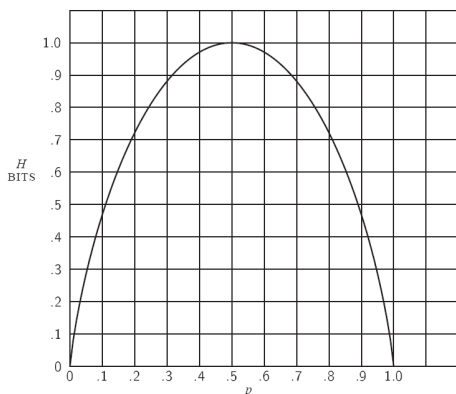
$$H(S) = - \sum_{i=1}^n P(s_i) \log_2 P(s_i) \quad (1)$$

Entropie je maximální, když jsou pravděpodobnosti výskytu vylučujících se stavů stejné (rozdělení je rovnoměrné): $H(S_{max}) = \log_2 n$, tedy právě tehdy, když

$$P(s_i) = \frac{1}{n} \quad \text{pro } \forall i$$

Entropie je minimální, pokud jsou všechny pravděpodobnosti $P(s_i)$ rovny nule, kromě jedné, jež nabývá hodnoty 1. Musí tedy platit: $H(S_{min}) = 0$ právě tehdy, když platí

$$\exists P(s_k) = 1 \quad \text{a} \quad P(s_i) = 0 \quad \text{pro } \forall i \neq k$$



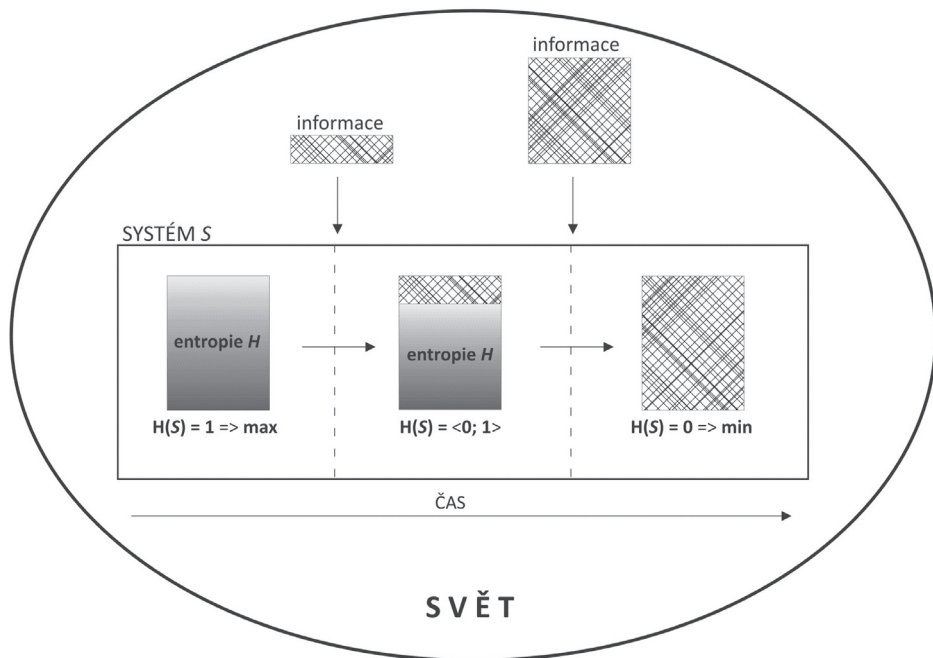
Obr. 1 – Entropie v případě dvou možností s pravděpodobnostmi p a $(1-p)$. Podle Shannona (1948).

Naopak minimální entropie nastane, jakmile jeden z výskytů získá 100% pravděpodobnost a všechny ostatní nulovou. Je jasné, která eventualita nastane, a není tedy potřeba dodávat informaci. Například při tenisovém zápase se hráč X zraní a vzdá souboj. Je jasné, že hráč Y vyhraje zápas kontumačně.

Oba výše uvedené příklady mají dvě alternativy. Obrázek 1 ukazuje entropii v případě dvou možností. Je z něj patrné, že se entropie zvyšuje s rostoucím vyrovnáváním možnosti/pravděpodobnosti výskytu obou eventualit, kdy v hodnotě $p = 0,5$ (tedy 50 %) je entropie maximální ($H = 1$). Naopak při převládání

Podle zvoleného logaritmu ve vztahu (1) se pak rozlišují jednotky entropie. Pokud se použije logaritmus o základu 2 (\log_2), tak jednotkou je bit. Za použití přirozeného logaritmu (\ln) nebo dekadického (\log) jsou jednotkami entropie nit, resp. dit.

Jak uvádí Shannon (1948) i Komenda (1991), maximální entropie nastává při vyrovnaných šancích všech výskytů souboru, tedy nastává nejvíce nejistá (neurčitá) situace. Je tak potřeba získat maximum informace pro odstranění této neurčitosti. Například při hodu mincí mohou nastat dvě situace – padne rub nebo líc. Obě možnosti mají pravděpodobnost 50 % – rovnoměrné rozdělení šancí.



Obr. 2 – Schéma procesu snižování entropie v systému

možnosti výskytu jedné či druhé eventuality entropie klesá, až nastane $H = 0$ a entropie je minimální (jedna z eventualit nastala).

Na obrázku 2 je schematicky znázorněno odstraňování entropie (neurčitosti) ze systému S . Uvedené platí i pro geografické systémy. Na počátku se v systému vyskytuje entropie rovna jedné. Systém je tedy maximálně neuspořádaný (neurčitý). Postupně je do systému dodávána informace, která neurčitost odstraňuje. Po doručení veškeré relevantní informace se entropie ze systému zcela vytrácí, klesá na hodnotu 0 a systém se stává plně uspořádaným.

Zatímco entropie diskrétní veličiny je absolutní mírou neurčitosti, ve spojitě verzi je entropie relativní mírou neurčitosti vzhledem ke zvolenému systému souřadnic (Pěchouček, Železný 2004).

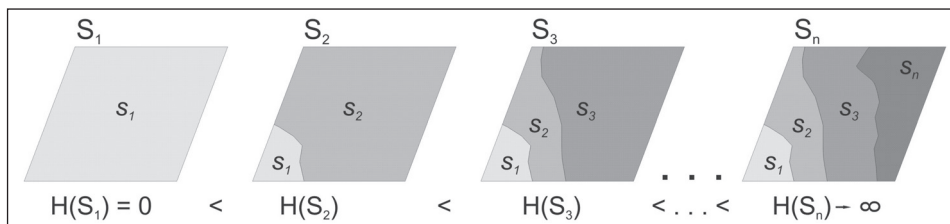
4. Příklady použití Entropie v geografii, kartografii a geoinformaticce

Murdych ve své práci (1988) popisuje využití entropie v kartografii. Analogicky lze entropii využít i v geoinformaticce (Voženílek 2005). Vůbec jedno z prvních použití entropie s implicitně zabudovanou prostorovou složkou (tzv. prostorová entropie) bylo aplikováno Battym (1974) při vymezení zón města. Stejně tak v geografii lze pomocí entropie hodnotit stupeň uspořádanosti a vzájemné shody prostorových jevů. S těmito možnostmi využití souvisí i aplikace entropie na stanovení optimálního počtu intervalů kvantitativních stupnic při znázorňování územního uspořádání geografických jevů (v závislosti na uživateli). Podobně lze výpočet entropie využít pro hodnocení závislosti geografických jevů. Výhodou použití entropie v geografii, kartografii a geoinformaticce je možnost jejího použití jak pro jevy kvantitativní (s číselnými charakteristikami), tak i pro jevy kvalitativní (biogeografické členění, mapy typů půd aj.).

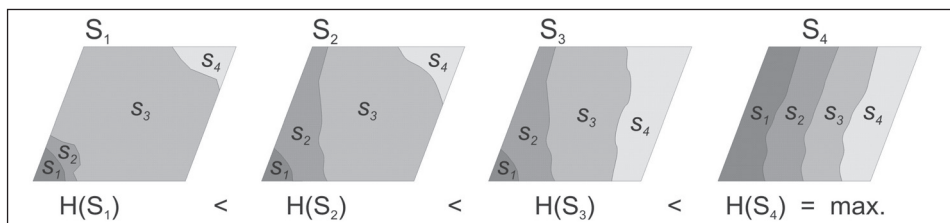
4.1. Hodnocení uspořádanosti geografických jevů

Uspořádanosti jevu se rozumí rozčlenění geografického jevu do jednotlivých kategorií. Vzhledem k charakteru vztahu (1) narůstá s rostoucím počtem kategorií i entropie. Pro názornost lze uvést následující příklad (upraveno podle Murdycha 1988): Mapa klimatické klasifikace znázorňuje n odlišných oblastí. Pokud je celé mapové pole mapy pokryto pouze jednou klimatickou oblastí, je území z pohledu klimatické klasifikace stejnorodé, tzn. že stupeň uspořádanosti klimatických oblastí je rovný 0, a entropie mapy je taky rovna 0 – z podstaty vztahu (1). Pokud se však zvýší počet odlišných klimatických oblastí (např. při změně měřítka či zvětšení zájmového území), pak narůstá i hodnota entropie a tedy i stupeň uspořádanosti geografického jevu. Klima v daném území, resp. mapa klimatických oblastí, již není stejnorodá, nýbrž více a více diferenciována (viz obr. 3).

Pokud má mapa konstantní počet ploch a zkoumá se podíl jednotlivých klimatických oblastí, pak nestejnorodost geografického jevu, resp. mapy, dosahuje maxima, jsou-li podíly jednotlivých ploch v mapě stejné. Toto opět plyne ze vztahu (1), kdy stejné podíly ploch v podstatě představují stejné pravděpodobnosti jejich výskytu v mapě, a tedy i maximální hodnotu entropie. Naopak minimální možná entropie, tedy maximální možná stejnorodost mapy, nastává, pokud podíl právě a pouze jedné plochy dosahuje maxima. To znamená, že podíl jedné klimatické oblasti natolik převyšuje podíl ostatních oblastí, které



Obr. 3 – Diferenciace mapy v souvislosti s entropií (s_1 až s_n jsou plochy klimatických oblastí, $H(S_n)$ je entropie klimatu na jednotlivých mapách)



Obr. 4 – (Ne)stejnorodost mapy v souvislosti s entropií (s_1 až s_4 jsou podíly jednotlivých ploch, $H(S_n)$ je entropie jednotlivých map)

však – narozdíl od předchozího příkladu – jsou stále zastoupeny. Je tedy zřejmé, že v tomto případě nemůže být entropie rovna 0 a ani nemůže nekonečně růst (viz obr. 4).

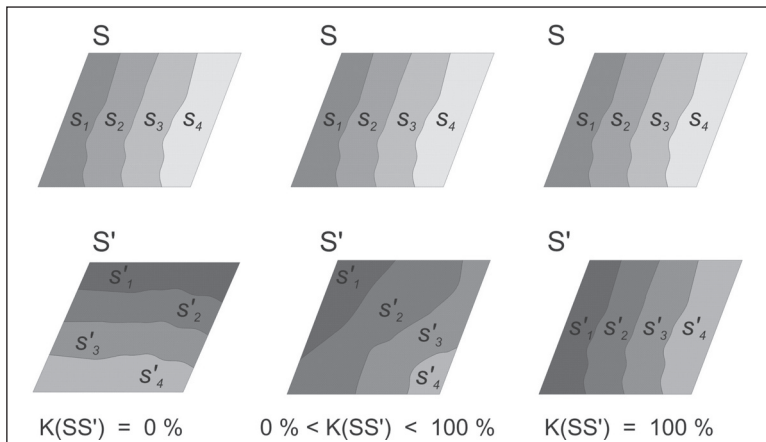
Pro hodnocení stupně uspořádanosti jevů se používá relativní entropie, jež se získá podílem absolutní entropie k maximální možné entropii. Právě relativní entropie najde využití při hodnocení nestejnorodosti kartografického znázornění geografického jevu v různých částech mapy při stejném počtu areálů.

4.2. Hodnocení vzájemné shody geografických jevů

Entropii lze rovněž využít k hodnocení závislosti dvou geografických jevů. Jsou-li oba jevy prostorově nezávislé, jejich společná entropie je rovna pouze součtu jednotlivých entropií. Nicméně řada geografických jevů je určitým stupněm na sobě závislá tak, že se některé areály obou map překrývají nebo splývají. Entropie této shody dvou korelovaných prvků (mapy S a S') se vypočítá užitím tzv. sdružené entropie

$$H(S, S') = - \sum_{i=1}^s \sum_{j=1}^r P(s_i, s'_j) \log_2 P(s_i, s'_j)$$

kde $P(s_i, s'_j) = P_{i,j}$ je sdružená pravděpodobnost (upraveno podle Pěchouček, Železný 2004). Pomocí sdružené entropie a prostého součtu jednotlivých entropií lze vypočítat koeficient vzájemné shody, jež se udává v procentech. Hodnoty koeficientu se pohybují od 0 do 100 %, kdy 0 % značí neshodnost jevů a 100 % úplnou shodu jevů (viz obr. 5).



Obr. 5 – Koefficient shody $K(SS')$ dvou geografických jevů S a S'

4.3. Stanovení optimálního počtu intervalů pro znázornění geografického jevu

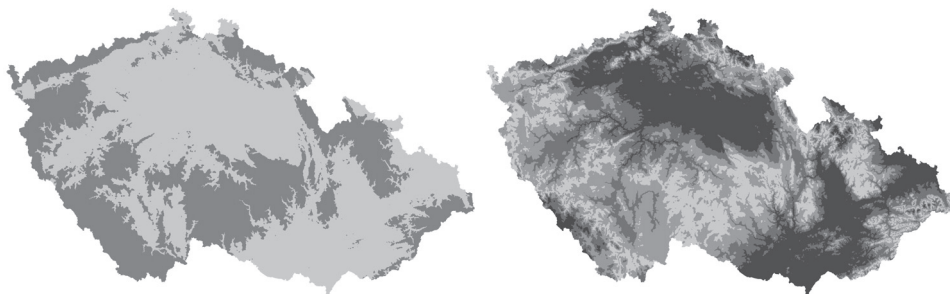
Optimálním počtem intervalů stupnice vyjádření kvantitativní charakteristiky jevu se rozumí počet intervalů vyšetřovaného jevu, který je podle teorie informace (podle výpočtu entropie) nejvhodnější pro zachování optimálního množství informace, jež vyšetřovaný jev poskytuje. Jinými slovy to znamená, na kolik intervalů je vhodné rozčlenit daný jev, aby bylo touto generalizací ztraceno co nejméně relevantní informace a aby dané rozdělení bylo v maximální možné míře informačně vydatné (Lin 1999).

Následující příklad demonstruje stanovení optimálního počtu intervalů pro rastrovou vrstvu (grid) nadmořských výšek České republiky. Pro výpočty byla použita data z Atlasu podnebí Česka (Tolasz a kol. 2007). Použitá datová rastrová vrstva má vertikální rozlišení jednotlivých hodnot pixelů 1 metr (prostorové rozlišení v tomto případě není důležité). Nejprve byly vypočítány entropie pro vrstvu rozdělenou do různých počtů intervalů (obr. 6 a obr. 7):

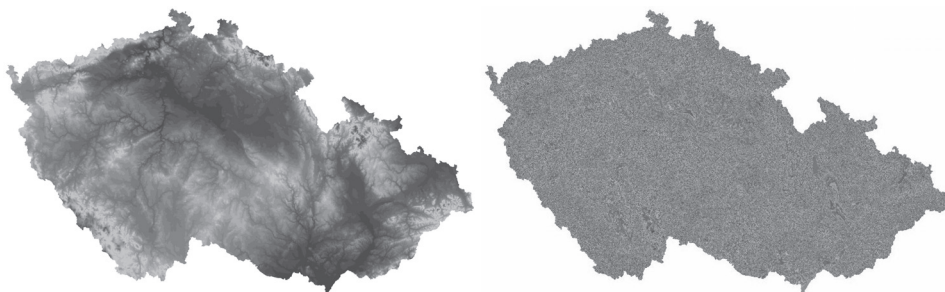
- grid se dvěma kategoriemi/intervaly ($I = 2$) – binární vyjádření
- grid se sedmi intervaly ($I = 7$) – vhodné kartografické rozdělení
- grid se 26 intervaly ($I = 26$) – intervaly stanovené na základě nalezeného inflexního bodu, který je zároveň minimem průběhu relativní entropie na zvolených datech)
- grid se 1337 intervaly ($I = 1337$) – po 1 metru výškovém.

Pro zjištění optimálního počtu intervalů na mapě byla použita relativní entropie. Vzhledem k užití modifikovaného vztahu pro výpočet značily nízké hodnoty relativní entropie ideální rozdělení pro zachování optimální informace. Naopak vyšší hodnoty relativní entropie ukazovaly na špatné rozdělení počtu intervalů vzhledem k nízkému či naopak vysokému (až nadbytečnému) množství informace, kterou čtenář mapy nedokáže efektivně využít.

Rozdělení gridu nadmořských výšek pouze do dvou intervalů (obr. 6 vlevo) se dá zjednodušeně charakterizovat rozdělením na nízké a vysoké polohy. Při rozdělení do sedmi kategorií (obr. 7 vpravo) je patrné, jak se informační hodnota gridu zvyšuje, a je možné rozeznat například nížinné oblasti, předhůří jednotlivých pohoří i samotné nejvyšší partie hor. Rozdělení do 26 kategorií



Obr. 6 – Grid nadmořských výšek rozdělený do dvou (vlevo) a sedmi kategorií (vpravo)



Obr. 7 – Grid nadmořských výšek rozdělený do 26 (vlevo) a 1 337 kategorií (vpravo)

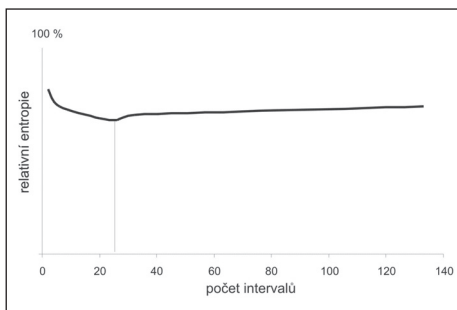
(obr. 7 vlevo) již ukazuje, jak se mapa podílí na interpretaci vlastního geografického jevu. Při detailnějším pohledu lze již rozeznat jednotlivé geomorfologické jednotky a množství informace obsažené v této mapě se opět zvětšilo. Jako poslední (obr. 7 vpravo) je znázorněna extrémní situace, kdy byl grid nadmořských výšek rozdělen do 1337 intervalů po jednom výškovém metru. Množství informace, které mapa tohoto gridu obsahuje, je maximální. Nicméně je potřeba rozlišit množství využitelné informace o geografickém jevu od celkového množství informace v mapě. Přes maximální podrobnost gridu není mapa pro uživatele interpretovatelná. Jedním z možných využití tohoto gridu může být detekce ploch se stejnou nadmořskou výškou (vodní plochy, plošiny, terasy, nivy apod.).

Po výpočtech relativní entropie u všech čtyř mapových vyjádření jednoho gridu bylo zjištěno, že největší informační hodnotu, tzn. nejvíce relevantní informace, udává grid s rozdělením na 26 intervalů. Ostatní gridy vykazují vyšší relativní entropii, takže podle teorie informace nezachovávají optimální množství využitelné informace. Přibližný průběh křivky relativní entropie ukazuje obrázek 8.

Lze tedy konstatovat, že pro grid nadmořských výšek je neoptimálnější použití rozdělení nadmořských výšek do 26 intervalů. Ovšem dalším důležitým faktorem ovlivňujícím rozdělení na intervalů je cílová skupina, která bude grid/obrázek/mapu číst. Například pro žáka prvního stupně základní školy bude grid se 2 intervaly (obr. 6 vlevo), žáci vyšších stupňů více rozumějí gridu se 7 intervaly (obr. 6 vpravo) a pro zkušenějšího geografa bude nejvhodnější grid s 26 intervaly (obr. 7 vlevo). Využití plnohodnotně grid nadmořských výšek s rozdělením na 1337 intervalů, resp. „hrubý“ grid (obr. 7 vpravo), dokáže pouze počítač při provádění GIS analýz.

S výše uvedeným tvrzením rovněž souvisí forma prezentace gridu či obrázku. Do toho vstupují i odborníci ostatních oborů (kartografové, meteorologové, biologové, lékaři, aj.). Pro kartografické znázornění na papíře či v atlase není vhodný vypočtený optimální grid s 26 intervaly, nýbrž je nutné tento počet zredukovat podle kartografických pravidel (např. Voženílek 2001). Při zobrazení v GIS, na obrazovce monitoru s možností přiblížení či oddálení, je vhodné použít grid s rozdělením na 26 intervalů (v závislosti na řešené úloze). Z těchto důvodů by bylo vhodné přizvat k řešení problematiky kartografa a psychologa, který by dokázal alespoň přibližně stanovit, co je pro cílovou skupinu optimum využitelné informace.

Jedním z důležitých faktorů míry entropie je vliv charakteru zkoumaného geografického jevu na jeho rozdělení do intervalů, stejně tak, jako použití statistických funkcí rozdělování do intervalů, jež nabízí GIS software (rozdělení podle Jenkse, přirozené intervaly, ekvivalentní intervaly, intervaly podle směrodatné odchylky a další). Charakter geografického jevu může také určovat, mají-li být v mapě zobrazeny extrémy jevu, nebo mají-li být naopak potlačeny. Vzhledem k charakteru digitálních dat je také potřeba rozeznat reálné extrémy od „outlierů“ (mimo statistický soubor ležících údajů), které mohou vzniknout např. nesprávným sběrem dat, chybou ve zpracování dat, náhodnou chybou aj. (Voženílek, Kaňok, Tuček 2008).



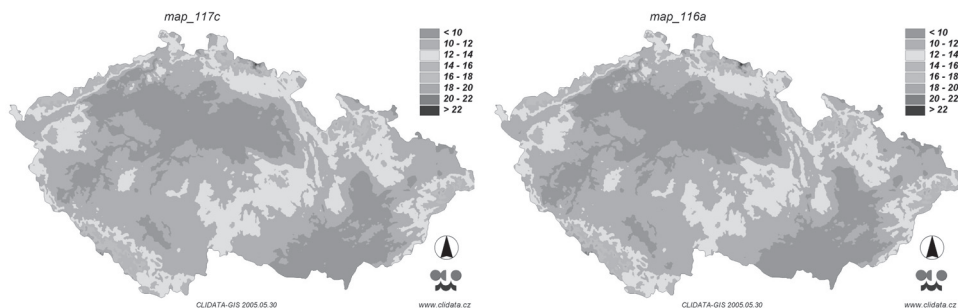
Obr. 8 – Průběh křivky relativní entropie vyšetřovaných gridů nadmořských výšek Česka

4.4. Studium vzájemné závislosti geografických jevů

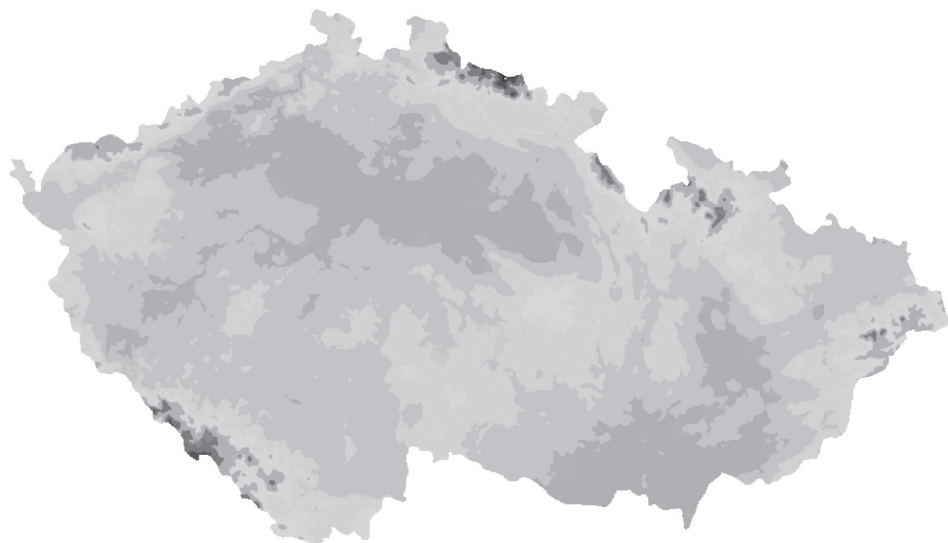
Pomocí entropie lze studovat vzájemné prostorové závislosti geografických jevů odlišnou metodou. Tato metoda spočívá v tom, že se pro dva geografické jevy vypočítá relativní entropie, podle které se určí vhodné rozdělení intervalů. Tyto dva gridy se následně mezi sebou vynásobí a podle počtu nově vzniklých jedinečných intervalů lze určit míru souvislosti, resp. závislosti gridů, tedy i znázorňovaných jevů.

Následující příklad ukazuje způsob použití této metody. Příklad je zjednodušen tím, že počet intervalů (včetně jejich mezních hodnot) je předem dán. Použity byly dvě gridové vrstvy (obr. 9) z Atlasu podnebí Česka (Tolasz a kol. 2007), které znázorňují vybrané klimatické charakteristiky. Prostorová závislost těchto charakteristik může být zkoumána bez vědomí, co který grid představuje. Závěry mohou být formulovány až po provedení výpočtů.

Nechť jsou dány gridy Grid A a Grid B s definovanými intervaly rozdělení, a to následovně: do 10–12–14–16–18–20–22 a více. Reklasifikací rastrového souboru gridu bylo vytvořeno 8 kategorií pro každý grid. Následně byly vypočteny absolutní, maximální a relativní entropie pro každý grid samostatně i pro grid po jejich vynásobení (obr. 10 a hodnoty v tabulce 2). Byla zjištěna rovnoměrnost rozdělení hodnot v jednotlivých kategoriích (pouze pro srovnání). Pro výpočet byl použit přirozený logaritmus (jednotka nit).



Obr. 9 – Gridová vrstva A (Grid A) a gridová vrstva B (Grid B)



Obr. 10 – Výsledný grid vzniklý násobením Gridu A a B

Z tabulky 2 je patrné, že rozdělení hodnot na základě výpočtu entropie rastrov Grid A a Grid B je velice podobné, nicméně v rámci těchto jednotlivých vrstev relativně nerovnoměrné. Lze vysledovat menší rozptýlení hodnot, tedy vyšší koncentraci hodnot v určité jedné kategorii. To samé lze tvrdit i pro spojenou vrstvu (Grid A + Grid B), kde je rozptýlení hodnot ještě menší a ještě vyšší koncentrace v určitém jednom (či více, ale ne v mnoha) intervalu (kategorii).

Poté byl Grid A vynásoben rastrov Grid B za vzniku nového gridu s 21 kategoriemi (intervaly). Na základě tohoto sloučení (Grid A + Grid B), resp. na nově vytvořených kategoriích, lze usuzovat o podobnosti či závislosti jevů. Maximální počet nově vzniklých intervalů by byl 8×8 , tedy 64. V tom případě by byly vyšetřované klimatické charakteristiky na sobě nezávislé. Opačný případ by nastal, kdyby nově sloučená vrstva měla minimální počet kategorií – 8. V tomto případě by byly jevy shodné, tedy absolutně na sobě závislé.

Nicméně získaný počet intervalů (kategorií) byl 21, což tvoří asi 23 % rozsahu možných kategorií, jinými slovy tento počet intervalů spadá do prvního kvartilu rozdělení. Na základě těchto zjištění lze tvrdit, že vrstvy (a tedy

Tab. 2 – Entropie jednotlivých vrstev (hodnoty entropií v nit)

Vrstva: počet kategorií	Vypočtená entropie $H(X)$	Maximální entropie $H_{max}(X)$	Relativní entropie $H_{rel}(X)$
Grid A: 8	1,39	2,08	66,8 %
Grid B: 8	1,36	2,08	65,2 %
Grid A + Grid B: 21	1,58	3,05	51,8 %

jevy, které vyjadřují) jsou si spíše podobné a více se shodují, nežli by se lišily. Vzhledem ke skutečnosti, že vrstvy reprezentují průměrný měsíční počet dní se sněžením v Česku v březnu (Grid A), resp. v prosinci (Grid B), za období 1961–2000, je zjištění logické, přesto přínosné, protože kvantifikuje nestejnoro-
dost dat reprezentující zkoumané prostorové jevy. Při dalším výzkumu, ve kterém by byly vyšetřovány entropie dalších jevů (např. měsíční počet dní se sněžením v ostatních měsících, průměr sezónních úhrnů výšky nového sněhu apod.), lze získat poznatky o jejich nestejnoro-
dosti, resp. shodnosti.

5. Závěr

Entropie vyjadřuje míru informace, kterou v sobě obsahuje sdělení. Kartografie používá pro sdělování informací mapovou tvorbu, přičemž míra informace v mapách se dá zkoumat právě použitím entropie. Autoři ukázali, jak lze za pomoci různých přístupů a postupů využít vypočtenou hodnotu entropie k dosažení informativně hodnotného kartografického díla vyjadřující nestejnoro-
dost znázorňovaného prostorového jevu. Výše uvedené příklady jsou pouze fragmentem použití entropie pro studium geografických jevů, resp. jejich znázor-
nění na mapě. Tyto příklady slouží tedy spíše pro naznačení směru, jakým lze entropii v prostředí GIS použít. Zvláště pak metoda stanovení optimálního počtu intervalů pro znázornění libovolného geografického jevu má potenciál vnést objektivitu (ve smyslu exaktního stanovení počtu intervalů na základě teorie informace) do procesu určování počtu kategorií, na niž se velkou měrou podílí autor mapy, tedy osoba víceméně subjektivně se rozhodující, a to i přes veškeré vědecké znalosti. Lze tímto přístupem rozšířit řadu geografických studií, např. Daňhelka (2004); Dobrovolný, Kepřtová (2006) nebo Hudeček (2008). Kartograf ve spolupráci s odborníkem v tématu dokáže vizualizaci přizpůsobit cílové skupině uživatelů a při použití poznatků z teorie informace a výpočtu entropie také optimalizovat množství informace v mapě, jež předkládá uživateli.

V příspěvku uvedené metody a příklady mohou uvést do pohybu diskuzi o problematice informační entropie a v budoucnu vést i k definování nových zákonitostí a postupů, jak precizněji, exaktněji a pomocí GIS také daleko efektivněji studovat různorodé geografické jevy a objekty, potažmo jejich znázor-
nění v mapě.

Literatura:

- BATTY, M. (1974): Spatial Entropy. *Geographical Analysis*, 6, s. 1–31.
BRILLOUIN, L. (2004): *Science and Information Theory*. Dover Publications, 2nd edition, 368 s.
DANHELKA, J. (2004): August 2002 Flood in the Czech Republic: Meteorological Cause and Hydrological Response. *Geografie–Sborník ČGS*, 109, č. 2, s. 84–92.

- DOBROVOLNÝ, P., KEPRTOVÁ, K. (2006): Spatial analysis of damage caused by strong winds and gales in the Czech lands since ad 1500. *Geografie–Sborník ČGS*, 111, č. 1, s. 51–69.
- HUDEČEK, T. (2008): Model časové dostupnosti individuální automobilovou dopravou *Geografie–Sborník ČGS*, 113, č. 2, s. 140–153.
- KAŇOK, J. (1999): Tematická kartografie. Ostravská univerzita v Ostravě, Ostrava, 318 s.
- KOLÁČNÝ, A. (1969): Cartographic information: a Fundamentals concept and term in modern cartography. *Cartographic Journal*, 6, s. 47–49.
- KOMENDA, S. (1991): Základy statistiky ve zdravotnictví. Univerzita Palackého v Olomouci, Olomouc, 53 s.
- KUČEROVÁ, H. (2006): Definice informace: data – informace – znalosti, [www.<http://info.sks.cz/users/ku/UIS/inform1.htm>](http://info.sks.cz/users/ku/UIS/inform1.htm), cit. 20. 4. 2008.
- LIN, S.-K. (1999): Diversity and Entropy. *Entropy*, 1, č. 1, s. 1–3.
- MacEACHREN, A. M., KRAAK, M. J. (2001): Research challenges in Geovisualization. *Cartographic and Geographic Information System*, 28, č. 1, s. 3–12.
- MURDYCH, Z. (1988): Tematická kartografie. SPN, Praha, 248 s.
- PECHANEC, V. (2006): Nástroje podpory rozhodování v GIS. Univerzita Palackého v Olomouci, Olomouc, 104 s.
- PĚCHOUČEK, M., ŽELEZNÝ, F. (2004): 33KUI – Kybernetika a umělá Inteligence, [www.<http://cyber.felk.cvut.cz/gerstner/teaching/kui/>](http://cyber.felk.cvut.cz/gerstner/teaching/kui/), cit. 12. 4. 2008.
- PRAVDA, J. (1990): Základy koncepcie mapového jazyka. Geografický ústav SAV, Bratislava, 168 s.
- PRAVDA, J. (2003): Mapový jazyk. Univerzita Komenského v Bratislavě, Bratislava, 88 s.
- SHANNON, C. E. (1948): A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, s. 379–423, 623–656.
- SLOCUM, T. et al. (2008): Thematic Cartography and Geovisualization. Upper Saddle River, Prentice Hall Series in Geographic Information Science, 561 s.
- TOLASZ, R. a kol. (2007): Atlas podnebí Česka. Český hydrometeorologický ústav/Univerzita Palackého v Olomouci, Praha, Olomouc, 256 s.
- VOŽENÍLEK, V. (2001): Aplikovaná kartografie I – tematické mapy. Univerzita Palackého v Olomouci, Olomouc, 187 s.
- VOŽENÍLEK, V. (2005): Cartography for GIS – geovisualisation and map communication. Univerzita Palackého v Olomouci, Olomouc, 140 s.
- VOŽENÍLEK, V., KAŇOK, J., TUČEK, P. (2008): Detekce, prokazatelnost a vizualizace extrémů demografických dat ve statistických souborech. *Kartografické listy*, 16, s. 12–32.
- YOCKEY, H. P. (2005): Information Theory, Evolution, and the Origin of Life. Cambridge University Press, Cambridge, 242 s.

Summary

REGULAR USE OF ENTROPY FOR STUDYING DISSIMILAR GEOGRAPHICAL PHENOMENA

Information theory, entropy and its spatial form are widely used in geographical research. Hand in hand with development in informatics, geoinformatics and GIS must go forward as well. As it is known, geoinformatics deals with spatial information and it is easier to analyze this information using methods described in the information theory, i.e. entropy function and spatial entropy function.

Entropy was applied in many other fields of science (medicine, biology, chemistry, etc.) and throughout the time geography and geoinformatics have also been included. Spatial entropy, i.e. entropy with spatial aspects, was firstly used in the 1970s, and took into account spatial distribution of geo-phenomena.

The paper refers to basic applications of using entropy in cartography, when supported by geoinformatics and GIS. Robust analysis and visualization tools provided in GIS give a brand new meaning to entropy and enable studying and displaying geographical phenomena and processes in a new relationship as well as their expression via maps, atlases and etc.

It is possible to find information on setting the optimal amount of intervals of any spatial phenomenon in a map when using the entropy function. It is also possible to evaluate the orderliness of geographical phenomena or to evaluate the correspondence between two

phenomena shown in the map, and many more. And this all with the help of the entropy function.

The entropy function brings new possibilities in understanding nature around us. Further and deeper research in this problem is being carried on at the Department of Geoinformatics at Palacký University in Olomouc and new findings are awaited.

Fig. 1 – Entropy in the case of two possibilities with probabilities p and $(1-p)$. Source: Shannon (1948).

Fig. 2 – Schematic diagram of eliminating entropy. In scheme from above: information, system S , entropy H .

Fig. 3 – Map differentiation in context of entropy (s_1 to s_n are climate classification regions, $H(S_n)$ is the entropy of particular maps).

Fig. 4 – (In)-homogeneity of the map in context of entropy (s_1 to s_4 are climate areas quotients, $H(S_n)$ is the entropy of particular maps).

Fig. 5 – Conformity coefficient $K(SS')$ of two geographic phenomena S a S' .

Fig. 6 – Elevation grid divided into two (on the left) and seven categories (on the right).

Fig. 7 – Elevation grid divided into 36 (on the left) and 1,337 categories (on the right).

Fig. 8 – Behaviour of relative entropy waveform of the elevation grid of Czechia. Axis x – number of intervals, axis y – relative entropy.

Fig. 9 – Grid layer A (Grid A) and grid layer B (Grid B).

Fig. 10 – Final grid layer obtained by multiplying by grids A and B.

Pracoviště autorů: katedra geoinformatiky, Přírodovědecké fakulty Univerzita Palackého v Olomouci, tř. Svobody 26, 771 46 Olomouc; e-mail: pavel.tucek@upol.cz; vit_p@atlas.cz; vit.vozenilek@upol.cz.

Do redakce došlo 11. 1. 2009