

Shlukovací analýza údajů z meteorologických stanic Ruska a zemí někdejšího Sovětského svazu

Ivan Farský

farsky@sci.ujep.cz

katedra geografie, Přírodovědecká fakulta UJEP v Ústí nad Labem

Ivan Farský: Cluster analysis of meteorological data stations of Russia and some-time Soviet Union.

This article brings on base of 19 meteorological stations from Russia and other surrounding states confrontation of Köppen climatic regionalisation and cluster analysis of other data from this 19 stations. Confrontation shows significant conformity. The problems come on in case of non typical neighbourhood of station area, it means in border position (Saratov) or climatic anomaly (Stavropol, Kaunas).

Key words: Meteorological station, cluster analysis, Köppen climatic regionalisation, comparison.

1 Úvod

Slovo matematika si řada lidí nedovede představit ve spojitosti s problematikou geografie, přesněji pak fyzické geografie. Přitom jedním ze společných rysů metodiky těchto „nematematických“ věd je také potřeba hodnotit zkoumanou skutečnost na základě nějaké vnitřní podobnosti a diferenciací. Posuzování nějaké podobnosti je věc diskuse a osobního náhledu, proto se začaly prosazovat pokusy o vyhodnocení a objektivizaci této podobnosti, spolu se stanovením hranic mezi těmito podobnými si shluky, v geografii tedy snaha např. o vymezení regionů, kde jistá podobnost panuje a jinde ne. Historií vývoje užití shlukovací analýzy se podrobněji zabývá LUKASOVÁ, ŠARMANOVÁ (1985). Hledalo se jak označit některé jevy a jak je seskupit s jistou podobností a určit, kdy tato podobnost již nenastává (regionalizační kritérium). Tak se objevují v matematice metody, které dnes jsou nazývány shlukovací (shluková) analýza – cluster analysis. První práce se na toto téma objevují již před druhou světovou válkou. Problémem byla velká pracnost s nejistým výsledkem. Zvrat nastal v době, kdy bylo možné jednoduše použít výpočetní techniku, která byla schopna velké množství početních výkonů provést v krátkém čase a výzkumníkům dát možnost rychlé interpretace výsledku. Tyto metody nejsou žádným „zázrakem“ a někdy mohou vést k výsledku velmi nepřirozenému. Proto je nutné nebrat je absolutně, stane se, že se nedají ani logicky použít. Záleží vždy na vložených datech. Metody shlukovací analýzy se objevují pro potřebu taxonomie nejprve ve vědách jako biologie, později i ve společenských vědách a v geografii se s nimi setkáme nejčastěji v ekonomické geografii. Fyzická geografie zatím těchto metod pro potřebu třídění a regionalizace příliš nepoužívá.

2 Procedury shlukovací analýzy

Shlukovací analýza se provádí na množině objektů, kde každý je definován prostřednictvím stejné množiny znaků, které má smysl zjišťovat. Shlukování lze provádět i na množině znaků, které jsou reprezentovány objekty – nositeli těchto znaků. Shluková

analýza je hlavně prostředek vytváření hypotéz o klasifikaci, které je pak nutné konfrontovat se skutečností a vyloučit nepravděpodobné stavy.

Současná literatura předkládá mnoho metod shlukování. Také výpočetní technika má ve svých programech řadu možností. Je to grafická podoba obsahující výsledek hledání podobnosti jednotlivých členů zkoumané množiny a uživateli neznalému přesné matematické struktury výpočtu ukazuje, jak jsou jednotlivé členy přiřazeny do podobnostních shluků. Stupeň shlukování pak v sobě obsahuje podstupně, které reprezentují vnitřní stavbu shluku. Je pak na zkušenosti pracovníka, aby určil, na kterém stupni definovat regionalizační hranici, co je region a co je již jeho vnitřní diferenciací. V tom je práce uživatele nezastupitelná. Shlukovací analýza pracuje na principu vytvoření tzv. shluků (klastr, cluster), které tvoří skupiny objektů uvnitř nějaké větší skupiny (množiny). V našem případě to je skupina meteorologických stanic. Tyto shluky nemají nahodilý ani rovnoměrný výskyt. Vzájemná vzdálenost těchto shluků (nepodobnost) je menší než vzdálenost objektů jiných shluků (MARHOLD, SUDA 2002). Shlukovacích procedur je řada a záleží na tom, jak je objekt shlukování definován. V našem případě jde o shlukování vyjádřené dendrogramem. Je to grafická podoba obsahující výsledek hledání podobnosti jednotlivých členů zkoumané množiny a uživateli neznalému přesné matematické struktury výpočtu ukazuje, jak jsou jednotlivé členy přiřazeny do podobnostních shluků. Základní shlukovací procedura spočívá v tom, že se zjišťují dva objekty, které mají k sobě nejbližší (jsou si nejvíce podobné), takto vznikne nový objekt a k němu se hledá další nejpodobnější. Celý proces končí, když jsou všechny objekty zapojeny do některého shluku. Ke zjištění vzdálenosti jednotlivých objektů se používá několik postupů, například metoda nejbližšího souseda (single linkage) nebo nejbližšího souseda (complete linkage), průměrné vzdálenosti (average linkage) a další. Pro kvantitativní data ve stanicích se pro výpočet vzdáleností používá Eukleidovská vzdálenost nebo vzdálenost Manhattan a další, které program nabízí (STATISTICA, SPSS aj.). Celý proces výpočtu probíhá v prostoru vícerozměrném. Záleží na tom, kolik proměnných objekt obsahuje. Čím je proměnných víc, tím je proces zjišťování blízkosti (shody) nebo vzdálenosti přesnější a ve skutečnosti taktéž pravděpodobnější. (LUKASOVÁ, ŠARMANOVÁ 1985). Podrobný popis matematické stránky věci nalezne čtenář v níže uvedené literatuře.

3 Definice problému

Jedním z úkolů práce geografa je problematika vymezení (vyčleňování) teritoriálních (akvatoriálních) jednotek, které se vyznačují nějakým společným znakem, což je odlišné od ostatního prostoru, kde definovaný znak neplatí a nebo se vyskytuje jen ve velmi malé intenzitě. Zmíněný znak je nazván v učebnicích regionalizačním kritériem.

V tomto prvním kroku jde o proces, avšak v konečném provedení jde o výsledek tohoto procesu, což je možné znázornit například tématickou mapou. Regionalizační kritérium lze definovat ze všech sfér přírodního prostředí a výsledkem jsou základní regionalizace obsažené příkladně v Národním atlase nebo ve velkých atlasech pro celý svět (Kol. 2000).

Za účelem provedení praktického příkladu s pomocí shlukovací analýzy byly nahodile vybrány meteorologické stanice v Rusku a v zemích, které v minulosti byly součástí Sovětského svazu. Důvodem takto vymezeného prostoru, není vzpomínka na minulost, nýbrž získání dat pro analýzu z níže zmíněné publikace. Jsou to hodnoty, uvedených

stanic, dlouhodobých normálů pro maximální a minimální teplotu v měsících leden, duben, červenec, září a dále pak pro průměrné dlouhodobé úhrny srážek v těch že měsících, doplněných o roční úhrn, ze kterých byla sestavena tabulka. Údaje pocházejí z publikace: LUDEN, F. a kol.(1990), Zde jsou hodnoty teploty uvedeny ve stupních Farenheita a úhrny srážek v palcích. Zmíněná publikace obsahuje několik stovek stanic po celém světě.

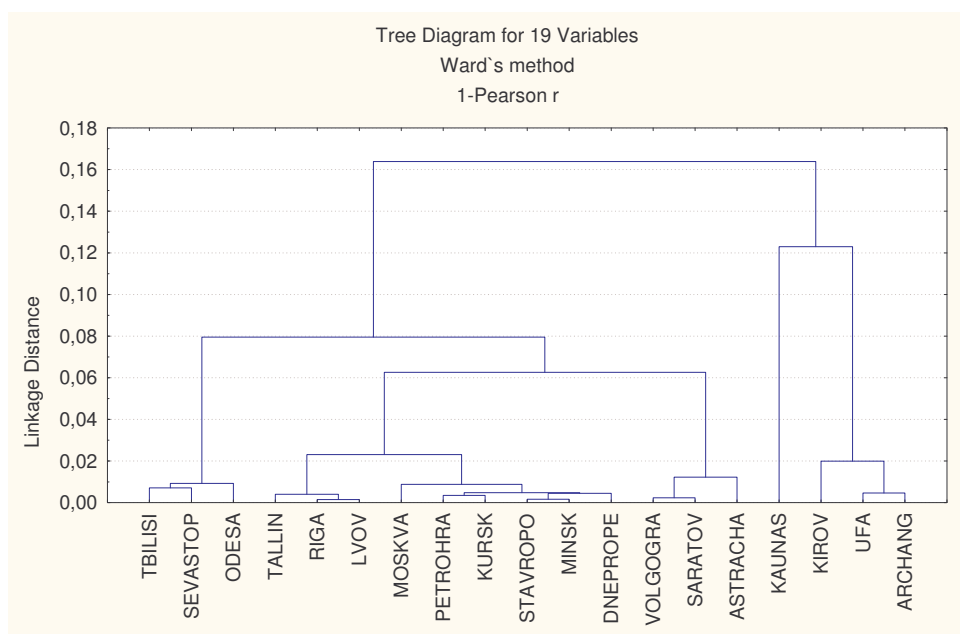
Tab.1: Příklad hodnot obsažených v publikaci a zpracovaných shlukovací analýzou

Stanice	teplota (°F)				srážky (palce)								
	max.		min.		měsíční průměr								
	leden,	duben,	červenec,	září	leden,	duben,	červenec,	září,	rok				
Saratov	15	7	55	35	82	64	48	36	1,0	1,0	1,2	1,4	14,5
atd.													

Zdroj: LUDEN, F. a kol.(1990)

Tabulka sestává ze 19 stanic, které tvoří sloupce obsahující 13 údajů v řádcích (viz příklad). Vložením dat a výpočtem programem STATISTIKA vychází seskupení, které zobrazuje shlukovací diagram na obrázku č. 1.

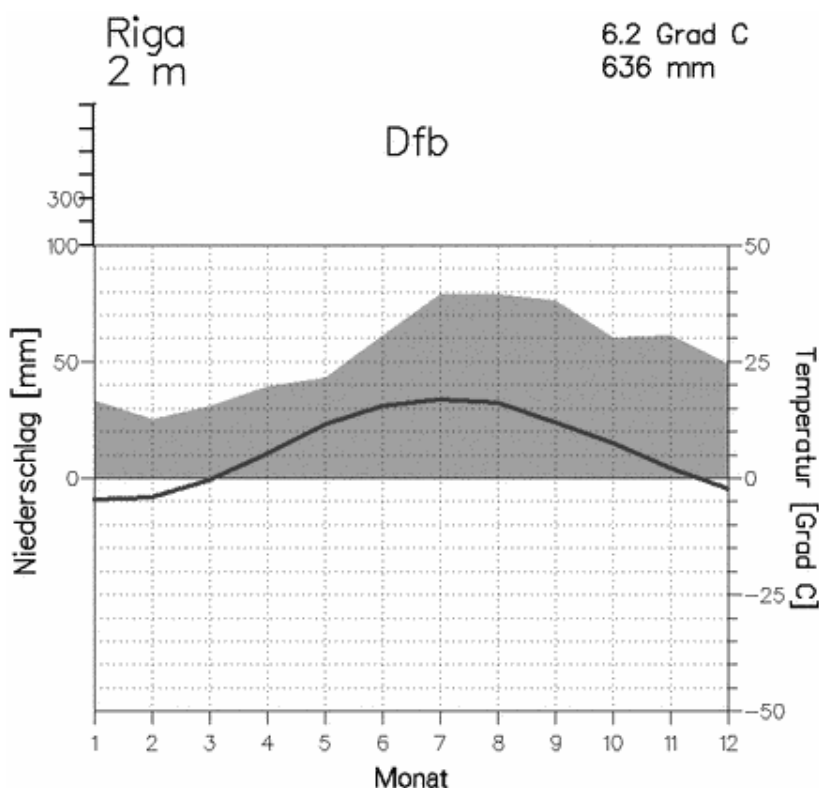
Obr. 1: Příklad shlukovacího diagramu pro stanice v Rusku a okolních zemích



Jsou zde patrné dva základní shluky. **První základní shluk** tvoří stanice Archangelsk, Ufa, Kirov a mimo stojící Kaunas. **Druhý základní shluk** je již diferencován na 3 základní skupiny. Jsou tu stanice Volgograd, Astrachaň a Saratov v jednom shluku, blízko kterého je shluk dvou shluků stanic, které mají k sobě poměrně blízko. Jsou to Tallinn, Riga a Lvov, ke kterým se blíží druhý shluk stanic Dněpropetrovsk, Minsk, Stavropol, Kursk s Petrohradem a již poněkud se vzdalující Moskvou. Jednoznačně je vymezen poslední shluk tvořený stanicemi Tbilisi, Sevastopol a Odesa. Tyto shluky jsou reálně možné a nahlédnutím do mapy klimatických oblastí a porovnáním s klasifikací dle

www.klimadiagramme.de, se shodují s provedenou regionalizací. Příklad diagramu je na obrázku 2.

Obr. 2: Klimatický diagram pro stanici Riga



Zdroj: <http://www.klimadiagramme.de>

Zařazení stanic dle *www.klimadiagramm* do klimatických oblastí je v tabulce 2.

Tab. 2: Zařazení stanic do shluků a jejich identifikace s Köppenovou klasifikací

Stanice ve shlucích	Zařazení do klimatické oblasti	Poznámka
První hlavní shluk Kirov, Ufa, <i>Kaunas</i> , <i>Archangelsk</i>	Dfb, Dfc	Archangelsk leží již za hranicí v příbuzné oblasti
Druhý hlavní shluk Moskva, Petrohrad, Kursk, Minsk Dněpropetrovsk, <i>Stavropol</i>	Dfb, BSk	Stavropol je sice v oblasti BSk, ale nachází se ve vyšší poloze
Druhý hlavní shluk, detailní části Lvov, Riga Tallinn	Dfb	Jednoznačné určení
Druhý hlavní shluk, vzdálenější součást Volgograd, Astrachaň, <i>Saratov</i>	BSk, Dfb	Saratov leží v přechodné zóně na hranici obou klim. oblastí
Vzdálená součást druhého shluku Odesa, Sevastopol, Tbilisi	Cfa	Jednoznačné určení

Vysvětlivky:

B – suché,

S – stepní,

k – chladné a suché, průměrná roční teplota pod 18 °C

D – mírně vlhké a s chladnou zimou,

f – celé období vlhké, průměrná teplota chladné sezony pod -3 °C,

b – průměrná teplota letních měsíců nedosahuje 22 °C.

c – krátké a chladné léto, průměrná teplota všech měsíců pod 22 °C, 1 až 3 měsíce s průměrnou teplotou nad 10 °C

C – teplé, vlhké,

f – celé období vlhké,

a – dlouhé horké léto, průměrná teplota nejteplejšího měsíce nad 22 °C, alespoň 4 měsíce s průměrnou teplotou nad 10 °C.

4 Interpretace shlukovací analýzy.

Zařazení stanic do jednotlivých shluků ukazuje výraznou shodu s Köppenovou klimatickou klasifikací (KOL. 2000). První shluk je ze všech nejdiferencovanější. Jednoznačně přiřazení mají stanice Kirov a Ufa. Stanice Archangelsk se nachází nepříliš daleko hranice podoblasti, proto je podobná a v tomto shluku. Pravděpodobné je ovlivnění od Atlantiku, avšak leží již výše na severu. Problém nastává ve shluku pouze jeden, a to Kaunas. Podle mapy patří sice do oblasti Dfb, ale nachází se blízko hranice s vedlejší oblastí Cfb. Navíc jeho poloha (stanice na letišti) na plochém, vyvýšeném terénu nad městem s větší možností ovlivnění od západu (Cfb) může znamenat jeho odlehlou polohu ve shluku.

Hlavní částí druhého shluku tvoří stanice Moskva, Petrohrad, Kursk, Stavropol, Minsk, Dněpropetrovsk. Zde je problémem stanice Stavropol. Je sice zařazena do BSk, ale opět její poloha ve Stavropolské vrchovině může být příčinou zařazení analýzou ze stepní oblasti, kde neleží, do vlhčí a chladnější v podhůří Kavkazu. Ve shluku jsou dvě odlehlejší části. Jednak jednoznačná, tvořená stanicemi Lvov, Riga a Tallinn. Zde je podobnost s předešlou evidentní z polohy stanic. Ty jsou jen víc posunuty k západu a tedy poněkud víc ovlivněné od Atlantiku. Druhá podskupina stanic Volgograd, Astrachaň a Saratov je opět vydělená prakticky jednoznačně do stepní oblasti. Problémem je stanice Saratov. Sice je značená již do skupiny Dfb, ale leží přímo na hranici oblastí, je možné se domnívat, že přesnější analýza by tuto stanici posunula do podskupiny BSk nebo kresba hranice na mapě je generalizovaná a tak vzniká tato nesrovnalost. Zcela mimo v tomto hlavním shluku stojí stanice Tbilisi, Sevastopol a Odesa. Je to pochopitelné, neboť tento kompaktní shluk představuje již jinou oblast Cfa.

5 Závěr

Výsledek shlukovací analýzy, provedený z jiných klimatických údajů než použil Köppen ke své klimatické klasifikaci, představuje téměř totožné umístění stanic do jednotlivých klimatických oblastí a podoblastí. Problémy se zařazením vznikly pouze v případech, že se stanice nacházela v netypickém prostředí pro celou oblast (podoblast), tedy v hraniční poloze (Saratov) nebo v klimatické anomálii (Stavropol, Kaunas).

Tento metodický příspěvek užívá metodu shlukovací analýzy na verifikovatelném příkladu. Použitá metoda není ve své podstatě nová. Ukazuje, že existují i jiné cesty řešení problému regionalizace v klimatologii a lze jí na základě ověřeného vzoru použít i v případech jiných, prvotního charakteru, s předpokladem úspěšného řešení.

Literatura

Kol. (2000): Philip's University Atlas. George Philip Limited, London. 304s.

Luden, F. a kol.(1990): The Water Encyclopedia. Geraghy and Miller Ground-Water, series Geography, Levis publishers New York. 807s.

Lukasová, S., Šarmanová, J., (1985): Metody shlukové analýzy, Academia, Praha. 212 s.

Marhold, K., Suda, J. (2002): Statistické zpracování mnohorozměrných dat v taxonomii.

Učební texty UK v Praze, Karolinum, Praha, 159 s.

<http://www.klimadiagramme.de/>