

**Univerzita Pardubice
Fakulta ekonomicko-správní**

**Analýza klasifikačních a rozhodovacích modelů ve vybrané
oblasti životního prostředí**

Bc. Marcela Geregová

Diplomová práce

2009

Univerzita Pardubice
Fakulta ekonomicko-správní
Ústav systémového inženýrství a informatiky
Akademický rok: 2008/2009

ZADÁNÍ DIPLOMOVÉ PRÁCE
(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: **Marcela GEREGOVÁ**
Studijní program: **N6209 Systémové inženýrství a informatika**
Studijní obor: **Informatika ve veřejné správě**

Název tématu: **Analýza klasifikačních a rozhodovacích modelů ve vybrané oblasti životního prostředí**

Z á s a d y p r o v y p r a c o v á n í :

- Předpokládá se, že v diplomové práci se bude řešit:
- problematika vybrané části životního prostředí;
 - problematika klasifikačních a rozhodovacích modelů;
 - sběr a příprava dat;
 - návrh a analýza vybraných modelů.

Rozsah grafických prací:

Rozsah pracovní zprávy:

Forma zpracování diplomové práce: tištěná/elektronická

Seznam odborné literatury:

FOTR, Jiří, DĚDINA, Jiří. Manažerské rozhodování. 1. vyd. Praha 4 : EKOPRESS, s.r.o., 1997. 207 s. ISBN 80-901991-7-8.

MAŘÍK, Vladimír etc.: Umělá inteligence 1. Praha : Academia, 1993. ISBN 80-200-0496-3.

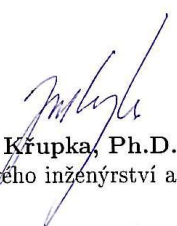
MAŘÍK, Vladimír etc.: Umělá inteligence 2. Praha : Academia, 1997. ISBN 80-200-0504-8.

CARLSON Christer, FULLĚR Robert. Fuzzy Reasoning in Decision Making and Optimization. New York : Physica-Verlag, 2002. ISBN 3-7908-1428-8.

CENIA, Česká informační agentura životního prostředí. Statistická ročenka životní prostředí České republiky 2007 [online]. 2007 [cit. 2008-05-09]. URL: http://www.cenia.cz/web/www/web-pub2.nsf/pid/CENMSFMVMTMNS/FILE/kap_99rej.pdf.

ZAPLATÍLEK Karel, DOŇAR Bohumil. MATLAB : tvorba uživatelských aplikací. 1. vyd. Praha : BEN - technická literatura, 2005. 216 s. ISBN 80-7300-133-0.

Vedoucí diplomové práce:



doc. Ing. Jiří Krupka, Ph.D.
Ústav systémového inženýrství a informatiky

Datum zadání diplomové práce:

6. října 2008


Termín odevzdání diplomové práce:

1. května 2009


doc. Ing. Renáta Myšková, Ph.D.

děkanka

L.S.


doc. Ing. Jiří Krupka, Ph.D.

vedoucí ústavu

V Pardubicích dne 6. října 2008

Prohlašuji:

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorský zákon, zejména ze skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně.

V Pardubicích dne 30. dubna 2009

Bc. Marcela Geregová

Poděkování

Na tomto místě bych chtěla poděkovat panu doc. Ing. Jiřímu Křupkovi Ph.D. za vedení diplomové práce, za půjčení materiálů potřebných ke zpracování této práce, za odbornou pomoc, za poskytnutí cenných rad a podporu při její vypracování.

Také bych ráda poděkovala svým rodičům za podporu a trpělivost v průběhu mého studia.

Anotace

Diplomová práce se zabývá analýzou klasifikačního modelu a rozhodovacích modelů. Zástupce rozhodovacího modelu je analytický hierarchický proces, zástupcem klasifikačních modelů jsou rozhodovací stromy a shluková analýza. Modelování je provedeno v programovém prostředí MATLAB, SPSS Clementine 11.1. Data použitá pro modelování jsou naměřené dešťové srážky a jejich chemické složení ve vybraných lokalitách. Pro přípravu dat je využito aplikace MS Excel 2007.

Klíčová slova

Životní prostředí, klasifikace, rozhodovací stromy, shluková analýza, rozhodování, vícekritériální rozhodování, analytický hierarchický proces, citlivostní analýza, rozhodovací pravidla.

Title

Analysis of classification and decision models in an environment area

Annotation

This thesis concerns with analysis of classification and decision models. Representative of impressionism of decision models is Analytic hierarchy process, representative of impressionism of classification models is decision trees and cluster analysis. Modeling is made in software solution MATLAB, SPSS Clementine 11.1. Data used to modeling rainfall and their chemical composition in selected localities. For preparation data was used MS Excel 2007.

Keywords

Environment, classification, decision trees, cluster analysis, decision, multiple creational decision, analytic hierarchy process, sensitivity, decision rule.

Obsah

1. Úvod	8
2. Životní prostředí	9
2.1. Popis dat	11
2.2. Příprava dat	17
2.3. Závěr kapitoly.....	22
3. Klasifikační modely.....	26
3.1. Shluková analýza	28
3.2. Model metodou K-means v prostředí Clementine.....	30
3.3. Dílčí závěr.....	32
3.4. Rozhodovací stromy.....	33
3.5. Model rozhodovacího stromu CRT v prostředí Clementine.....	35
3.6. Dílčí závěr.....	38
3.7. Závěr kapitoly.....	39
4. Modely vícekritériálního rozhodování.....	41
4.1. Vícekritériální rozhodování.....	43
4.2. AHP tři vrstvy.....	48
4.3. Verifikace modelu v prostředí MATLAB a MS Excel	50
4.4. Dílčí závěr.....	52
4.5. AHP čtyři vrstvy	55
4.6. Verifikace modelu v prostředí MATLAB	56
4.7. Dílčí závěr.....	57
4.8. Závěr kapitoly.....	61
5. Závěr	63
Seznam literatury.....	65
Seznam elektronických zdrojů.....	67
Seznam zkratk.....	68
Seznam obrázků.....	69
Seznam tabulek	71
Seznam vzorců	72

1. Úvod

Životní prostředí ovlivňuje náš každodenní život a jeho kvalitu. Oblast životního prostředí je velice rozsáhlá a rozmanitá. Do životního prostředí zahrnujeme všechny věci, které nás obklopují, například ovzduší, voda, půda, věci kolem nás (odpad). Životní prostředí by se dalo popsat jako jeden velký složitý systém, rozdělující se na subsystémy, ovzduší, vodu, půdu a jiné, které jsou mezi sebou navzájem propojené. Pokud máme v ovzduší nežádoucí prvky, dostanou se přes srážky do půdy. Pokud proniknou škodlivé látky do půdy, dostanou se i do spodních vod. Člověk, který svou činností narušuje tento systém, se musí následně postarat o návrat do původního stavu (rovnováhy).

Cílem mé diplomové práce je analýza a porovnání klasifikačního a rozhodovacího modelu. Jako zástupce klasifikačních modelů jsem vybrala shlukovací analýzu metodu K-Means a rozhodovací stromy. Zástupce rozhodovacích modelů jsem zvolila vícekritériální rozhodování metodu AHP. Data, která jsem vybrala, představují měření chemického složení srážek a atmosférické depozice na třech stanicích v okrese Trutnov v letech 2000 – 2007.

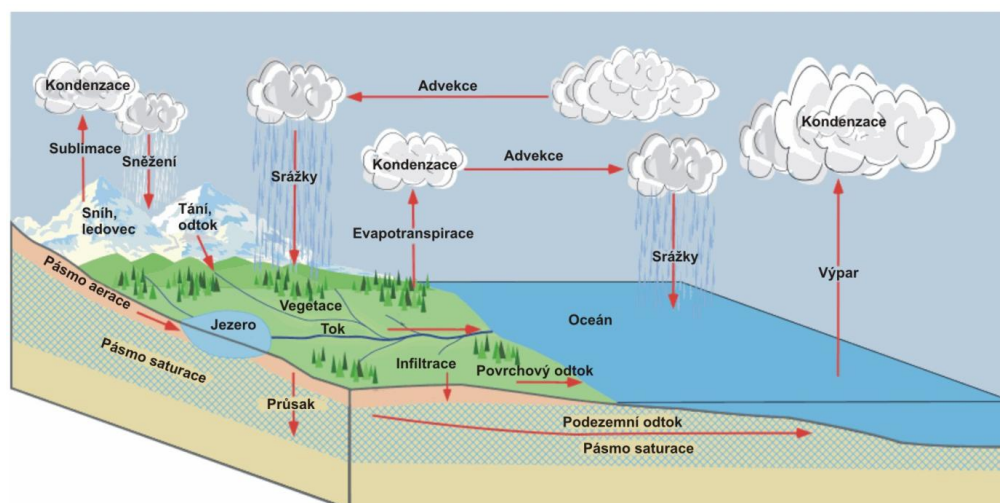
Ve své práci se budu zabývat návrhem klasifikačního stromu pro klasifikaci jednotlivých měření na základě lokality, srážek a pH. Nejprve budu muset upravit vstupní data, na kterých provedu shlukování a následně naučím klasifikační stromy. Následně použiji i původní naměřená data. Využiji programové prostředí Clementine. Použiji stejná data jako pro rozhodovací problém, návrh stromu verifikuji v prostředí Clementine, který ho optimalizuje. Potom se budu zabývat návrhem analytického hierarchického procesu, který následně ověřím v programovém prostředí MATLAB. Nejprve analyzuji data, předzpracuji je v prostředí EXCEL, návrh strukturu AHP verifikuji v prostředí MATLAB, který nalezne optimální řešení.

Klasifikace a rozhodování je součástí našeho života. Hledáme podobnost mezi situacemi, ve kterých se musíme rozhodovat, zdali zvolíme jednu nebo druhou variantu rozhodnutí. Pokud se rozhodujeme podle jednoho kritéria je to jednoduché, ale ne vždycky je rozhodování a klasifikace tak jednoduché a jednoznačné.

2. Životní prostředí

Pro člověka je asi nejdůležitější složkou životního prostředí (ŽP) ovzduší, protože se bez něho nemůže žít. Proto je kvalita ovzduší věnována velká pozornost jak na národní, tak na evropské a mezinárodní úrovni. Další důležitou částí ŽP je voda, kterou potřebujeme pro přežití. S těmito složkami je úzce spojena i třetí důležitá složka půda, kterou člověk také potřebuje ke svému životu. V půdě se pěstují plodiny, které slouží člověku jako potrava. Ovzduší ovlivňuje na půdu v podobě srážek. Látky obsažené v půdě se mohou dostat do okolních a spodních vod, které se vypařují a dostávají se zpátky do ovzduší.

Koloběh vody představuje nepřetržitou cirkulaci vody, mezi půdou, vodou a ovzduším, způsobenou slunečním zářením a gravitací. Zahrnuje výparu, přenos vodní páry, kondenzaci, atmosférické srážky a odtok. Malý oběh vody probíhá nad hladinou oceánů a velký oběh vody probíhá mezi oceánem a pevninou [11]. Názorně to popisuje obrázek (Obrázek 1).



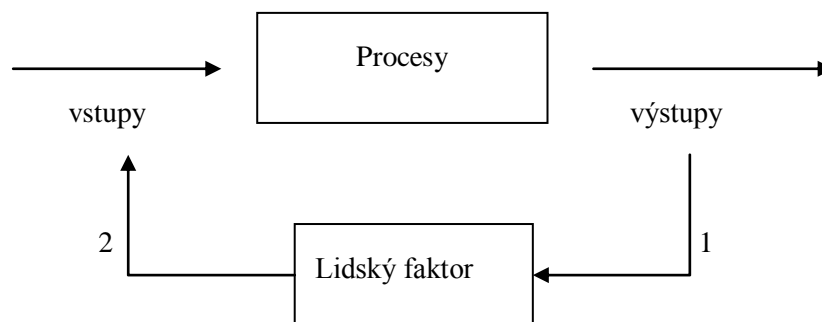
Obrázek 1 Koloběh vody [28]

Činností člověka se zvyšuje znečištění. Vyrůstá kontaminace vody, půdy a ovzduší kovy. Největšími znečišťovateli jsou kadmium, olovo, nikl a rtuť. Tyto látky se začínají v tkáních organismů akumulovat a jsou pro organismus nebezpečné [21].

Nárůst dopravy a průmyslu po roce 2000 způsobily zhoršení kvality ovzduší v České republice. Svou měrou k tomu přispívá chování lidí, kteří k topení v domácnostech používají nekvalitní paliva

či dokonce komunální odpad a vypouští tak do ovzduší nebezpečné látky. Největší problém v současné době představuje jemný prach [30].

Vstupní látkové toky atmosféry jsou plynné látky, tuhé látky a mořský aerosol. Výstupem je atmosférická depozice neboli usazené látky atmosféry. Jedním typem atmosférické depozice jsou dešťové srážky. Procesy jsou změny a přeměny probíhající v atmosféře z hlediska fyzikálního nebo fyzikálně chemického [26]. Látkové toky v atmosféře jsou znázorněny na obrázku (Obrázek 2). Část zpětné vazby s označením 1, představuje měření výstupů z atmosféry například měření složení mokrých a suchých depozit srážek. Blok lidský faktor se snaží o zmírnění znečišťování životního prostředí (ovzduší) a v přesvědčení znečišťovatelů o nápravu vzniklých škod. Část zpětné vazby s označením 2, zahrnuje konkrétní kroky a opatření na snížení znečištění ovzduší jednotlivými znečišťovateli.



Obrázek 2 Látkové toky v atmosféře

Voda pokrývá více než dvě třetiny zemského povrchu. Její objem se odhaduje na 1 400 mil. km³. Podstatnou část cca 97 % tvoří slané vody moří a oceánů. Voda je základní složkou životního prostředí a nezbytnou podmínkou existence života na Zemi. Z tohoto pohledu je voda nezastupitelná a plní následující funkce [22]:

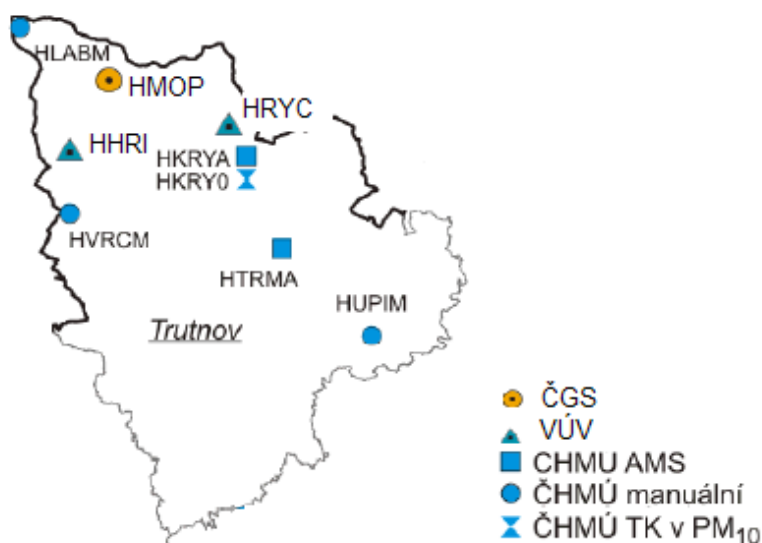
- biologickou,
- zdravotní,
- kulturní a estetickou.

Půda je svrchní, biologicky oživená část zemského povrchu. Vznikla působením půdotvorných faktorů a plní významné ekologické a produkční funkce. Je to omezený a nenahraditelný přírodní zdroj. V případě postupující degradace, ztráty ploch, se půda stává, jako zdroj v mnoha částech světa limitem dalšího rozvoje společnosti [31].

2.1. Popis dat

Chemické složení atmosférických srážek a atmosférické depozice se sledují na území České republiky dlouhodobě. Stanice ČHMÚ měří ve většině případů čisté srážky v týdenním intervalu. Od roku 1997 je na těchto stanicích zaveden týdenní odběr srážek typu bulk. Bulk představuje srážky s blíže nedefinovatelným obsahem prašného spadu na analýzu těžkých kovů Pb, Cd a Ni, Mn. Měření se provádí po týdnech, měsících popř. nepravidelných intervalech, při kterých se sleduje koncentrace ve srážkách typu bulk na volné ploše nebo pod korunami stromů [32]. Obsah látek ve srážkách představuje cond, pH, fluoridových, chloridových, síranových a dusičnanových aniontů a sodných, hořčíkových, vápenatých, draselných a amonných kationtů se provádí z čistých srážek. Pro zjištění obsahu zinku a železa, kadmia, olova, niklu a manganu se používá odběr typu srážky s prašným spadem.

Mokrá depozice je vytvořena pro vybrané ionty z celkových chemických analýz odebraných vzorků čistých srážek, a to konkrétně pro SO_4^{2-} - S, NO_3^- - N, NH_4^+ - N, H^+ (pH), F^- a Cl^- , a dále jsou tvořeny mokré depozice s blíže nedefinovatelným množstvím suché depozice pro Pb, Cd, Ni.

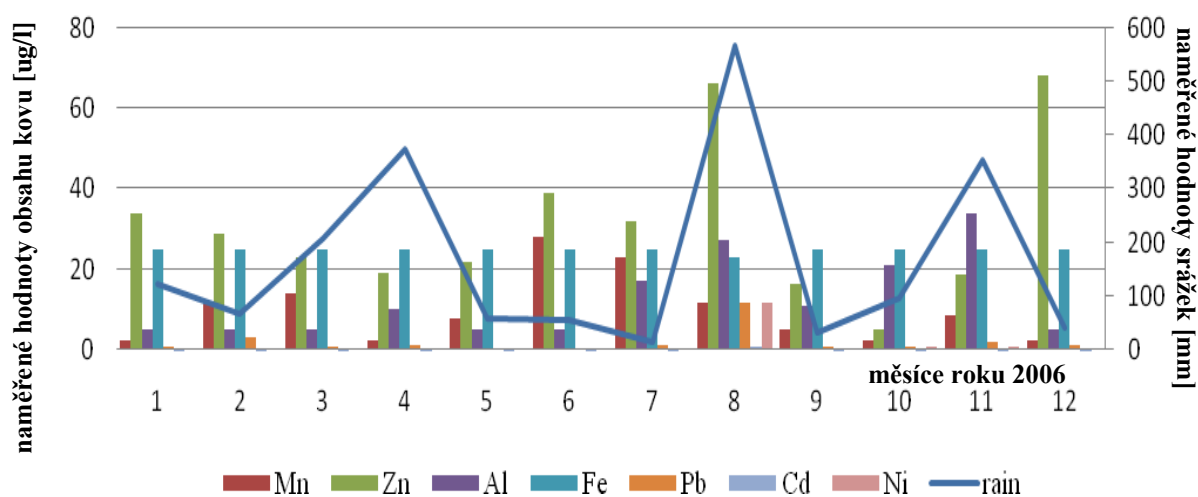


Obrázek 3 Staniční síť v okrese Trutnov [32]

Pro svoji diplomovou práci jsem si vybrala tři, z osmi měřících stanic. Stanice zabývající se měřením chemického složení srážek a atmosférické depozice jsou v okrese Trutnov čtyři, ale pouze tři používají typ odběru měsíční bulk. Určení lokality znázorňuje obrázek (Obrázek 3). První stanice HHRI se nazývá Hřibecí, je v nadmořské výšce 842 m, měření zajišťuje Výzkumný ústav vodohospodářský

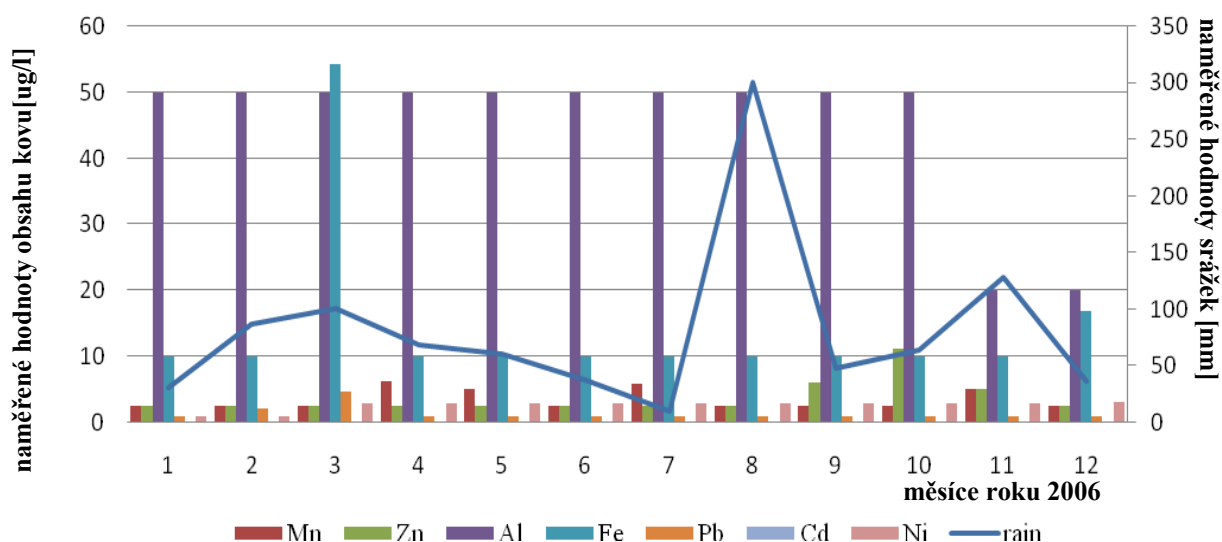
T.G.M. HMOP stanice s názvem Modrý potok se nachází v nadmořské výšce 1010 m, na této stanici měření zajišťuje Česká geologická služba. Na poslední třetí měřicí stanici HRYC měření zajišťuje Výzkumný ústav vodohospodářský T.G.M., tuto stanici nazýváme Rýchory, je v nadmořské výšce 1003 m. Ve všech třech lokalitách se provádí stejné měření Bulk (srážky s prašným spadem), já jsem si zvolila měsíční měření. Vybrala jsem si data z let 2000 – 2007. V průběhu sledovaného období došlo ke změně zkratk lokalit, já pracuji s označením bez určení metody. Lokalita Hřibčičí HHRIB, je ve zdrojích označena jako HHRIB a HHRI2, lokalita Modrý potok HMOPB je označována jako HMOPB nebo HMOP2 a třetí lokalita Rýchory HRYCB je označena jako HRYCB a HRYC2 [29]. V grafickém vyjádření používám označení HHRIB, HMOPB a HRYCB v tomto pořadí používám číselné označení 1,2,3.

Na grafech (Obrázek 4, Obrázek 5 a Obrázek 6) jsou znázorněny závislosti jednotlivých kovů na velikosti srážek za sledované měsíce roku 2006 ve všech třech lokalitách. Rok 2006 jsem vybrala z důvodu výpadku měření v roce 2007. Měření roku 2008 zatím nebyla zveřejněna. Z následujících grafů vyplývá, že obsah srážek je zcela nezávislý na úhrnu srážek naměřených ve sledovaném období. Na obrázcích jsou znázorněny dva grafy. Sloupcové grafy představují naměřené hodnoty obsahu kovů v ug/l, přesné hodnoty jsou čitelné na hlavní ose. Spojnicový graf spojuje hodnoty naměřených srážek v mm, jejichž hodnoty se ukazují na vedlejší ose.



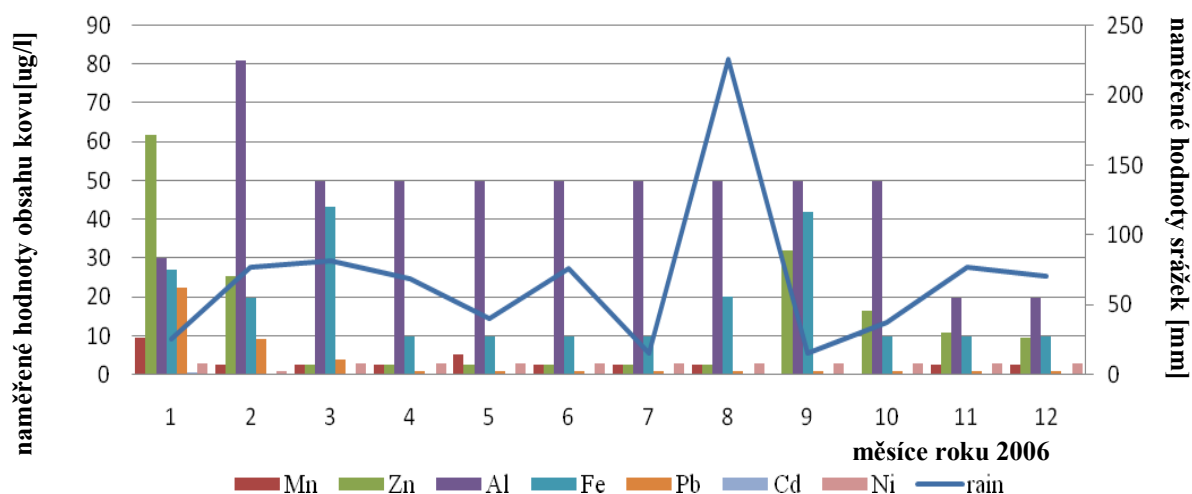
Obrázek 4 Naměřené srážky a obsažené kovy za jednotlivá měření v roce 2006 v lokalitě HMOPB

Graf na obrázku (Obrázek 4) názorně dokazuje, že kovy navzájem mezi sebou ani s objemem srážek nevykazují závislost v lokalitě Modrý potok, proto nevidím žádnou příčinu, proč bych tato data nemohla použít pro rozhodovací modely. V lokalitě Modrý potok na velikost objemu srážek reaguje hliník a nikl, obsah železa je konstantní vzhledem k naměřeným srážkám, nezávislý na objemu naměřených srážek je mangan a zinek.



Obrázek 5 Naměřené srážky a obsažené kovy za jednotlivá měření v roce 2006 v lokalitě HHRIB

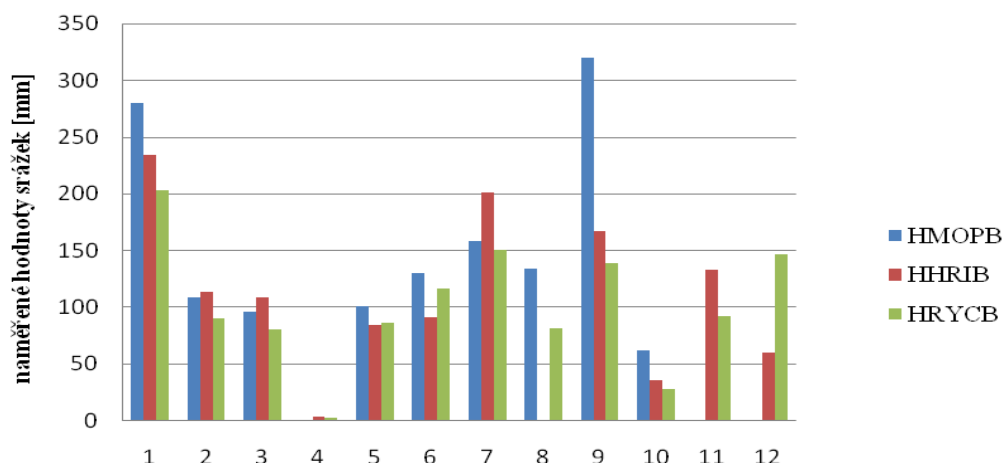
Graf na obrázku (Obrázek 5) názorně dokazuje, že kovy navzájem mezi sebou ani s objemem srážek nevykazují závislost v lokalitě Hříběcí, proto nevidím žádnou příčinu, proč bych tato data nemohla použít pro rozhodovací modely. V lokalitě Hříběcí na velikost objemu srážek výrazně nereaguje žádný z kovů, obsah hliníku je konstantní vzhledem k naměřeným srážkám, nezávislý na objemu naměřených srážek je mangan a zinek, železo, nikl, kadmium a olovo.



Obrázek 6 Naměřené srážky a obsažené kovy za jednotlivá měření v roce 2006 v lokalitě HRYCB

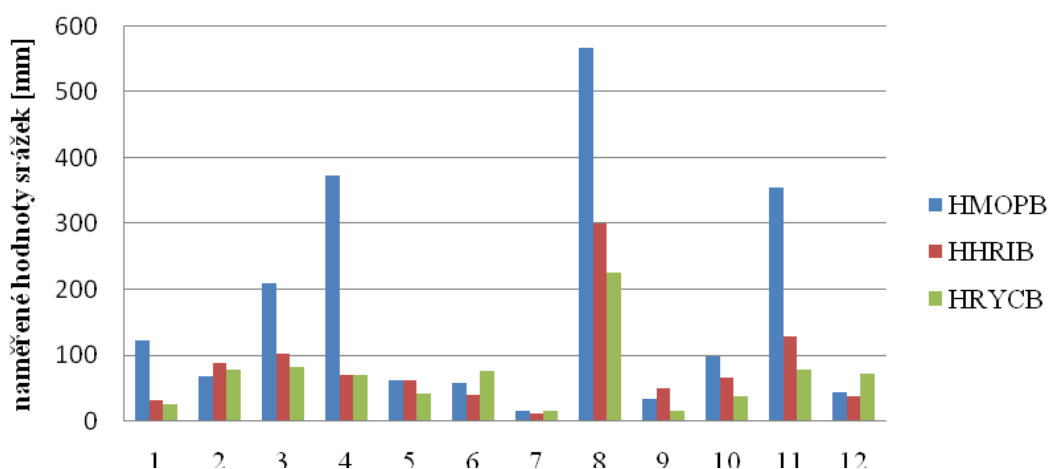
Graf na obrázku (Obrázek 6) názorně dokazuje, že kovy navzájem mezi sebou ani s objemem srážek nevykazují závislost v lokalitě Rýchory, proto nevidím žádnou příčinu, proč bych tato data nemohla použít pro rozhodovací modely. V lokalitě Rýchory na velikost objemu srážek výrazně nereaguje žádný z kovů, obsahy niklu a hliníku jsou konstantní vzhledem k naměřeným srážkám, nezávislý na objemu naměřených srážek je mangan a zinek, železo, kadmium a olovo.

Grafy na obrázcích (Obrázek 7, Obrázek 8, Obrázek 9, Obrázek 10) ukazují, že srážky nejsou závislé na ročním období. I když data představují časové řady, neobsahují zřejmou sezonní složku. Větší důležitost hraje náhodná složka.



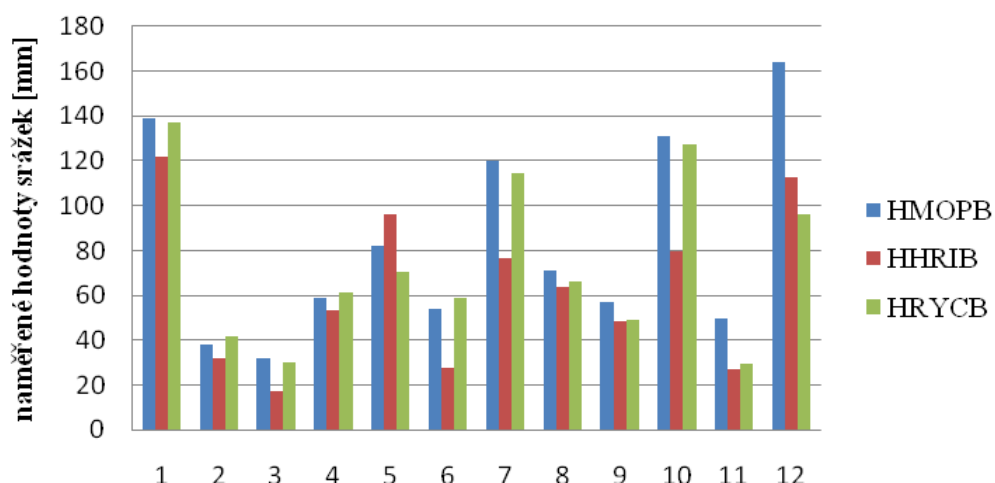
Obrázek 7 Naměřené srážky v mm za rok 2007 ve vybraných lokalitách

Graf na obrázku (Obrázek 7) znázorňuje, že v měření srážek v roce 2007 docházelo k výpadkům. V dubnu, listopadu a prosinci nedošlo k měření srážek v lokalitě Modrý potok, v srpnu neproběhlo měření v lokalitě Hřiběcí. Dešťové srážky ve všech třech lokalitách byly přiměřeně stejné, pouze v září bylo v lokalitě Modrý potok naměřeno o polovinu více srážek než u dvou ostatních lokalit. Minimum srážek bylo naměřeno v říjnu. Maximum srážek u zbylých dvou lokalit bylo naměřeno v lednu a minimum v dubnu. Průměrná hodnota srážek v roce 2007 v lokalitě Modrý potok bylo 154,11 mm, Hřiběcí 111,48 mm a Rýchory 100,93 mm. V červenci došlo k záměně pořadí v obsahu naměřených srážek.



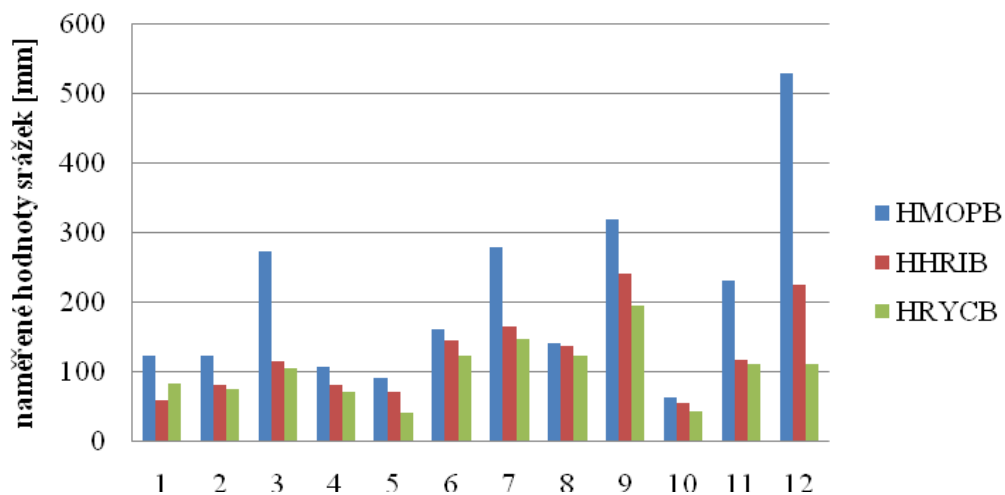
Obrázek 8 Naměřené srážky v mm za rok 2006 ve vybraných lokalitách

Graf na obrázku (Obrázek 8) ukazuje srážky v jednotlivých měsících roku 2006. V roce 2006 bylo maximum naměřených srážek v srpnu a minimum v červenci. Průměrná hodnota srážek v roce 2006 v lokalitě Modrý potok bylo 166,5 mm, Hřiběcí 80,91 mm a Rýchory 67,25 mm. Největší objem přeháněk byl v lokalitě Modrý potok, v měsících únor a září byly největší srážky v lokalitě Hřiběcí a v červenci a v prosinci byl největší objem srážek v lokalitě Rýchory.



Obrázek 9 Naměřené srážky v mm za rok 2003 ve vybraných lokalitách

Graf na obrázku (Obrázek 9) ukazuje srážky v jednotlivých měsících roku 2003. V roce 2003 byly nejnižší průměrné srážky v lokalitě Modrý potok 83,02 mm, Hřibčecí 63 mm a Rýchory 73,47 mm. Průměrné srážky byly nejvyšší v lokalitě Modrý potok, druhé v lokalitě Rýchory a jako nejnižší v lokalitě Hřibčecí. Došlo k záměně druhého a třetího pořadí v množství srážek ve sledovaném období 2000 - 2007. Lokalita Modrý potok byla lokalita s nejvyšším obsahem srážek. V únoru a v dubnu bylo naměřeno nejvíce srážek v lokalitě Rýchory, v květnu na Hřibčecím, v ostatních měsících bylo naměřeno nejvíce srážek v lokalitě Modrý potok.



Obrázek 10 Naměřené srážky v mm za rok 2001 ve vybraných lokalitách

Graf na obrázku (Obrázek 10) ukazuje srážky v jednotlivých měsících roku 2001. V roce 2001 byly naměřené hodnoty v lokalitě Rýchory pod 200 mm, v lokalitě Hřiběcí byla tato hranice dvakrát překročena a to v září a v prosinci. Lokalita Modrý potok tuto pomyslnou hranici překročila v březnu, v červenci a v září 300 mm srážek a v prosinci překročila hranici 500 mm. Průměrná hodnota srážek v roce 2001 v lokalitě Modrý potok byla 202,96 mm, Hřiběcí 123,67 mm a Rýchory 101,33 mm. Nejvyšší objem srážek byl naměřen za každé sledované období pouze v lokalitě Modrý potok.

2.2. Příprava dat

Při předzpracování dat jsem provedla korelaci, zjistila jsem čtyři vyšší hodnoty než 0,7. Po provedení korelace do klasifikačního modelu nezahrnuji hořčnaté kationty, železo, nikl a síranové anionty. Při statistické analýze jsem nezjistila žádné problémy, proč bych nemohla pracovat s těmito daty. Toto jsem provedla v MS Excel, dokládám v přílohách (Příloha 7, Příloha 8). Všechny atributy a jejich minimální a maximální naměřené hodnoty názorně znázorňuje datový slovník viz tabulka (Tabulka 1). Zkrácenou ukázkou matice vstupních dat představuje příloha (Příloha 6).

Kyselost srážek se uvádí podle stupnice od hodnoty 14 do hodnoty 1. Stupeň 14 představuje extrémně zásadité prostředí, stupeň 7 je neutrální prostředí a stupeň 1 je extrémně kyselé prostředí. Alkalické kovy tvoří homogenní skupinu nesmírně reaktivních prvků. Přestože jsou Na a K chemicky podobné, nenacházejí se v přírodě společně, což je způsobeno především rozdíly v jejich rozměrech. Vlastnosti prvků alkalických zemin, jsou závislé na jejich atomovém čísle, podobně jako tomu je u prvků alkalických kovů. V mořské vodě je obsaženo asi 0,13 % Mg. Pokud bude výroba pokračovat milión let, klesne obsah Mg asi na 0,12 %. Hořčík podobně jako Ca se vyskytuje v horninách ve formě nerozpustných uhličitanů, síranů a křemičitanů. Odhad jejich celkového zastoupení závisí na geochemickém modelu. Křemík je po kyslíku nejrozšířenějším prvkem v zemské kůře. Olovo je nejrozšířenějším těžkým prvkem, kterému se do výskytu blíží jen thallium a uran [8].

Tabulka 1 Měřené látky

Zkratka	Měřená látka / veličina	Hodnoty	
		min,	max,
Al [ug/l]	hliník	0,1	213
Ca [ug/l]	vápenaté kationty	0,03	15400
Cd [ug/l]	kadmium	0,01	21,2
Cl [ug/l]	chloridové anionty	0,08	6880
cond [μS/cm]	vodivost (konduktometrické stanovení)	1,04	312
F [ug/l]	fluoridové anionty	0,01	390
Fe [ug/l]	železo	0,1	270
K [ug/l]	draselné kationty	0,02	9310
Mg [ug/l]	hořečnaté kationty	0,01	1120
Mn [ug/l]	mangan	0,02	375
Na [ug/l]	sodné kationty	0,02	2270
NH ₄ [ug/l]	amonné kationty	0,01	13900
Ni [ug/l]	nikl	0,01	40,1
NO ₃ [ug/l]	dusičnanové anionty	0,13	36300
Pb [ug/l]	olovo	0,2	39
pH	pH	3,69	7,85
rain [mm]	srážky	2,1	567
SO ₄ [ug/l]	síranové anionty	0,02	16700
Zn [ug/l]	zinek	0,03	9277

Dalším problémem jsou výpadky v měření, proto je potřeba provést substituci chybějících hodnot. Rozhodla jsem se pro substituci průměrné hodnoty měření jednotlivých vlastností měření.

Měření úhrnu srážek se provádí každý den pomocí automatického srážkoměru. Ostatní hodnoty se měří v intervalu jednou za týden. Na zjištění vodivosti se používá analytická metoda C-metr, na hodnoty pH se používá pH metr, na zjištění hodnot fluoridových, chloridových, síranových a dusičnanových aniontů se používají analytické metody iontové chromatografie. K určení hodnot sodných, hořčíkových, vápenatých, draselných, amonných kationtů a zinku, železa se používá atomová absorpční spektrometrie. Kadmium, olovo, nikl a mangan se stanovují pomocí atomové absorpční spektrometrie s elektrotermickou atomizací.

Následující tabulky (Tabulka 2, Tabulka 3) znázorňují velké rozptýlení naměřených hodnot v rámci lokalit. Při porovnání průměrných hodnot a směrodatné odchylky zjistíme, že směrodatná odchylka je

větší než průměrná hodnota. Při důkladnějším prozkoumání všech vlastností se dá konstatovat odlehlost hodnot. Jednotlivé vlastnosti jsou po testování Dixonovým testem odlehlé. Použila jsem Dixonův test, jako kritickou hranici jsem stanovila hodnotu 0,341. Tato hodnota je stanovena pro 30 hodnot, ale pro větší hodnoty dále klesá, proto ji mohu použít.

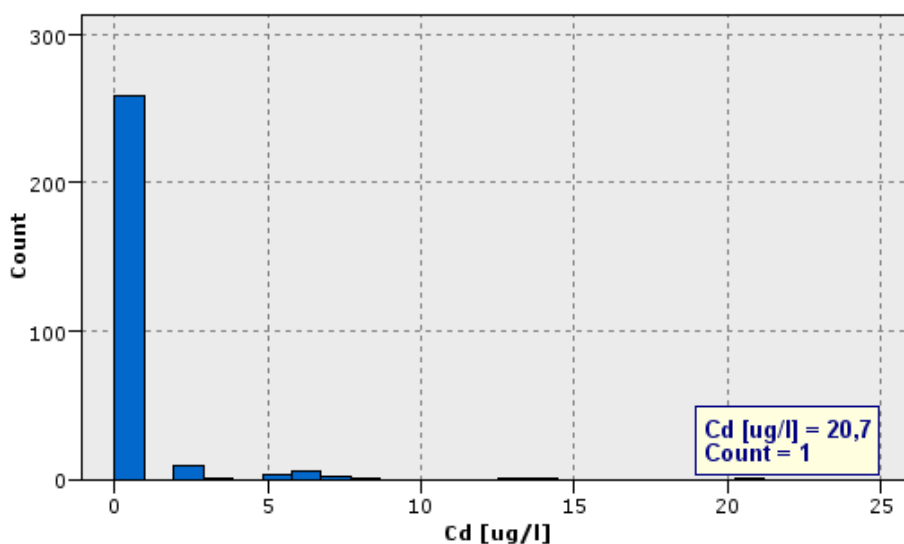
Tabulka 2 Průměrné hodnoty naměřených kovů ve srážkách v letech 2000 - 2007

Lokalita	Al	Cd	Fe	Mn	Ni	Pb	Zn
HHRIB	31,07	0,89	29,6	6,95	3,76	2,92	14,44
HMOPB	14,06	0,27	20,05	20,58	1,07	2,51	65,86
HRYCB	37,01	0,95	35,33	7,6	3,45	5,78	21,78

Tabulka 3 Směrodatná odchylka neměřených kovů ve srážkách v letech 2000 - 2007

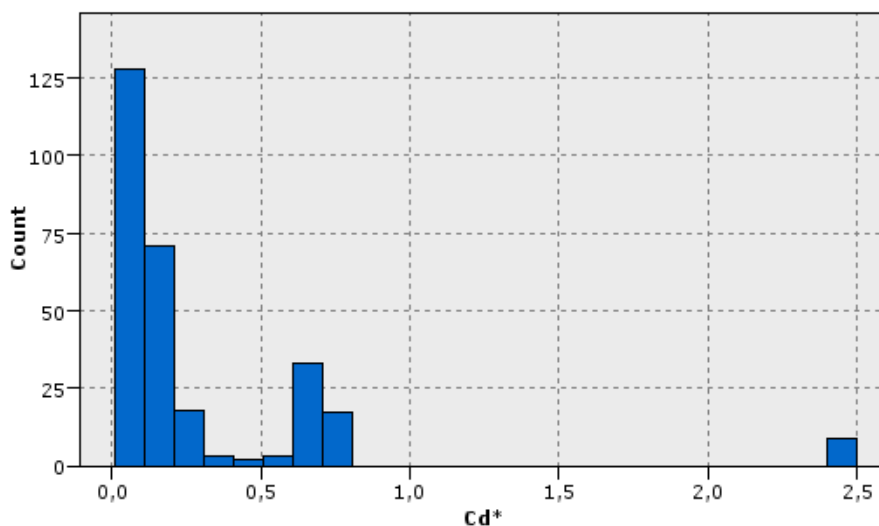
Lokalita	Al	Cd	Fe	Mn	Ni	Pb	Zn
HHRIB	32,28	2,17	43,65	9,47	7,76	3,40	17,48
HMOPB	22,03	0,27	17,60	54,54	1,19	1,76	83,72
HRYCB	37,53	2,81	50,80	12,89	7,61	7,71	30,86

Na obrázku (Obrázek 11) jsou znázorněny statisticky odlehlé hodnoty kadmia. Podle Dixonova statistického testu pro odlehlé hodnoty jsou tyto hodnoty kadmia a dalších měřených hodnot mimo srážek a pH. Tyto hodnoty jsem nahradila viz tabulka (Tabulka 4) a její popis.



Obrázek 11 Odlehlé hodnoty kadmia

Na obrázku (Obrázek 12) jsou znázorněné upravené hodnoty. I tak by se mohlo zdát, že hodnota 2,5 je odlehlá, tuto hodnotu již nebudu dále nahrazovat. Pro měřené vlastnosti a jejich hodnoty provedu pouze první substituci.



Obrázek 12 Upravené hodnoty kadmia

Na obrázku (Obrázek 13) můžeme názorně vidět změnu rozdělení, ke kterému došlo z důvodu nahrazení odlehlých hodnot průměrnou hodnotou. Odlehlé hodnoty bych mohla vyloučit, ale z důvodu zachování velikosti matice se dopouštím této chyby. V levé části obrázku je původní rozdělení a v pravé části je upravené rozdělení. Ke změně rozdělení dochází i z důvodu zkrácení sledovaného intervalu.

Field	Graph	Type	Field	Graph	Type
cond [μS/c...		Range	cond*		Range
pH		Range	pH		Range
Na [ug/l]		Range	Na *		Range
K [ug/l]		Range	K*		Range
NH4 [ug/l]		Range	NH4 *		Range
Ca [ug/l]		Range	Ca*		Range
Mn [ug/l]		Range	Mn*		Range
Zn [ug/l]		Range	Zn*		Range
Al [ug/l]		Range	Al*		Range
Pb [ug/l]		Range	Pb*		Range

Obrázek 13 Porovnání změny rozdělení pravděpodobností výskytu jednotlivých vlastností

V tabulce (Tabulka 4) jsou uvedeny průměry a směrodatné odchylky jednotlivých vlastností před úpravou a po úpravě dat. Pomlčky znázorňují hodnoty, které nebyly upravovány. Odlehle hodnoty jsem nahradila průměrnou hodnotou. Pokud bych na základě Dixonova testu stanovila statisticky odlehle hodnoty, kterého bych vyřadila celé měření, neměla bych dostatečně velkou datovou matici pro klasifikaci.

Tabulka 4 Průměry a směrodatné odchylky pro neupravená a upravená data

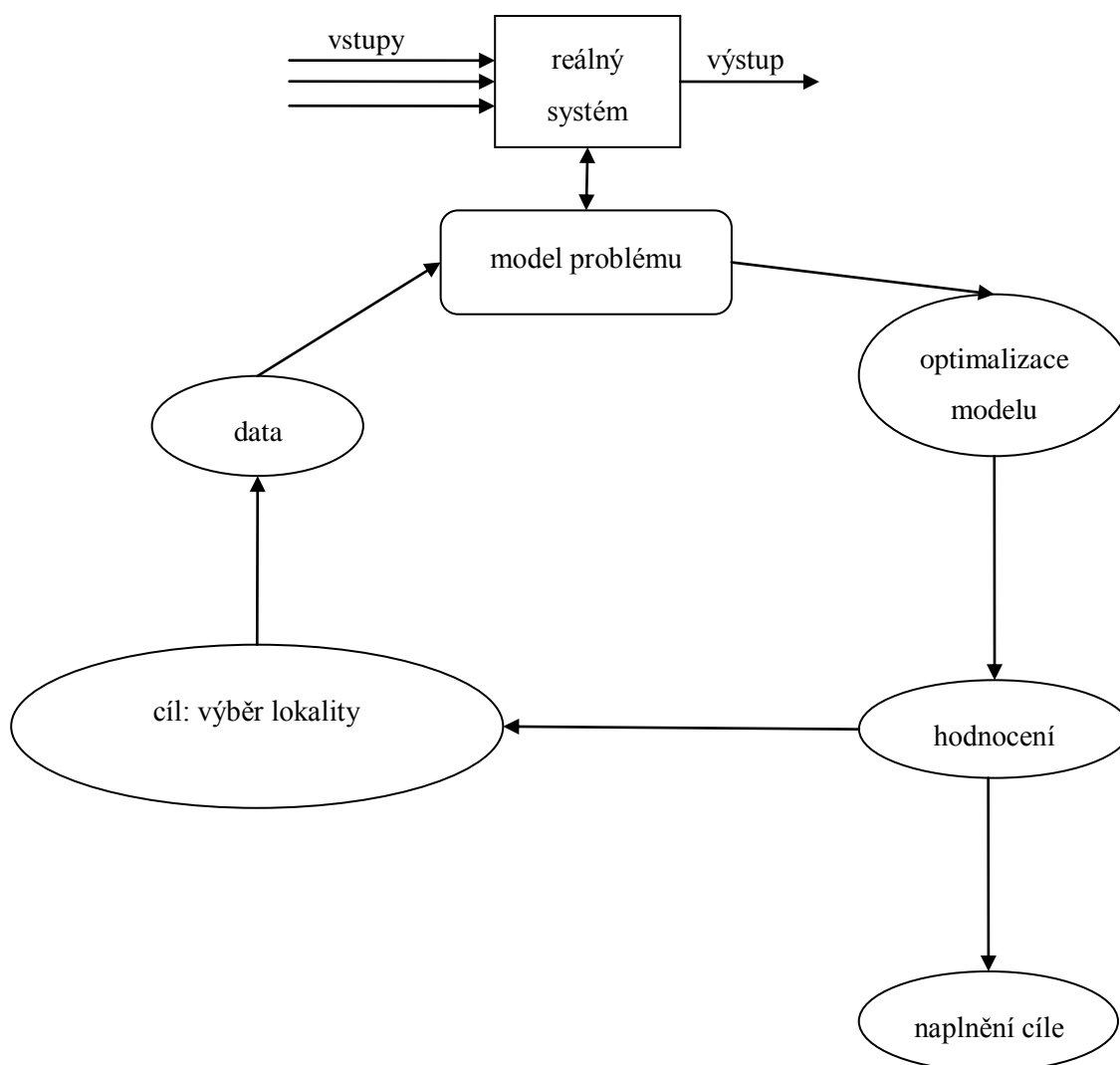
Vlastnosti	původní data		upravená data	
	průměr	odchylka	průměr	odchylka
rain	114,22	85,20	-	-
cond	23,14	25,04	18,36	7,95
pH	4,91	0,67	-	-
Na	188,25	276,47	124,98	128,05
K	199,65	649,29	88,30	84,46
NH4	674,77	1670,43	289,32	289,11
Ca	423,32	1141,70	195,12	191,54
Mn	11,63	33,04	6,58	5,67
Zn	31,93	84,13	16,93	11,07
Al	27,51	32,72	19,54	16,14
Pb	3,75	5,19	2,93	2,72
Cd	0,71	2,08	0,30	0,46
F	19,77	31,12	11,52	9,88
Cl	412,60	631,31	290,01	285,16
NO3	2135,01	4168,53	1170,25	1395,96

2.3. Závěr kapitoly

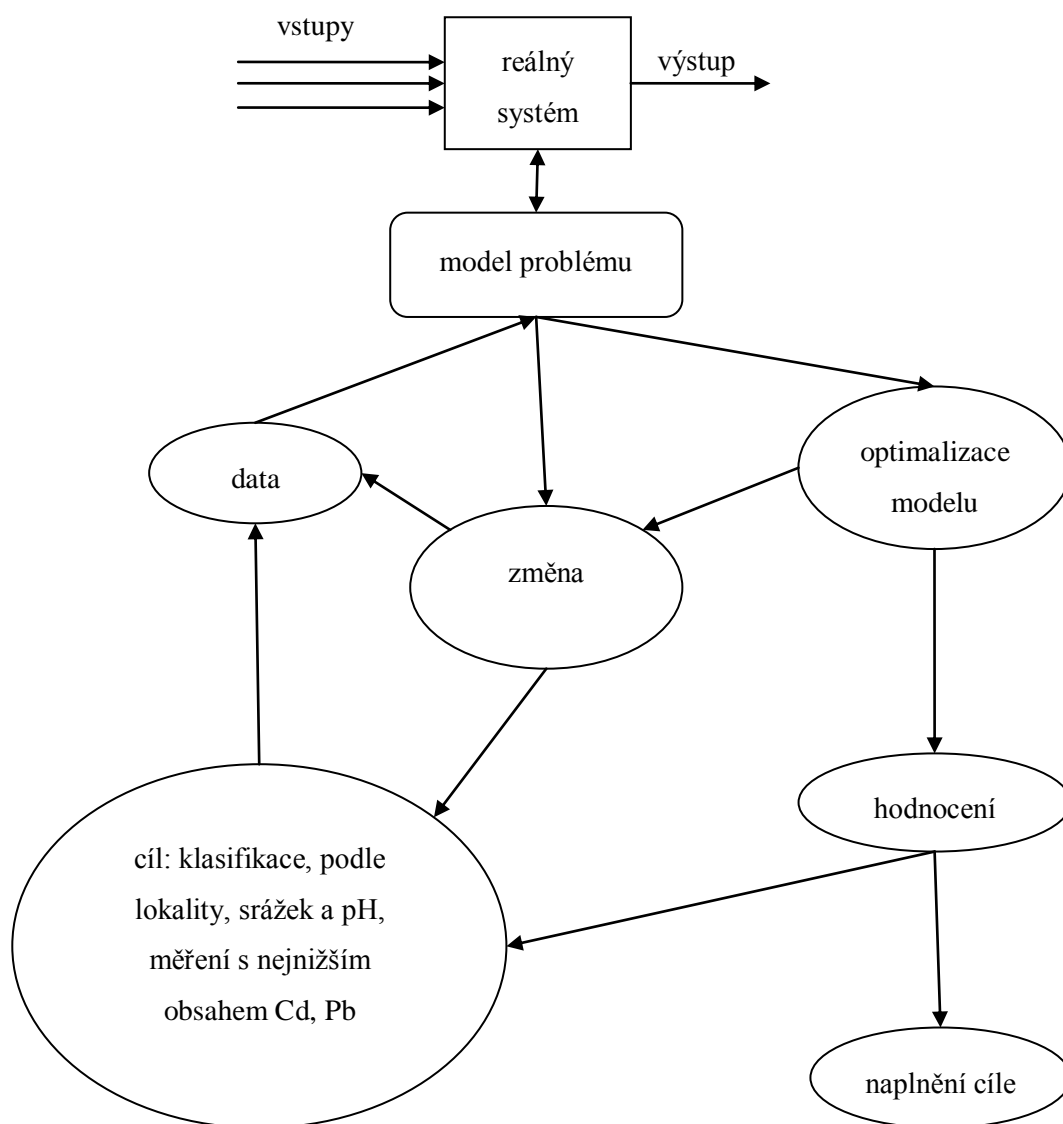
Pro klasifikaci jsem použila jako objekty jednotlivá měření a vlastnosti budou hodnoty složení srážek, dále lokalita, organizace, měsíc a rok měření. V lokalitách se sledují následující látky: hliník, vápenaté kationty, kadmium, chloridové anionty, vodivost, fluoridové anionty, železo, draselné kationty, hořečnaté kationty, mangan, sodné kationty, amonné kationty, nikl, dusičnanové anionty, olovo, pH, srážky, síranové anionty a zinek. Cílem je, posoudit zdali se jednotlivé lokality výrazně odlišují složením látek obsažených ve srážkách bližší specifikace viz [29]. Po provedení korelace jsem do modelu nezahrnula hořečnaté kationty, železo, nikl, síranové anionty viz příloha (Příloha 8).

Pro rozhodovací proces jsou tři alternativy představující lokality měření a kritéria představují sedm vybraných kovů. Vítězná alternativa bude lokalita s minimálním obsahem kovů naměřených srážek. Vzhledem k větší podobnosti obsahu látek za kratší časové období jsem nevyřadila z rozhodovacího modelu železo a nikl jako v předchozím klasifikačním modelu. Kritéria budou minimalistická.

Svůj postup řešení daného problému jsem nastínila na obrázku (Obrázek 14, Obrázek 15). Prvním krokem je stanovení cíle. Druhým krokem je sběr a úprava vhodných dat. Já jsem si z oblasti životního prostředí zvolila měření dešťových srážek s prašným spadem, informace o lokalitě a naměřené hodnoty sledovaných látek. Klíčovým krokem je model problému, který je znázorněn na obrázku (Obrázek 16). V dalších dvou krocích zvolený model optimalizuji a zhodnotím. Pokud po zhodnocení dospěji k výsledku, postup ukončím, zdali nedojdu k požadovanému výsledku, pokračuji od začátku.

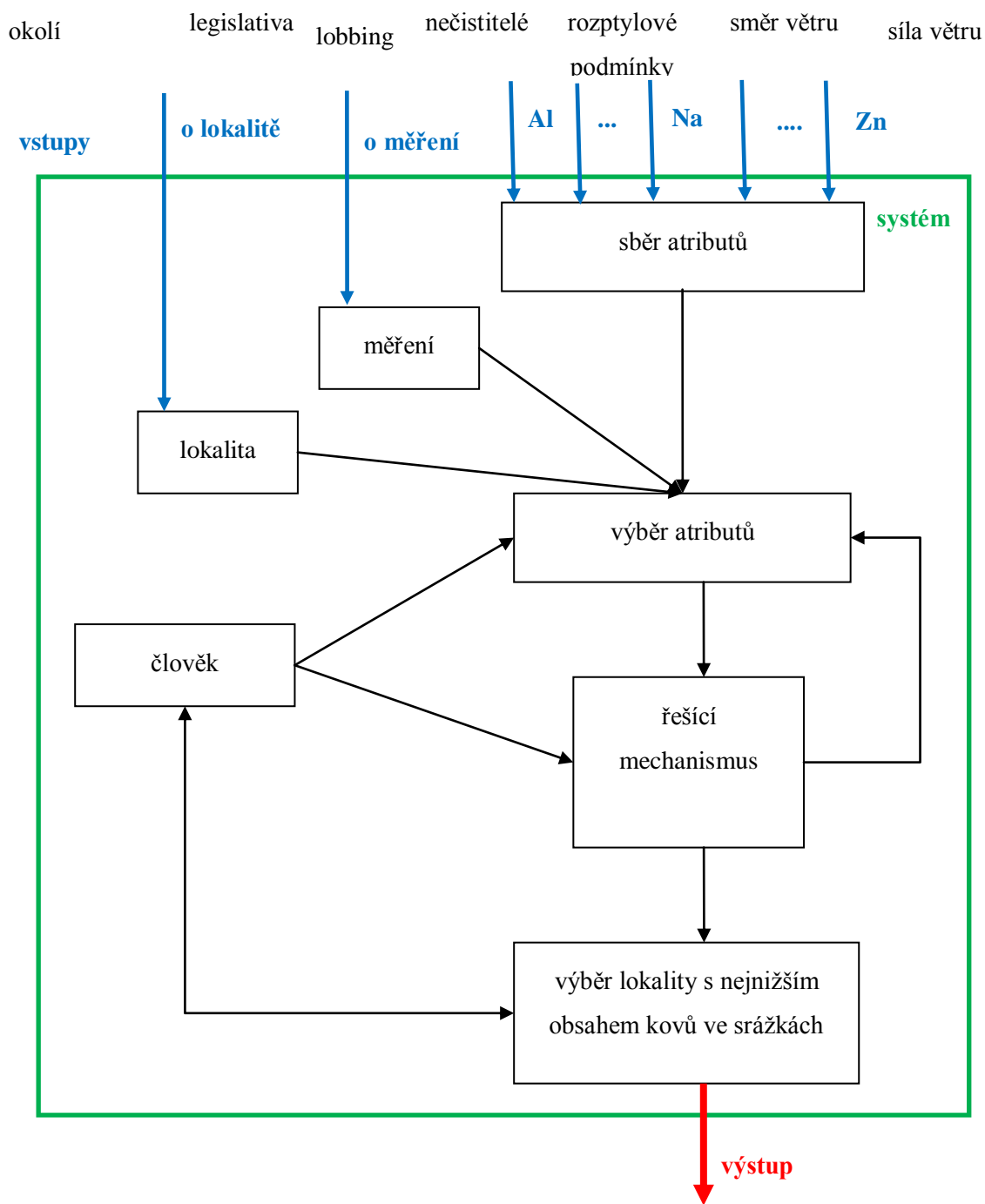


Obrázek 14 Návrh rozhodovacího modelu



Obrázek 15 Návrh klasifikačního modelu

V reálném životě existuje hodně složitých jevů, k jejichž pochopení využíváme obecných modelů. „Model musí vyjadřovat ty stránky daného jevu, které jsou z hlediska studia a zkoumání důležité.“ [13] Obrázek (Obrázek 16) popisuje můj pohled na vybraný problém.



Obrázek 16 Model systému

3. Klasifikační modely

Data Mining se zabývá získáváním informací z dat. Tento výraz znamená průzkum a analýzu velkého množství dat a zjišťování smysluplných modelů a pravidel. Úlohy, jejichž řešením se Data Mining zabývá, jsou různorodé a mají velké zastoupení v oblasti ekonomiky a obchodu. Mnoho úloh může být řešeno do následujících šesti úkolů [3] :

- klasifikace,
- odhadování,
- predikce,
- rozřazení do zájmových skupin,
- seskupování,
- popis a profilování.

Při řešení úlohy rozpoznávání provádíme zařazení objektů do tříd. Každý objekt je svým popisem jedinečný a zařazení do třídy představuje generalizaci objektů. Úloha rozpoznávání má dvě části [17] :

- zpracování objektů, aby byla maximalizována diskriminační schopnost při minimalizaci dat,
- přiřazení indikátoru třídy na základě předchozího bodu tj. klasifikace.

Klasifikace představuje zařazení prvků do tříd nebo seřazení neuspořádaných skupina prvků do shluků [9] . „Člověk je schopen objevovat a zobecňovat pravidelnosti ve zdánlivě chaotických souborech pozorování. Tuto schopnost nazýváme induktivním odvozováním z faktů poskytovaných učitelem nebo prostředím [3] .“ S vývojem systémů schopných automatické indukce se rozvíjí oblast znalostního inženýrství. Od těchto systémů očekáváme podstatný příspěvek k řešení problému ve vývoji znalostních systémů umělé inteligence. Při induktivním získávání znalostí se na základě konečného počtu vyřešených rozhodovacích situací získává obecný návod jak postupovat i v neznámých situacích [3] .

Klasifikací nazýváme činnost vytvářející rozklad množiny objektů, vedoucích k vytvoření rozkladu tříd. Systém tříd rovněž nazýváme klasifikací. „Klasifikace je činnost, jejímž výsledkem je klasifikace [20] .“ Dalším úskalím tohoto slova je „klasifikace“ ve škole. Učitel klasifikací pouze zařazuje znalost žáka do tříd, v tomto případě se jedná o identifikaci[20] . Algoritmus klasifikace je založen na určitých předpokladech o vlastnostech klasifikovaných objektů, například rozdělení pravděpodobností [19] .

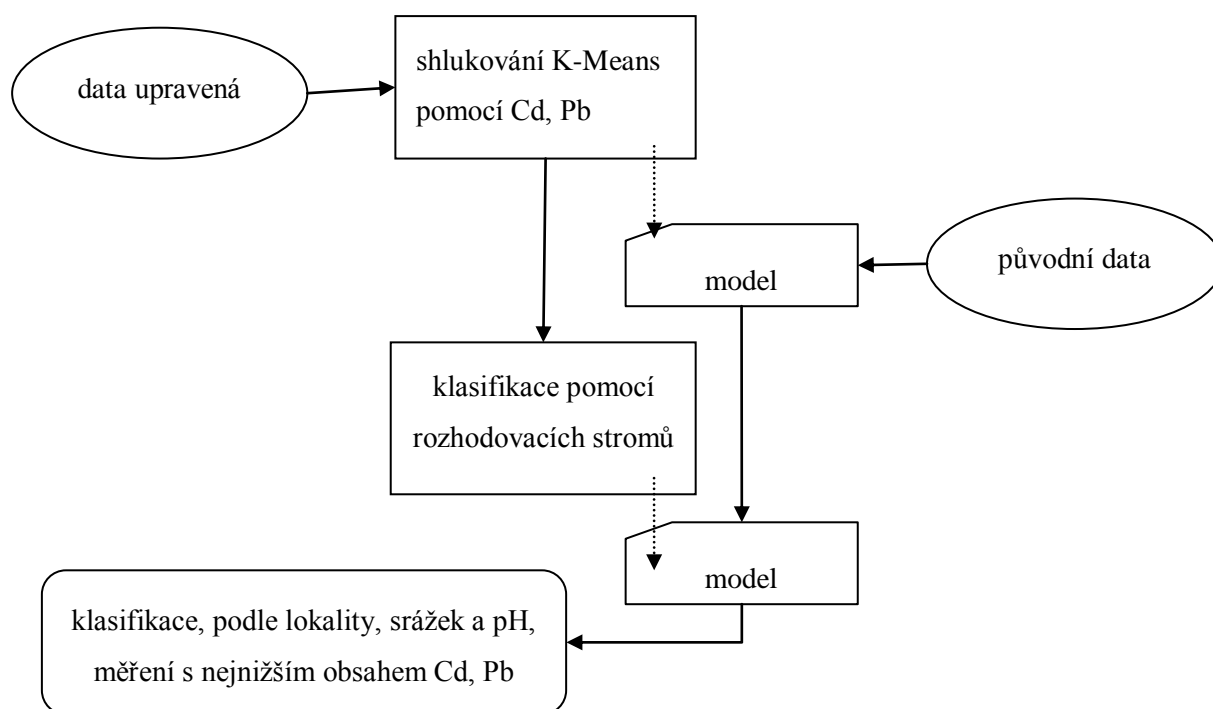
Mezi klasifikační metody patří:

- diskriminační analýza,
- regresní analýza,
- neuronové sítě,
- shluková analýza,
- faktorová analýza,
- klasifikační stromy.

Vstupem pro modely jsou předzpracovaná data, výstupem jsou znalosti. Modely se od sebe odlišují [1] :

- způsobem reprezentování hledaných znalostí,
- vhodností typu úloh,
- srozumitelnost pro uživatele,
- efektivností při klasifikaci nových případů,
- vhodností typu dat.

Následující obrázek (Obrázek 17) objasňuje algoritmus řešení klasifikačního problému. Zde jsem musela řešit problém rozptýlených hodnot vlastností (obsah látek naměřených ve srážkách) v jednotlivých lokalitách. Proto jsem přistoupila k úpravě dat. Dopustila jsem se chyby, a to náhradou odlehlých hodnot za průměrné hodnoty. Pro vytváření modelů jsem použila upravená data. Na základě těchto modelů se mi podařilo provést klasifikaci i na původních datech. Snažila jsem se seskupit jednotlivá měření pomocí všech kovů. Bohužel na základě tohoto shlukování se mi nepodařilo naučit rozhodovací strom na více než 70%. Výsledky uvádí příloha (Příloha 5). Na základě publikací viz [22] , [8] , [10] , jsem zvolila shlukování pouze pomocí dvou kovů a to kadmia a olova. Pro klasifikaci jsem použila jako vstupní hodnoty lokalitu, srážky a pH. Výstupní hodnotou pro mě byly skupiny shluků podle obsahu kadmia a olova. Výsledkem je klasifikace měření s nejnižším obsahem kadmia a olova na základě naměřených srážek a pH v jednotlivých lokalitách.



Obrázek 17 Algoritmus klasifikace

3.1. Shluková analýza

Shlukování je významný nástroj, který se používá, když máme velkou složitou množinu dat s množstvím vlastností a velkou vnitřní strukturou[3]. Analýza shluků patří mezi metody, které se zabývají zjišťováním podobností vícerozměrných objektů. Je vhodné ji použít tam, kde objekty projevují přirozenou tendenci se seskupovat. Při odhalování struktury objektů je shluková analýza velmi citlivá na přítomnost nevýznamných hodnot vlastností. Shlukovou analýzu citlivě ovlivňuje odlehlost objektu, které se výrazně odlišují [19]. Proto je nutno použít Dixonův test odlehlých hodnot [12]:

$$Q_i = \frac{x_n - x_{n-1}}{R_{\text{var}}} \quad (1)$$

kde R_{var} je variační rozpětí a vypočítáme ho jako $R_{\text{var}} = x_{\text{max}} - x_{\text{min}}$, Q_i je testovací kritérium, $x_n - x_{n-1}$ rozdíl porovnávaných hodnot. Používám kritickou hranici pro hodnotu 30 a více uvažuji na hladině významnosti 5 % 0,260 a na hladině významnosti 1 % 0,341.

Podle způsobu shlukování se algoritmy dělí na hierarchické a nehierarchické. Hierarchické algoritmy se dělí na aglomerativní a divizní. U aglomerativních postupů je na začátku každý objekt samostatným shlukem a postupně se tyto objekty slučují. Divizní algoritmy pracují na opačném principu [19]. Nehierarchické shlukování také dělíme do dvou skupin na optimalizační metody a analýzu modelů. Tato část je podrobněji rozebrána v publikaci [20] v kapitole Metody nehierarchického shlukování. Dalším typem shlukování je fuzzy shlukování, které doplňuje všechny shlukovací metody tím, že umožňuje shlukování jednoho objektu do více shluků. Podrobněji tuto problematiku rozebírá publikace [19] v kapitole Fuzzy shlukování.

Analýzu shluků je možné sledovat a vyšetřovat podobností objektů, pomocí dendogramu objektů nebo dendogramu proměnných. Dendogram podobnosti objektů je standardní výstup hierarchického shlukování, ze které je patrná struktura objektů ve shlucích. Dendogram podobnosti proměnných odhaluje m -tice proměnných, které jsou si velice podobné a silně spolu korelují [19].

Shluková analýza odpovídá na otázku, lze-li rozdělit zvolené objekty do skupin (shluků) navzájem si blízkých na základě vlastností. Důležitým předpokladem je, že dokážeme změřit vzdálenosti mezi objekty. Každý příklad je charakterizován m numerickými veličinami. Vzdálenost mezi vektory \bar{x}_1 a \bar{x}_2 lze vyjádřit různými způsoby například [1]:

- Hammingova vzdálenost,
- Euklidovská vzdálenost,
- Čebyševova vzdálenost.

Při hierarchickém shlukování se postupuje metodou „zdola nahoru“. Vzdálenost mezi shluky se dá vyjádřit několika způsoby [1]:

- metoda nejbližšího souseda,
- metoda nejvzdálenějšího souseda,
- metodou průměrné vzdálenosti,
- centroindní metodou.

Algoritmus hierarchického shlukování [1]:

1. určení vzájemné vzdálenosti mezi všemi objekty, každý příklad je samostatný shluk,
2. nalezení dvou nejbližších shluků a jejich spojení,
3. vypočítání vzdáleností mezi novým shlukem a ostatními shluky.

Další metodou je metoda nejbližších těžišť (K-means = K-středu). Algoritmus metody K-means:

- 1) zvolíme rozklad do k shluků,
- 2) určíme středy pro všechny shluky v aktuálním rozkladu,
- 3) pro každý objekt
 - a) určí vzdálenost středu shluku od objektů,
 - b) porovnání vzdálenosti mezi novým středem shluku a vzdálenosti od stávajícího středu, pokud je nová vzdálenost menší,
 - c) není-li součástí shluku nového, dojde k přesunu objektu do nového shluku,
- 4) došlo-li k nějakému přesunu, potom jdi na krok 2, jinak je konec.

Kritérium věrohodnosti je kritérium těsnosti proložení, které je založeno na srovnávání rozličných shluků a vychází z mezi shlukové sumy čtverců WSS_K , která je definovaná vztahem [19] :

$$WSS_K = \frac{n * m}{n * m - m} \sum_{k=1}^k \sum_{i=1}^m \sum_{j=1}^{n_k} (1 - \delta_{ijk}) * (y_{ij} - c_{ik})^2$$

kde: k – počet shluků, i – index i -té proměnné, m – počet proměnných, n – počet objektů, δ_{ijk} - chybějící hodnota i -té proměnné v j -tem řádku a k -tem shluku y_{ij} - vstupní údaje, c_{ik} - průměr i -té proměnné v k -tém shluku.

3.2. Model metodou K-means v prostředí Clementine

Rozhodla jsem se rozdělit lokality na základě jednotlivých naměřených vlastností pomocí rozhodovacích stromů. Bylo to velice obtížné, protože naměřený obsah jednotlivých vlastností ve srážkách je velice různorodý. Nakonec jsem zvolila rozdělení do shluků pouze podle dvou kovů v návaznosti na rozhodovací problém. Zvolila jsem kadmium a olovo, pro výběr jsem použila zdroj [22] . Zkoušela jsem také shlukování na základě všech pěti kovů, ale na základě těchto shluků jsem nedokázala naučit rozhodovací strom na více než 60%, což představuje náhodný výběr. Výsledky uvádí příloha (Příloha 5). Rozdělení do shluků všech kovů a pouze dvou vybraných kovů kadmia a olova, stanovení středů shluků je uvedeno na obrázku (Obrázek 19). Vstupy do uzlu pro vytvoření modelu jsou znázorněny na obrázku (Obrázek 18). Pro vytvoření modelu používám upravená data.

Field	Type	Values	Missing	Check	Direction
měření	Set	<Read>		None	None
rain	Range	[2.1,567.0]		None	None
cond	Range	[1.04,39.9]		None	None
pH	Range	[3.69,7.85]		None	None
Na	Range	[0.02,495.0]		None	None
K	Range	[0.02,200.0]		None	None
NH4	Range	[0.01,700.0]		None	None
Ca	Range	[0.03,594.0]		None	None
Mn	Range	[0.02,32.0]		None	None
Zn	Range	[0.03,34.0]		None	None
Al	Range	[0.1,50.0]		None	None
Pb	Range	[0.2,12.7]		None	In
Cd	Range	[0.01,2.5]		None	In
F	Range	[0.01,40.0]		None	None
Cl	Range	[0.08,980.0]		None	None
NO3	Range	[0.13,6330.0]		None	None
rok	Set	2000,2001,2002...		None	None
měsíc	Set	1,2,3,4,5,6,7,8,9...		None	None
lokality	Set	1,2,3		None	None
organizace	Set	1,2,3		None	None

Obrázek 18 Vstupy do uzlu K-means

Následující obrázek (Obrázek 19) znázorňuje tři shluky, do kterých se rozdělili jednotlivá měření podle dvou vybraných kovů a podle všech pěti kovů. Z tohoto obrázku je patrné, že pokud porovnáme měření pomocí dvou kovů, je největší početní zastoupení převážně v prvním shluku, je to shluk s nejnižším obsahem kadmia a olova. Shlukování na základě všech pěti kovů popisuje rovnoměrné zastoupení mezi prvním a třetím shlukem i v tomto případě je shluk druhý nejméně zastoupen. Dalo by se říci, že měření v druhém shluku jsou odlišná od ostatních měření a ne tak častá. Měření druhého shluku mají charakter vyššího obsahu kadmia než olova.

Cluster	Records	Elements
cluster-1: 236 records		Cd (0,233)
		Pb (2,029)
cluster-2: 9 records		Cd (2,5)
		Pb (2,402)
cluster-3: 39 records		Cd (0,218)
		Pb (8,496)
cluster-1: 88 records		Al (9,763)
		Cd (0,132)
		Mn (6,365)
cluster-2: 21 records		Pb (1,317)
		Zn (17,971)
cluster-3: 80 records		Al (1,02)
		Cd (1,298)
		Mn (0,639)
cluster-2: 21 records		Pb (3,106)
		Zn (0,885)
cluster-3: 80 records		Al (35,216)
		Cd (0,239)
		Mn (8,43)
cluster-2: 21 records		Pb (4,673)
		Zn (18,74)

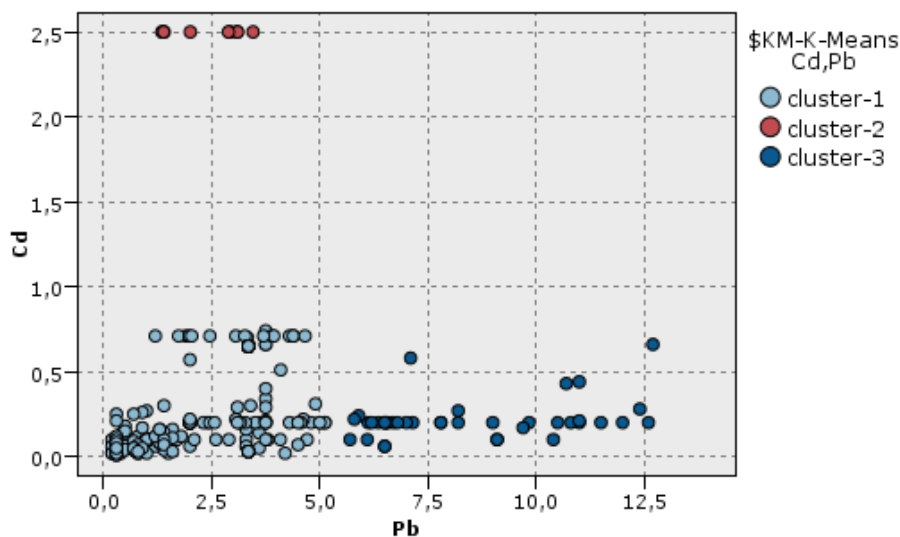
Obrázek 19 Porovnání shlukování

3.3. Dílčí závěr

Snažila jsem provádět shlukování dle největšího, logický možného počtu vlastností. Další krok rozhodovacího stromu bylo přínosné pouze shlukování na základě dvou kovů.

Na základě uzlu K-means se vytvoří shluky. U metody K-means musíme předem zadat počet shluků do kterých je potřeba shlukování provádět. Já jsem si vybrala tři shluky, pokud jsem zvětšovala počet shluků, docházelo k rozpadání krajních hodnot do menších shluků, nikoli k rozpadu velkého shluku. Shluky na obrázku (Obrázek 20) charakterizují:

- Cluster 1 (shluk 1) v tomto shluku se nachází nejvíce měření, shluk je charakteristický nízkým výskytem obsahu kadmia a olova ve srážkách. O tomto shluku můžeme říci, že je to shluk s nejnižším obsahem kovů.
- Cluster 2 (shluk 2) v tomto shluku se nachází nejméně měření, shluk je charakteristický vysokým výskytem obsahu kadmia a nízkým výskytem olova ve srážkách. O tomto shluku můžeme říci, že je to shluk s vysokým obsahem kovů.
- Cluster 3 (shluk 3) je charakteristický nízkým výskytem obsahu kadmia a vysokým obsahem olova ve srážkách. O tomto shluku můžeme říci, že je to shluk se středním obsahem kovů.



Obrázek 20 Grafické znázornění shlukování pomocí dvou kovů kadmia a olova

V důsledku nerovnoměrného rozdělení měření do shluků jsem sledovala rozdělení testovacích a trénovacích dat uzlu Partition. Graf na obrázku (Obrázek 21) ukazuje nerovnoměrné rozdělení do shluků, ale rovnoměrné rozdělení na testovací a trénovací data. Cluster 1 je rozdělen v poměru 114 : 122 na trénovací a testovací data. Cluster 2 je rozdělen v poměru 5 : 4 na trénovací a testovací data. Cluster 3 je rozdělen v poměru 19 : 20 na trénovací a testovací data.



Obrázek 21 Rozložení testovacích a trénovacích dat mezi shluky – upravená data

3.4. Rozhodovací stromy

Rozhodovací stromy jsou silnými a oblíbenými nástroji pro řešení úloh klasifikace a predikce [3]. Klasifikační stromy používáme k předpovídání členství jednotlivých objektů do tříd dle kategorií příslušnosti. Klasifikační stromy jsou jednou z technik používaných v Data Miningu. Výsledkem klasifikačních stromů je předpověď nebo vysvětlení reakce závislé proměnné, jsou to dostupné, běžně používané techniky jako u tradičních technik jako diskriminační analýza, shluková analýza, neparametrická statistika a nelineární odhadování. Pružnost klasifikačních stromů dělá z této metody dobrou metodu, avšak to není výlučným doporučením [9]. Existuje několik algoritmů pro tvorbu rozhodovacích stromů. Nejčastěji používané jsou algoritmy CHAID, CART, QUEST, C5.0 a obecný algoritmus TDIDT. Rozhodovací stromy se používají pro klasifikaci a predikci [2].

Mezi výhody rozhodovacích stromů patří schopnost vytvářet srozumitelná pravidla, možnost provádět klasifikaci bez velkých požadavků. Vstupem mohou být spojitá i kategorická data. Výsledkem stromu

je pouze veličina kategorická. Nevýhodou rozhodovacích stromů je potřeba velkého množství objektů a jejich vlastností, proto je méně vhodná pro časové řady.

Algoritmus TDIDT (Top-Down Induction of Decision Trees) byl navržen Quinlanem v roce 1966. TDIDT představuje algoritmus postupného rozkladu směrem dolů od kořene stromu k listům. První verze nesla název ID3 (Iterative Dichotomizer). ID3 vytváří iterativním způsobem rozhodovací strom tak, že v každém kroku [2] :

- vezme jeden atribut jako kořen stromu,
- rozdělí data v trénovací podmnožině na další podmnožiny podle hodnot tohoto atributu,
- nepatří-li všechna data podmnožiny do téže třídy, pro tuto podmnožinu opakuje postup od začátku.

Rozhodovací stromy jsou bi-nární, to vyjadřuje schopnost stromu větvit se pouze na dvě další podmnožiny stávající nebo kořenové množiny objektů. N-ární stromy mají schopnost větvení do více než dvou podmnožin podle hodnot atributů.

Volba vhodného atributu v kroku dvě se provádí na základě entropie, která se pro j-tou hodnotu atributu spočítá jako:

$$H_j = -\sum_{t=1}^T p_t \log_2 p_t \quad (2)$$

kde p_t je pravděpodobnost výskytu třídy t pro j -tou hodnotu atributu. Entropie pro celý atribut se spočítá jako vážený součet entropií pro jeho jednotlivé hodnoty:

$$H = -\sum_{j=1}^K p_j H_{ji} \quad (3)$$

kde p_j je relativní četnost j -té hodnoty atributu.

V každém kroku se pro dělení uvažované podmnožiny vybírá atribut s nejmenší hodnotou H [2] .

Každý rozhodovací strom můžeme převést na pravidla. Převod rozhodovacích stromů na pravidla se provádí průchodem stromu od kořenového uzlu k listovému uzlu. Převod rozhodovacího stromu na pravidla zvyšuje srozumitelnost [1] .

3.5. Model rozhodovacího stromu CRT v prostředí Clementine

Vstupem uzlu rozhodovacího modelu CRT jsou srážky, pH a lokalita měření. Následující obrázek (Obrázek 22) představuje výčet atributů zobrazujících vstupní a výstupní hodnoty modelu rozhodovacího stromu CRT.

Field	Type	Values	Missing	Check	Direction
metření	Set			None	None
rain	Range	[2.1,567.0]		None	In
cond	Range	[1.04,39.9]		None	None
pH	Range	[3.69,7.85]		None	In
Na	Range	[0.02,495.0]		None	None
K	Range	[0.02,200.0]		None	None
NH4	Range	[0.01,700.0]		None	None
Ca	Range	[0.03,594.0]		None	None
Mn	Range	[0.02,32.0]		None	None
Zn	Range	[0.03,34.0]		None	None
Al	Range	[0.1,50.0]		None	None
Pb	Range	[0.2,12.7]		None	None
Cd	Range	[0.01,2.5]		None	None
F	Range	[0.01,40.0]		None	None
Cl	Range	[0.08,980.0]		None	None
NO3	Range	[0.13,6330...]		None	None
rok	Set	2000,2001...		None	None
měsíc	Set	1,2,3,4,5,6,...		None	None
lokalita	Set	1,2,3		None	None
organizace	Set	1,2,3		None	None
\$KM-K-Means ...	Set	cluster-1,cl...		None	None
\$KM-K-Means ...	Set	cluster-1,cl...		None	Out
cluster Cd,Pb	Discrete	<Read>		None	In
cluster Al,Cd,M...	Discrete	<Read>		None	None
lokalita2	Discrete	<Read>		None	None
Partition	Set	"1_Trainin...		None	Partition

Obrázek 22 Vstupy do uzlu pro vytvoření rozhodovacího stromu

V modelu rozhodovacího stromu CRT je 8 pravidel pro určení tří shluků při použití všech třech vlastností, lokalita, srážky, pH. Rozhodovací strom je přiložen v příloze (Příloha 4). Čtyři pravidla popisují první shluk s nejnižším obsahem kadmia a olova. Jedno pravidlo charakterizuje druhý shluk s nízkým obsahem olova a vyšším obsahem kadmia a tři pravidla na určení třetího shluku s nízkým obsahem kadmia a vysokým obsahem olova. Pravidla popisující první shluk:

if lokalita2 in [1 2] and pH <= 5,750 then cluster-1

if lokalita2 in [1 2] and pH > 5,750 and rain > 78,400 then cluster-1

if lokalita2 in [3] and pH <= 6,300 and rain <= 109,500 then cluster-1

if lokalita2 in [3] and pH <= 6,300 and rain > 131,150 then cluster-1

Pravidlo popisující druhý shluk:

if lokalita2 in [3] and pH <= 6,300 and rain > 122,300 then cluster-2

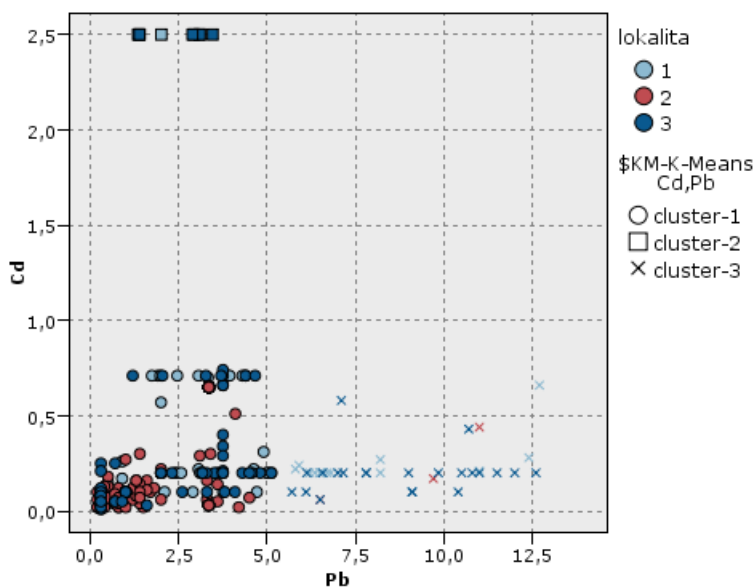
Pravidla popisující třetí shluk:

if lokalita2 in [1 2] and pH > 5,750 and rain <= 78,400 then cluster-3

if lokalita2 in [3] and pH <= 6,300 and rain > 109,500 then cluster-3

if lokalita2 in [3] and pH > 6,300 then cluster-3

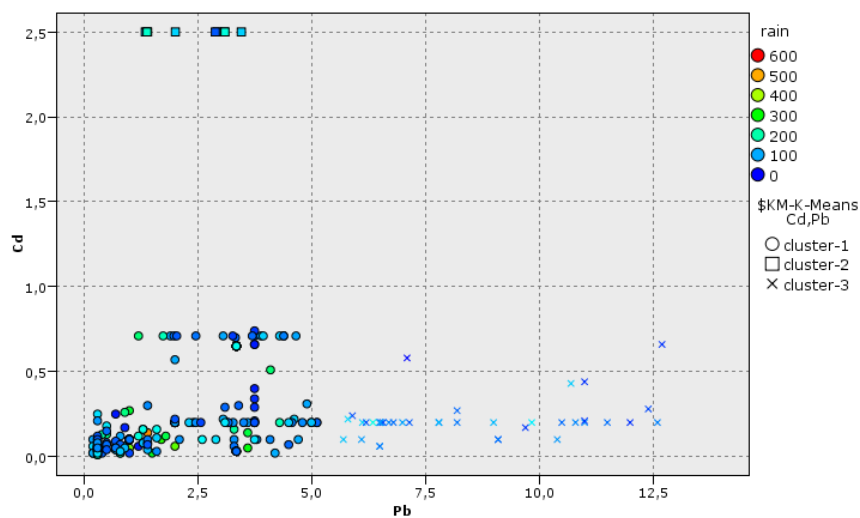
Obrázky (Obrázek 23, Obrázek 24, Obrázek 25) znázorňují vliv lokality, srážek a pH na zařazení jednotlivých měření do shluků. Na obrázku (Obrázek 23) je názorně vidět, že ve shluku 1, který je určen jako shluk s nejnižším obsahem kadmia a olova ve srážkách, převládají měření z lokality (2,3) Modrý potok a Hřibčecí. Ve shluku 2, který je charakterizován vysokým obsahem kadmia ve srážkách, je zastoupena převážně lokalita (3) Hřibčecí. Třetí shluk je charakterizován vysokým obsahem olova a jsou v něm zastoupeny všechny tři lokality.



Obrázek 23 Grafické znázornění shlukování a lokalit

Na obrázku (Obrázek 24) na základě srážek určíme shluky následujícím způsobem:

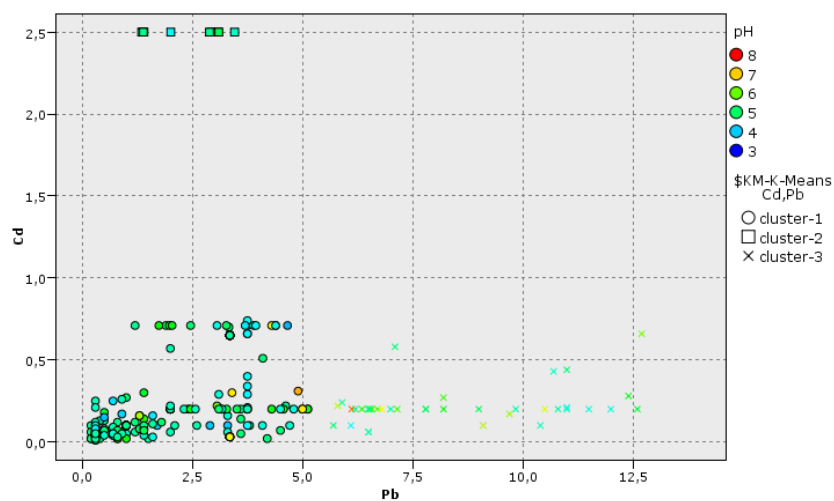
- shluk 1 je definován v intervalu 100 až 500 mm naměřených srážek,
- shluk 2 je definován v intervalu 0 až 200 mm naměřených srážek,
- shluk 3 je definován v intervalu 100 až 200 mm naměřených srážek.



Obrázek 24 Grafické znázornění shlukování a srážek

Na obrázku (Obrázek 25) na základě pH charakterizujeme shluky následujícím způsobem:

- shluk 1 je definován v intervalu 4 až 7 pH,
- shluk 2 je definován v intervalu 4 až 5 pH,
- shluk 3 je definován v intervalu 4 až 8 pH.



Obrázek 25 Grafické znázornění shlukování a pH

3.6. Dílčí závěr

Rozhodla jsem se měření podle tří shluků, které jsem si namodelovala podle obsahu kadmia a olova, klasifikovat je na základě lokality měření, naměřených srážek a pH. Zvolila jsem rozhodovací strom CART. Na obrázku (Obrázek 26) je patrné, že rozhodovací strom CRT, při shlukování do tří shluků podle dvou kovů kadmia a olova, dosahuje 81,5 % úspěšnosti na testovacích datech klasifikace shluků, vytvořených metodou K-means, na základě lokality, srážek a pH. Tuto schopnost má pouze na upravených datech. V tomto případě správně klasifikuje shluk 1, s nejnižším obsahem kadmia a olova, 115 případů ze 135 případů měření. Rozhodovací strom správně rozhodne o jednom případě z 5 případů měření, shluk 3 bude poznán třemi případy měření z 6 z daného shluku.

Results for output field cluster Cd,Pb

Comparing \$R-cluster Cd,Pb with cluster Cd,Pb

'Partition'	1_Training		2_Testing	
Correct	121	87,68%	119	81,51%
Wrong	17	12,32%	27	18,49%
Total	138		146	

Coincidence Matrix for \$R-cluster Cd,Pb (rows show actuals)

'Partition' = 1_Training		1	2	3
1		112	0	2
2		2	3	0
3		13	0	6
'Partition' = 2_Testing		1	2	3
1		115	4	3
2		3	1	0
3		17	0	3

Obrázek 26 Přehled správnosti klasifikace rozhodovacího stromu na upravených datech

Z obrázku (Obrázek 27) je patrné, že rozhodovací strom dosahuje 77,4 % úspěšnosti na testovacích datech klasifikace shluků, vytvořených metodou K-means, na základě lokality, srážek a pH. Na neupravených datech správně klasifikuje shluk 1 a 3. Shluk 2 není schopen určit, ale tuto skutečnost můžeme zanedbat, protože tento shluk můžeme označit jako výkyv. Shluk 1 klasifikuje správně ve 105 případech ze 135 případů měření. Rozhodovací strom správně rozhodne u osmi případech z 11 případů měření shluk 3, který charakterizujeme jako shluk s vyšším obsahem olova.

Results for output field \$KM-K-Means Cd,Pb

Comparing \$R-\$KM-K-Means Cd,Pb with \$KM-K-Means Cd,Pb

'Partition'	1_Training		2_Testing	
Correct	100	72,46%	113	77,4%
Wrong	38	27,54%	33	22,6%
Total	138		146	

Coincidence Matrix for \$R-\$KM-K-Means Cd,Pb (rows show actuals)

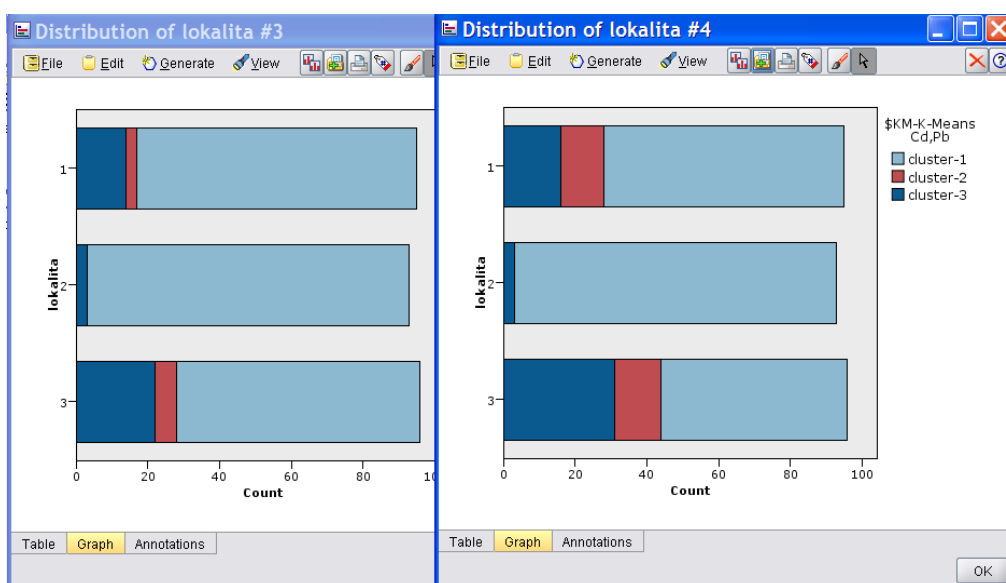
'Partition'= 1_Training	cluster-1	cluster-2	cluster-3
cluster-1	94	0	7
cluster-2	12	2	0
cluster-3	19	0	4

'Partition'= 2_Testing	cluster-1	cluster-3
cluster-1	105	3
cluster-2	11	0
cluster-3	19	8

Obrázek 27 Přehled správnosti klasifikace rozhodovacího stromu na upravených datech

3.7. Závěr kapitoly

Modely, které jsem získala modelováním upravených dat, jsem porovnávala s výsledky stejných modelů s původními daty. Z obrázku (Obrázek 28) je patrné, že převážná část měření byla zařazena do shluku 1, který je označen jako shluk s nejnižším obsahem kadmia a olova. Levý graf popisuje výsledky modelu na upravených datech a pravý graf je proveden s původními daty. Lokalita Hřibecí, označená jako 1, má v nejlepším shluku, označený jako cluster 1, 67 všech svých měření. Lokalita Mokřý potok, označená jako 2, má v nejlepším shluku, označený jako cluster 1, 90 všech svých měření. Lokalita Rýchory, označená jako 3, má v nejlepším shluku, označený jako cluster 1, 52 všech svých měření. Na základě tohoto zjištění můžeme lokalitu Modrý potok označit jako lokalitu s nejnižším obsahem kadmia a olova ve srážkách.

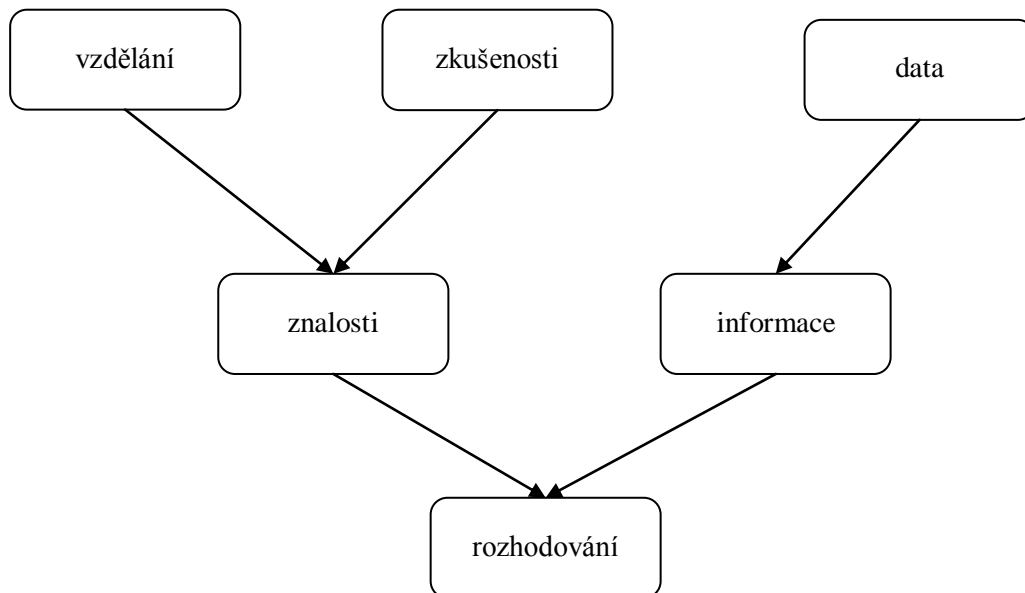


Obrázek 28 Porovnání klasifikování lokalit do shluků podle obsahu kadmia a olova

Tento problém by bylo možné řešit také faktorovou analýzou. Faktorová analýza je vícerozměrná technika k zjištění vnitřních souvislostí a vztahů a k odhalení základní struktury zdrojové matice. Ve faktorové analýze předpokládáme, že každý vstupní znak můžeme vyjádřit jako lineární kombinací nevelkého počtu společných skrytých faktorů a jediného specifického faktoru [19].

4. Modely vícekriteriálního rozhodování

Kriteria hodnocení představují hlediska zvolená účastníkem rozhodovací situace, která slouží k posuzování výhodnosti jednotlivých variant plnění dílčích cílů řešení rozhodovací úlohy. „Rozhodovací proces prováděný člověkem nezbytně vyžaduje přítomnost znalostí i dat, neboť kvalifikovaně rozhodnout bez znalostí z dané problémové oblasti, ale stejně tak rozhodovat bez konkrétních údajů o konkrétní situaci nelze.“ Prostřednictvím vzdělání a zkušenosti člověk získává znalosti. Vzdělání zahrnuje obecné učebnicové znalosti. Zkušenosti představují znalosti soukromé. Vazby mezi zkušenostmi, vzděláním, daty a rozhodováním znázorňuje obrázek (Obrázek 29). Experta dělá expertem velký potenciál znalostí získaných vlastními zkušenostmi [18].



Obrázek 29 Data a znalosti v procesu rozhodování [18]

Součástí řešení úlohy jsou často složité rozhodovací situace. Procesem rozhodování rozumíme akt výběru jedné z několika možných alternativ. Člověk nebo kolektiv, který provádí výběr optimální alternativy, se nazývá subjekt rozhodování nebo účastník rozhodovací situace. Účastník rozhodovací situace, jehož rozhodnutí je uvědomělé, zaměřené k určitému cíli a které při volbě vhodné varianty využívá všech objektivně dostupných informací o důsledcích volby alternativy, se nazývá racionální rozhodování. Iracionální rozhodování je na základě dostupných informací, vlastních zkušeností, znalostí a intuice.[5].

Výraz optimální rozhodnutí, optimální alternativa, nejlepší řešení zpravidla předpokládá, že je intuitivně jasné, co přívlastek optimální nebo nejlepší znamená. Při bližším zkoumání se však ukáže, že není tak jednoduché tento pojem přesněji definovat. Zdrojem obtíží je, že pojem podstatně závisí na typu situace, v níž se rozhoduje, přičemž klasifikace těchto situací je již problémem sama o sobě [14].

Rozhodovací situaci chápeme jako soubor vzájemně vázaných podmínek určujících nebo spolupracujících rozhodnutí. Model by měl zachycovat charakteristiky rozhodovatelů, charakteristiky podstatného okolí, které ovlivňují výsledky rozhodování. Při řešení složitých rozhodovacích úloh rozhodovatel zjišťuje některé charakteristiky rozhodovací situace během iterativního procesu [5].

Klasifikace rozhodovacích procesů:

- rozhodování při jistotě,
- rozhodování při riziku,
- rozhodování při nejistotě.

Rozhodování za jistoty je, když známe veškeré související informace ovlivňující rozhodování. O tomto rozhodování nebo vícekriteriálním rozhodování za jistoty a rizika se více zabývá Fotr ve své knížce viz [6]. Vícekriteriální rozhodování za nejistoty zpracoval Carlsoon ve své publikaci v kapitole Fuzzy Multicriterial Decision Making viz [4].

Rozhodování zahrnuje následující kroky [23]:

- vyhledávání vhodných příležitostí pro rozhodování,
- nalezení možných způsobů řešení rozhodovacích situací,
- výběr mezi alternativními situacemi.

Činnosti rozhodovacího procesu, které jsou vzájemně závislé a návazné, označujeme jako fáze nebo etapy. Rozhodovací proces lze rozdělit na etapy podrobněji nebo seskupeně. Příklad agregované dekompozice rozhodovacího procesu [23]:

- analýzy okolí (intelligence aktivity),
- návrh řešení (design aktivity),
- volba řešení (choice aktivity),
- kontrola výsledku (review aktivity).

Metody rozhodování:

- vícekriteriální metody rozhodování,
- Buridanův osel znamená povinnost se rozhodnout,
- Occamova břitva znamená, že zvolím tu jednodušší cestu z množiny možných variant,
- zdravý selský rozum, když člověk při řešení každodenních problémů vychází ze svých zkušeností a názoru.

4.1. Vícekriteriální rozhodování

Vícekriteriální rozhodování představuje rozhodování za jistoty. Rozhodování za jistoty je reálné za takových znalostí variant, kritérií a metod vyhodnocení, které se vyznačují minimální, pro hodnotitele zanedbatelnou rozptýleností a vysokými znalostmi variant [6].

Důležitým klasifikačním hlediskem je způsob zadání množiny přípustných variant. Je-li množina přípustných variant zadána ve formě konečného seznamu, mluvíme o úloze vícekriteriálního hodnocení variant. Je-li množina přípustných variant vymezena souborem podmínek, které rozhodovací alternativy musí splňovat, aby byly přípustné, mluvíme o úloze vícekriteriálního programování [15].

Metody vícekriteriálního hodnocení variant představují:

- metodu úměrného skórování,
- metody založené na párovém srovnání variant např. Saatyho metoda.

Převážná část metod vícekriteriálního hodnocení variant vyžaduje nejprve stanovit váhy kritérií:

- metoda alokace 100 bodů,
- metoda párového srovnání např. Fullerova metoda.

Fullerova metoda slouží k párovému porovnání kritérií vzájemně mezi sebou. Je to jednoduchá metoda a slouží pouze k černobílému vidění světa. Jednička určuje, jsou-li kriteria v řádku důležitější než kriteria ve sloupci. Nula naopak kriteria ve sloupci jsou důležitější, než kriteria v řádku. První Fullerův trojúhelník může nabývat pouze hodnot 1,0. Tato metoda zjišťuje počet preferencí jednotlivých kritérií vůči sobě navzájem, platí pro Fullerův trojúhelník. Výpočet provádíme dle následujících vztahů:

$$v_i = \frac{P_i}{\sum P_i} \quad (4)$$

kde: v_i – výpočet vah, P_i – součet preferencí v řádku.

Bohužel pomocí této metody dochází k vyloučení nejméně preferovaného kritéria, proto se musí provést vyrovnání. Fullerova metoda s vyrovnáním nulových vah

$$v_i = \frac{P_i}{\frac{n * (n-1)}{2}} \quad (5)$$

kde: v_i – výpočet vah, P_i – preference dle skóre, n - počet kritérií.

Saatyho metoda slouží ke stanovení vah kritérií nebo stanovení hodnot variant. Prvním krokem je sestavení hodnotící Saatyho stupnice relativních důležitostí:

- 1 kritérium je stejně důležité,
- 3 kritérium je méně než středně důležité,
- 5 kritérium je středně důležité,
- 7 kritérium je více než středně důležité,
- 9 kritérium má absolutní důležitost.

Druhým krokem je sestavení Saatyho matice S . Pro prvky Saatyho matice platí:

$$S = \begin{bmatrix} s_{11} & \dots & s_{1i} \\ \dots & & \dots \\ s_{j1} & \dots & s_{ji} \end{bmatrix} \text{ pokud platí } i = j \quad (6)$$

$$s_{ii} = 1 \quad (6)$$

$$s_{ji} = \frac{1}{s_{ij}} \quad (6)$$

kde: S – Saatyho matice, s_{ji}, s_{ij} - prvky matice.

$$S * \sigma = \sigma * \max \lambda \quad (7)$$

kde: S – Saatyho matice, σ – vektor vlastních čísel, $\max \lambda$ – maximální vlastní číslo.

Saatyho metoda používá pro výpočet vah kritérií a alternativ vlastních čísel a vlastních vektorů, algoritmus výpočtu vlastních čísel a vlastního vektoru je řešen v publikaci [12] v kapitole Praktické provedení metody hlavních komponent.

Index konzistence (CI) pro danou Saatyho matici představuje kontrolu správnosti sestavení matice. Pokud platí, že $CI \leq 0,1$, pak Saatyho matice byla sestavena správně. Vypočítá se pomocí vztahu [24] :

$$CI = \frac{\max \lambda - k}{k - 1} \quad (8)$$

kde: k – počet kritérií, $\max \lambda$ – maximální vlastní číslo.

Dalším možným ověřením správnosti Saatyho matice je konzistenční průměr (CR) výpočetní vztah pro matici prvního řádu [24] :

$$CR = \frac{CI}{RI} \quad (9)$$

kde: CR - konzistenční průměr, CI – index konzistence, RI – průměrný index.

Konzistenční index průměr (CR) výpočetní vztah pro matice druhého řádu [27] :

$$\overline{CR} = \frac{\sum_i w_i * CI_i}{\sum_i w_i * RI_i} \quad (9)$$

kde: CR - konzistenční průměr, CI – index konzistence, RI – průměrný index.

Průměrný index zjistíme na základě následující tabulky (Tabulka 5) podle příslušného počtu řádků matice.

Tabulka 5 Průměrný index [27]

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0,58	0,9	1,12	1,24	1,32	1,41	1,45	1,49

Pro výpočet vah kritérií a variant v AHP můžeme také použít metody aritmetického průměru, geometrického průměru vycházejících ze Saatyho matic prvního a druhého řádu.

Aritmetický průměr vypočítáme podle následujícího vztahu:

$$v_i = \sum \frac{s_{ij}}{k} \quad (8)$$

kde: k – počet kritérií, s_{ij} – hodnota Saatyho matice.

Geometrický průměr vypočítáme podle následujícího vztahu:

$$v_i = \sqrt[k]{\prod s_{ij}} \quad (11)$$

kde: k – počet kritérií, s_{ij} – hodnota Saatyho matice.

Váhy kritérií v_i jsou hodnoty vlastního vektoru vypočítaných ze Saatyho matice prvního řádu. Váhy ohodnocení variant h_i^j pro jednotlivá kritéria představují matice z vlastních vektorů, vypočítaných na základě Saatyho matice druhého řádu. Ohodnocení nejlepší alternativy určíme jako maximální H^j . Výsledky jednotlivých alternativ určíme podle vzťahu:

$$H^j = \sum v_i * h_i^j \quad (7)$$

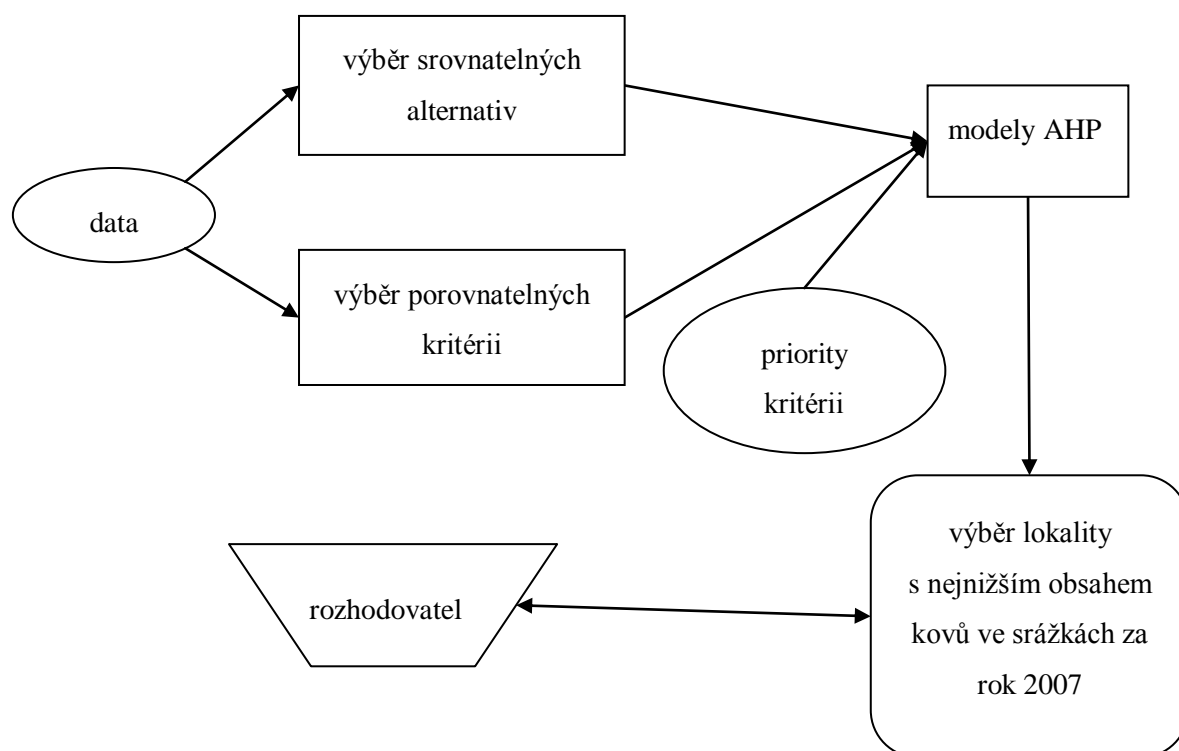
kde: H^j – celkové ohodnocení, v_i – váhy jednotlivých kritérií, h_i^j - dílčí ohodnocení j-té varianty vzhledem k i-tému kritériu.

Následující obrázek (Obrázek 30) znázorňuje postup řešení rozhodovacího problému. Na obrázku je určen algoritmus mé práce.

Vstupem je celá datová matice, z které je potřeba vybrat alternativy a kritéria, výstupem tohoto kroku je tabulka (Tabulka 6). Jako alternativy jsem zvolila tři lokality a jako kritéria jsem určila kovy. Priority k jednotlivým kritériím jsem přiřadila, na základě publikací viz [22], [8], [10]. Preference kritérií jsem uvedla v tabulce (Tabulka 7). Vybrala jsem vícekritériální metodu AHP, jejímž výsledkem je doporučení pro konečného rozhodovatele nebo skupinu rozhodvatelů.

Tabulka 6 Průměrné hodnoty roku 2007 pro vybrané lokality

	K1	K2	K3	K4	K5	K6	K7
	Al [ug/l]	Cd [ug/l]	Fe [ug/l]	Mn [ug/l]	Ni [ug/l]	Pb [ug/l]	Zn [ug/l]
HHRIB	32,22	0,15	17,11	16,02	1,23	3,50	13,44
HMOPB	15,00	0,08	18,45	7,45	0,73	2,94	49,71
HRYCB	27,10	0,12	23,91	8,59	1,10	1,57	74,81



Obrázek 30 Algoritmus AHP

Tabulky (Tabulka 6, Tabulka 7) představují vstupní údaje a jsou pro rozhodovací modely třívrstvý model AHP i čtyřvrstvý model AHP stejné.

Tabulka 7 Rozhodovací kritéria a jejich priority

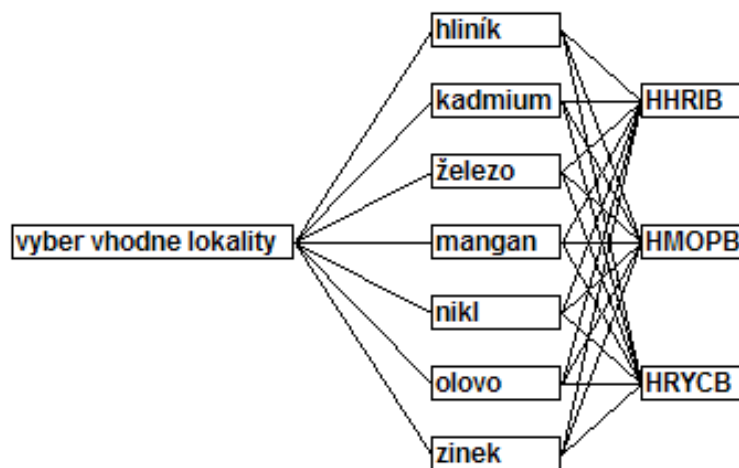
Kritérium		Priorita
K1	hliník	6
K2	kadmium	2
K3	železo	3
K4	mangan	7
K5	nikl	5
K6	olovo	1
K7	zink	4

4.2. AHP tři vrstvy

Tento model se zabývá rozhodovacím procesem z oblasti životního prostředí. Pro rozhodování jsem zvolila třívrstvý analytický hierarchický proces, kterým bych chtěla rozhodnout o lokalitě s nejmenším obsahem prašného spadu kovů v naměřených srážkách. Vybrané lokality jsou Hřibčící, Modrý potok, Rýchory znázorňuje obrázek (Obrázek 3). Lokality jsou v okrese Trutnov.

Hierarchie je specifický typ systému, který se skládá z předpokladů o objektech, jež identifikujeme, shlukujeme do množin na dvě skupiny. První skupina ovlivňuje druhou skupinu. Prvky všech vrstev hierarchie si myslí, že jsou nezávislé [24] .

Nadefinujeme hierarchickou strukturu systému. Strukturu tvoří tři vrstvy $S = L_1, L_2, L_3$. První vrstva obsahuje cíl neboli nejlepší alternativu, kterou označíme $L_1 = OV$. Druhá vrstva zahrnuje kritéria potřebná k rozhodnutí, značíme $L_2 = K_1, K_2, K_3, K_4, K_5, K_6, K_7$. Ve třetí vrstvě se nacházejí jednotlivé varianty, označíme $L_3 = A_1, A_2, A_3$. Jednotlivé vrstvy jsou mezi sebou propojeny systémem. Každý prvek z vrstvy je spojen s každým prvkem sousedící vrstvy. Celková struktura a její vrstvy jsou ukázané na obrázku (Obrázek 31).



Obrázek 31 Systém hierarchického třívrstvého modelu

Prvky tohoto modelu jsou:

- Ve třetí vrstvě jsou jednotlivé alternativy, v tomto případě to jsou jednotlivé lokality měření srážek (A1) Hřiběcí, (A2) Modrý potok a (A3) Rýchory.
- Prvky druhé vrstvy znázorňují jednotlivá kritéria pro rozhodování, jsou to kovy sledované v pravidelných intervalech v lokalitách. Při měření srážek sledujeme hodnotu následujících látek, které slouží k rozhodování hliník (K1), kadmium (K2), železo (K3), mangan (K4), nikl (K5), olovo (K6), zinek (K7). Kritéria jsou minimalistická.
- V první vrstvě je pouze jeden prvek a to optimální řešení dle zvolených kritérií a jejich vah, které je cílem našeho rozhodnutí. Je shodné s jednou z alternativních variant v nejnižší vrstvě.

Vazby existují mezi prvky sousedících vrstev, jsou číselně ohodnocené a nazýváme je váhou. Vazby mezi prvky první a druhé vrstvy jsou ohodnocené váhou přiřazené k určitému kritériu. Váhy zapisujeme do vektoru, značíme w_m , tyto vazby spojují kritérium a optimální alternativu. Váhy tvoří sedmi prvkový vektor viz tabulka (Tabulka 10). Každý řádek znázorňuje jeden vektor pro danou metodu. Pro výpočet vah kritérii pro Fullеровu metodu použijí Fullerův trojúhelník viz tabulka (Tabulka 8). Saatyho matici viz tabulka (Tabulka 9) pro výpočet vah kritérii pomocí Saatyho metody, aritmetického a geometrického průměru. Celý algoritmus Saatyho metody je v příloze (Příloha 1).

Vazby mezi prvky druhé a třetí vrstvy představují váhy jednotlivých variant pro jednotlivá kritéria. Váhy značíme $v_{m,n}$ a zapisujeme do matice V . Váha nebo-li ohodnocené vazby jsou mezi každou alternativou a každým kritériem. Vznikne nám matice o třech sloupcích a sedmi řádcích tuto matici sestavíme ze Saatyho matic druhého řádu viz příloha (Příloha 3)

Tabulka 8 Fullerův trojúhelník

	K1	K2	K3	K4	K5	K6	K7	Počet 1	Počet 0	Počet	v_i
K1		1	1	1	1	0	1	5	0	5	0,227
K2			0	1	0	0	0	1	0	1	0,045
K3				1	1	0	1	3	1	4	0,182
K4					0	0	0	0	0	0	0,000
K5						0	0	0	2	2	0,091
K6							1	1	6	7	0,318
K7								0	3	3	0,136
										22	1,000

Tabulka 9 Saatyho matice

	K1	K2	K3	K4	K5	K6	K7
K1	1,00	0,11	0,33	0,20	0,11	0,22	0,14
K2	9,00	1,00	8,00	5,00	3,00	5,00	3,00
K3	3,00	0,12	1,00	0,33	0,20	0,11	0,14
K4	5,00	0,13	0,33	1,00	0,20	0,33	0,20
K5	9,00	0,33	8,00	5,00	1,00	5,00	3,00
K6	5,00	0,20	5,00	3,00	0,20	1,00	0,20
K7	7,00	0,33	7,00	5,00	0,33	5,00	1,00

Aritmetický průměr se vypočítá po jednotlivých řádcích na základě hodnotící Saatyho matice. Tím vypočítáme vektor vah w_i pro AHP. Průměrem řádku Saatyho matice druhého řádu dostaneme hodnoty vah v_{ij} pro AHP zapsaných do matice. Vynásobením vektoru a matice získáme vektor výsledných hodnot pro tři alternativy.

Geometrický průměr se vypočítá pro jednotlivé řádky hodnotící Saatyho matice. Tím vypočítáme vektor vah w_i pro AHP. Zprůměrováním Saatyho matice druhého řádu dostaneme hodnoty vah v_{ij} pro AHP zapsaných do matice. Vynásobením vektoru a matice získáme vektor výsledných hodnot pro tři alternativy.

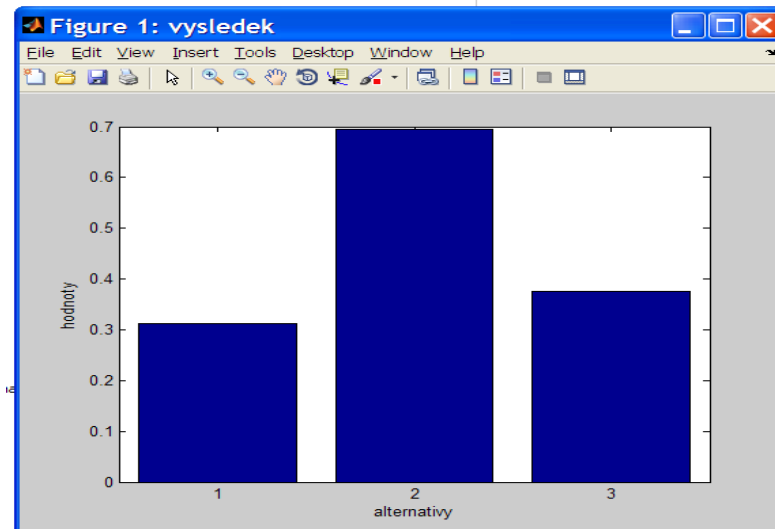
4.3. Verifikace modelu v prostředí MATLAB a MS Excel

Pro stanovení vah analyticky hierarchického procesu jsem si vybrala metodu aritmetického průměru, geometrického průměru a Fullerovu metodu, které jsem zpracovala v programu Excel. Dále jsem pro stanovení váhy pomocí Saatyho metody použila prostředí MATLAB. Pro výpočet metody aritmetický, geometrický průměr a Saatyho metodu jsem použila Saatyho matice prvního a druhého řádu. Pro výpočet Fullerovy metody bez vyrovnání nulových vah a Fullerovy metody s vyrovnáním nulových vah jsem použila Fullerův trojúhelník.

Tabulka 10 Váhy kritérií w_m pro jednotlivé metody

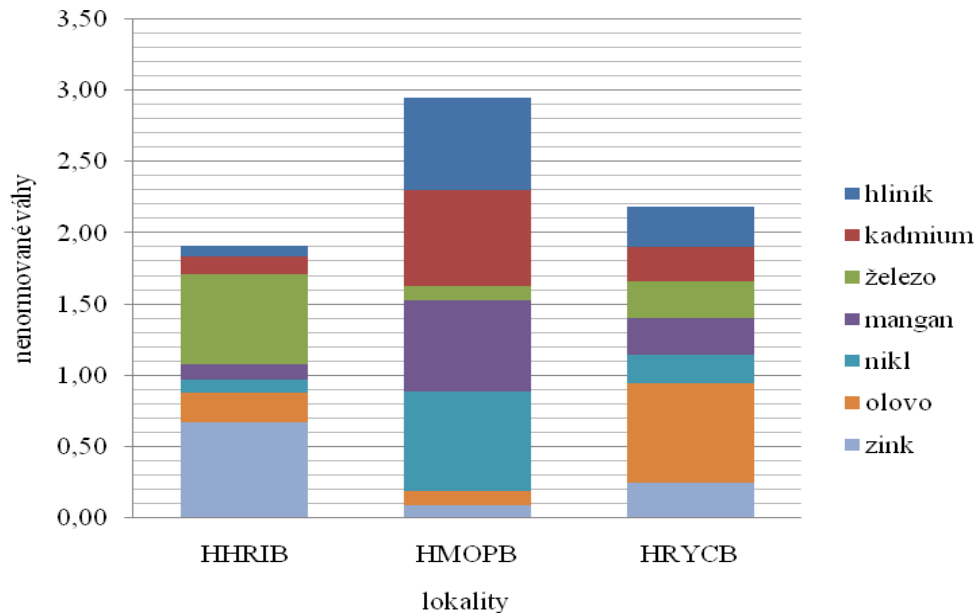
	K1	K2	K3	K4	K5	K6	K7
aritmetický průměr	0.020	0.370	0.032	0.042	0.267	0.084	0.184
geometrický průměr	0.020	0.370	0.032	0.042	0.267	0.084	0.184
Saatyho metoda	0.020	0.370	0.032	0.042	0.267	0.084	0.184
Fullerova metoda a	0.227	0.045	0.182	0.000	0.091	0.318	0.136
Fullerova metoda b	0.077	0.231	0.077	0.128	0.179	0.128	0.179

Graf na obrázku (Obrázek 32) ukazuje, že na základě zvolených vah je pro rozhodovací problém nejlepší alternativa Modrý potok. Rozhodovací proces nám doporučil jako lokalitu s nejnižším obsahem kovů v prašném spadu lokalitu Modrý potok na 50 %. Hřibčecí a Rýchory mají podobné doporučení na 24 a 26 %.



Obrázek 32 Saatyho metoda v prostředí MATLAB

Graf na obrázku (Obrázek 33) znázorňuje, kolik každé kritérium přispívá ke konečnému umístění dané alternativy pomocí Saatyho metody. Na vybrání Mokrého potoka jako vítězné alternativy se nejvíce podílí obsah niklu, hliníku, manganu a kadmia v prašném spadu srážek. Na umístění lokality Rýchory má největší podíl obsah olova v prašném spadu srážek. Na umístění lokality Hřibčecí má největší podíl obsah železa a zinku v prašném spadu srážek.



Obrázek 33 Váhy kritérií jednotlivých alternativ pomocí Saatyho metody

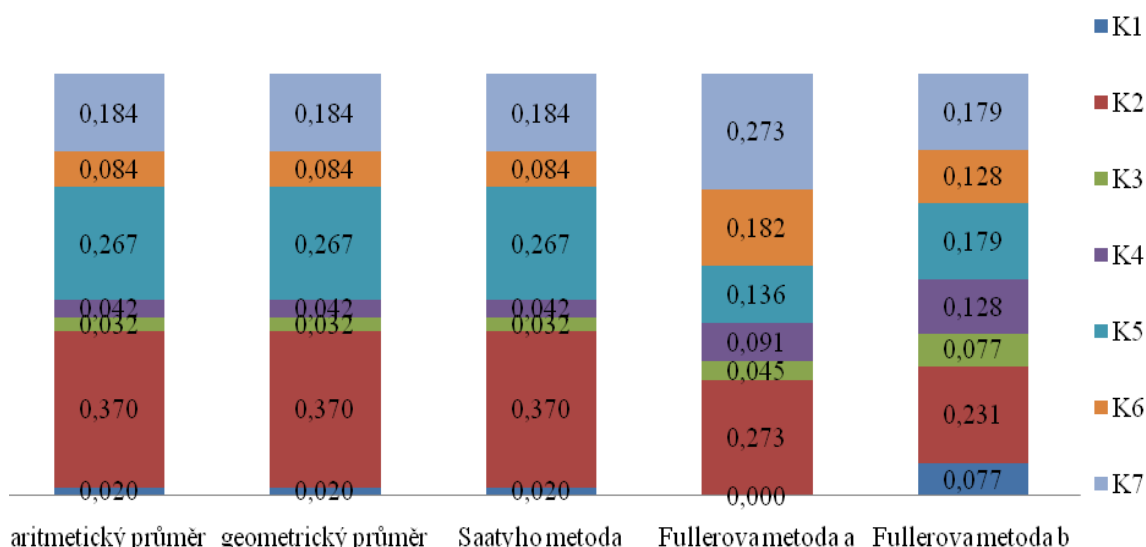
4.4. Dílčí závěr

Jak názorně ukazuje závěrečná tabulka (Tabulka 11), doporučuji lokalitu Modrý potok na základě uvedených metod, jako lokalitu s nejnižším obsahem kovů ve srážkách. Jak ukazují použité metody aritmetický průměr, geometrický průměr, Saatyho metoda, Fullerova metoda bez vyrovnání nulových vah a Fullerova metoda s vyrovnáním nulových vah tak i prostředí CDP.

Tabulka 11 Výsledky jednotlivých metod

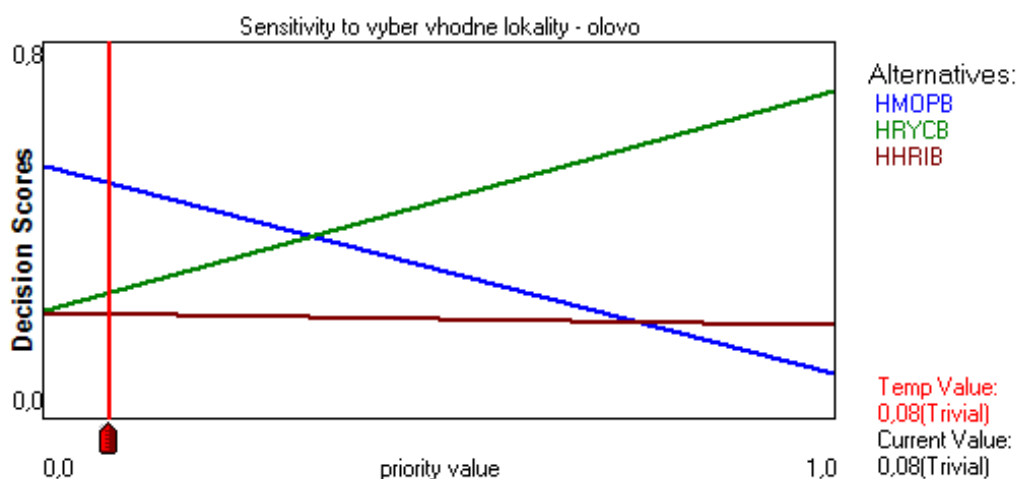
	HHRIB	HMOPB	HRYCB
aritmetický průměr	0,373	0,482	0,294
geometrický průměr	0,327	0,503	0,272
Saatyho metoda	0,238	0,503	0,272
Fullerova metoda a	0,248	0,524	0,227
Fullerova metoda b	0,247	0,434	0,319

Tento graf na obrázku (Obrázek 34) ukazuje důležitost vah kritérií v jednotlivých metodách. Jednotlivé metody se v hodnotách vah příliš neliší. Metoda aritmetický, geometrický průměr a Saatyho metoda mají výsledky vah totožné. Podobné je to u vyrovnané Fullerovy metody.



Obrázek 34 Normované váhy kritérií pro vybrané metody

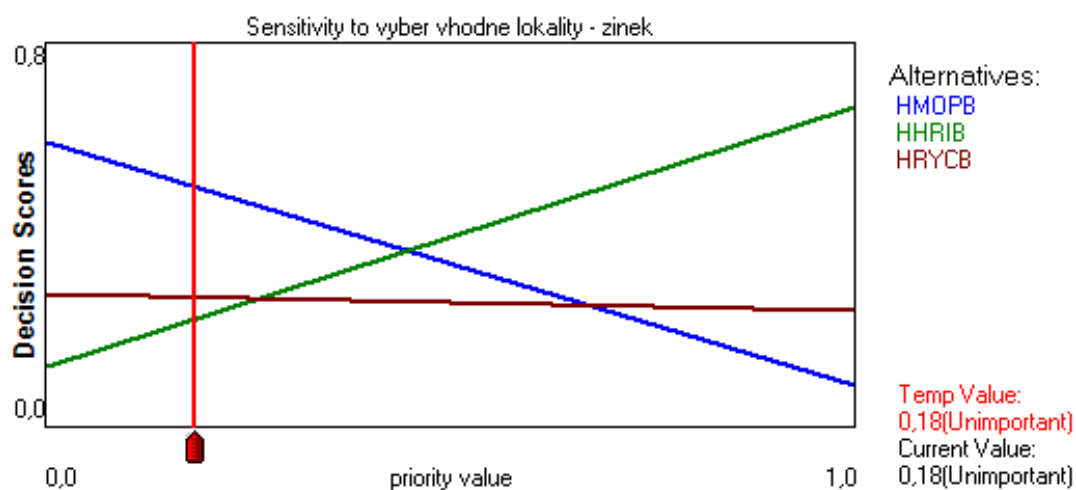
Následující graf na obrázku (Obrázek 35) popisuje, jak by se změnil výsledek jednotlivých alternativ, kdyby se změnila váha kritéria olova. Ať by se váha tohoto kritéria změnila jakkoliv lokalita Hříběcí by nebyla jako nejlepší alternativa nikdy. Pokud by se váha tohoto kritéria zvětšila nad 34 %, vítěznou alternativou by se stala alternativa Rýchory na místo Modrého potoka.



Obrázek 35 Výsledné rozhodnutí ovlivněné kritériem olovo

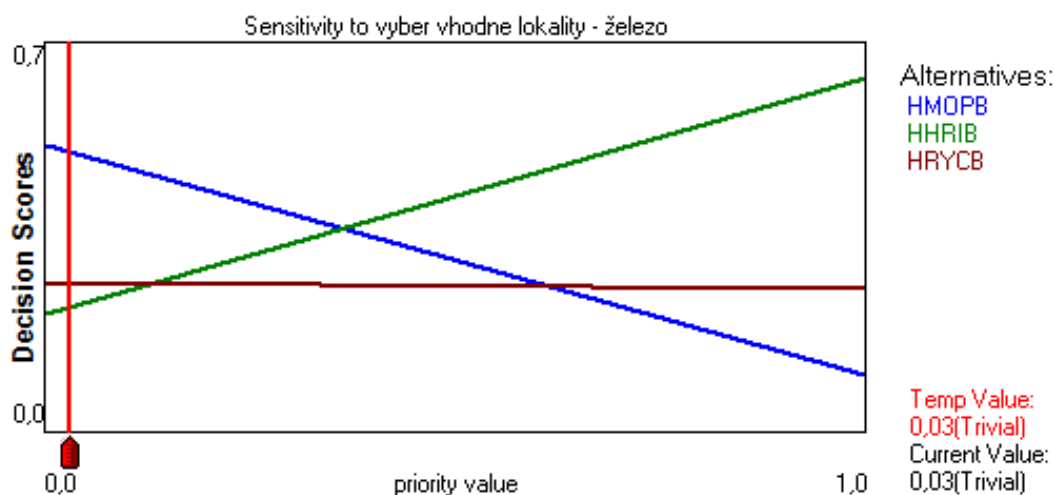
Graf na obrázku (Obrázek 36) znázorňuje, jak by se změnil výsledek jednotlivých alternativ, kdyby se změnila váha kritéria zinek. Ať by se váha toho kritéria změnila jakkoliv lokalita Rýchory by nebyla

jako nejlepší alternativa nikdy. Pokud by se váha tohoto kritéria zvětšila nad 45 %, byla by vítěznou alternativou Hříběcí na místo Modrého potoka.



Obrázek 36 Výsledné rozhodnutí ovlivněné kritériem zinek

Graf na obrázku (Obrázek 37) popisuje, jak by se změnil výsledek jednotlivých variant, kdyby se změnila váha kritéria obsahu železa. Váha tohoto kritéria je nízká, kdyby se váha tohoto kritéria zvedla nad 40 %, stala by se vítěznou lokalitou alternativa Hříběcí.



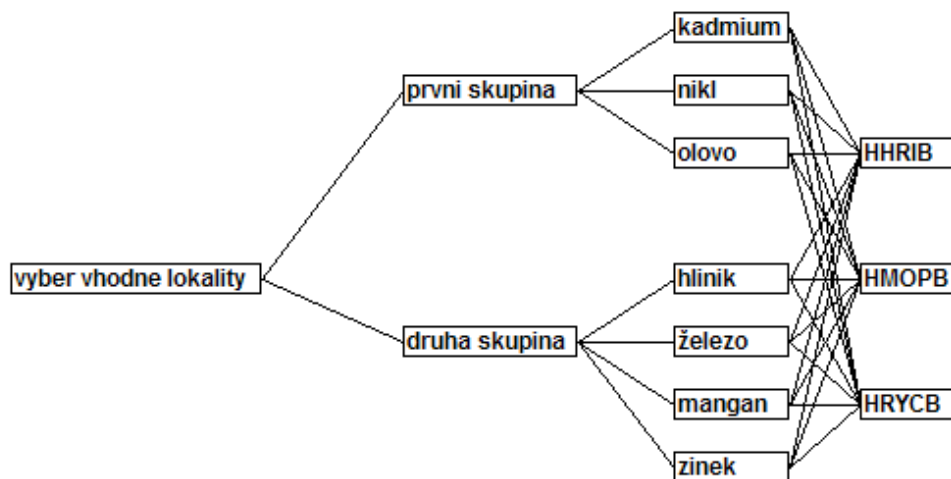
Obrázek 37 Výsledné rozhodnutí ovlivněné kritériem železo

4.5. AHP čtyři vrstvy

Tento model se zabývá stejně jako předchozí model rozhodovacím procesem z oblasti životního prostředí. Tentokrát zvolím pro rozhodování čtyřvrstvý analytický hierarchický proces, kterým bych chtěla rozhodnout o lokalitě s nejmenším obsahem prašného spadu kovů v naměřených srážkách. Zajímá mě, zdali se nejlepší lokalitou stane zase Modrý potok. Vybrané lokality jsou Hřiběcí, Modrý potok, Rýchory znázorňuje obrázek (Obrázek 3). Lokality jsou v okrese Trutnov.

Hierarchie je specifický typ systému, který se skládá z předpokladů o objektech, které identifikujeme, shlukujeme do množin na dvě skupiny. První skupina ovlivňuje druhou skupinu. Prvky všech vrstev hierarchie si myslí, že jsou nezávislé [24] .

Nadefinujeme hierarchickou strukturu systému. Strukturu tvoří čtyři vrstvy $S = L_1, L_2, L_3, L_4$. První vrstva obsahuje cíl neboli nejlepší alternativu, označujeme $L_1 = OV$. Druhá vrstva zahrnuje subkritéria značíme $L_2 = S_1; S_2$. Třetí vrstva obsahuje kritéria potřebná k rozhodnutí, značíme potřebná k rozhodnutí, značíme $L_3 = K_1; K_2; K_3; K_4; K_5; K_6; K_7$. Ve čtvrté vrstvě se nacházejí jednotlivé varianty, označíme $L_4 = A_1; A_2; A_3$. Jednotlivé vrstvy jsou mezi sebou propojeny systémem, každý prvek z vrstvy je spojen s každým prvkem sousedící vrstvy. Celková struktura a její vrstvy jsou ukázané na obrázku (Obrázek 38).



Obrázek 38 Systém hierarchický čtyřvrstvého modelu

Prvky tohoto modelu jsou:

- Ve čtvrté vrstvě jsou jednotlivé varianty, v tomto případě to jsou jednotlivé lokality měření srážek (A1), (A2) a (A3).
- Prvky třetí vrstvy představují subkritéria první skupina představuje těžké kovy, druhá skupina seskupuje ostatní kovy.
- Druhá vrstva obsahuje jednotlivá kritéria pro rozhodování, jsou to kovy sledované v pravidelných intervalech v lokalitách. Při měření srážek sledujeme hodnotu následujících látek, které slouží k rozhodování hliník (K1), kadmium (K2), železo (K3), mangan (K4), nikl (K5), olovo (K6) a zinek (K7). Kritéria jsou minimalistická.
- V první vrstvě je pouze jeden prvek a to je optimální řešení dle zvolených kritérií a jejich vah, které je cílem našeho rozhodnutí. Je shodné s jednou z alternativních variant v nejnižší vrstvě.

4.6. Verifikace modelu v prostředí MATLAB

Pro model tří vrstev AHP jsem pro výpočet vah kritérií vybrala více druhů metod. Pro čtyřvrstvý model AHP použiji pouze Saatyho metodu, protože se ukázalo, že ostatní metody si jsou velice podobné. Postup je stejný jako u předchozího modelu. Tabulka (Tabulka 12) představuje Saatyho matici prvního řádu rozdělenou pro první a druhou skupinu. Celý algoritmus Saatyho metody je v příloze (Příloha 2).

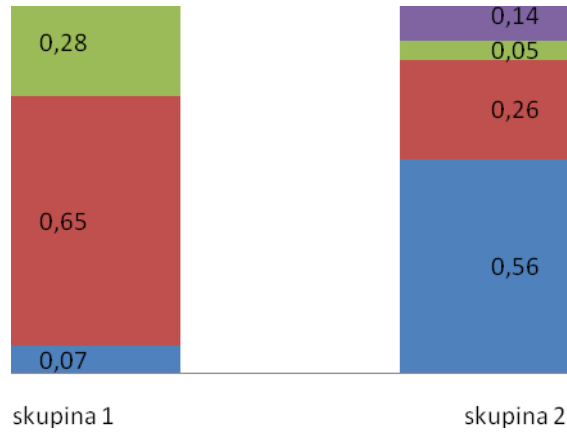
Tabulka 12 Saatyho matice prvního řáda

	K2	K5	K6
K2	1,00	0,14	0,20
K5	7,00	1,00	3,00
K6	5,00	0,33	1,00

	K1	K3	K4	K7
K1	1,00	3,00	7,00	5,00
K3	0,33	1,00	5,00	3,00
K4	0,14	0,20	1,00	0,20
K7	0,20	0,33	5,00	1,00

Graf na obrázku (Obrázek 39) znázorňuje, kolik každé kritérium přispívá ke konečnému umístění dané alternativy pomocí Saatyho metody. První skupina obsahuje prvky kadmium, které má váhu 0,28,

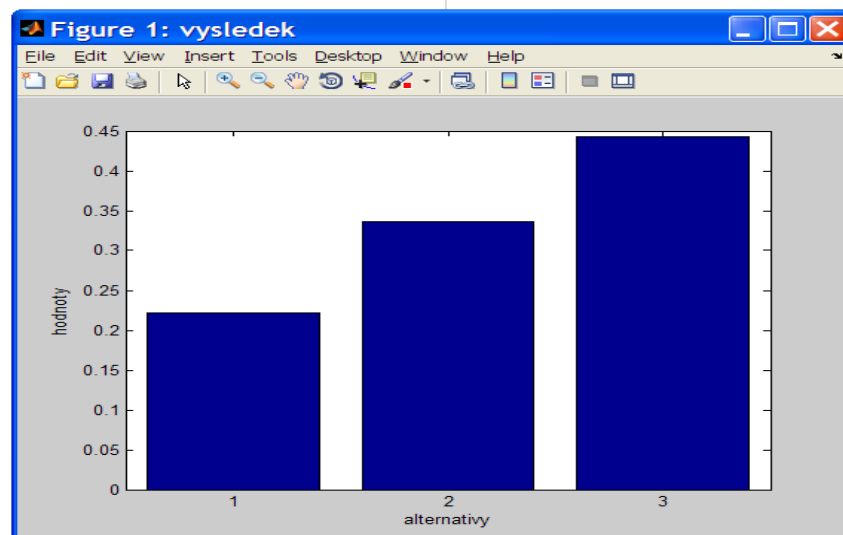
nikl, které má váhu 0,65, olovo, jehož váha je 0,07. Druhá skupina obsahuje prvky hliník, mající váhu 0,56, železo, jehož váha je 0,26, mangan, který má váhu 0,05 a zinek, který má váhu 0,14.



Obrázek 39 Normované váhy kriterií

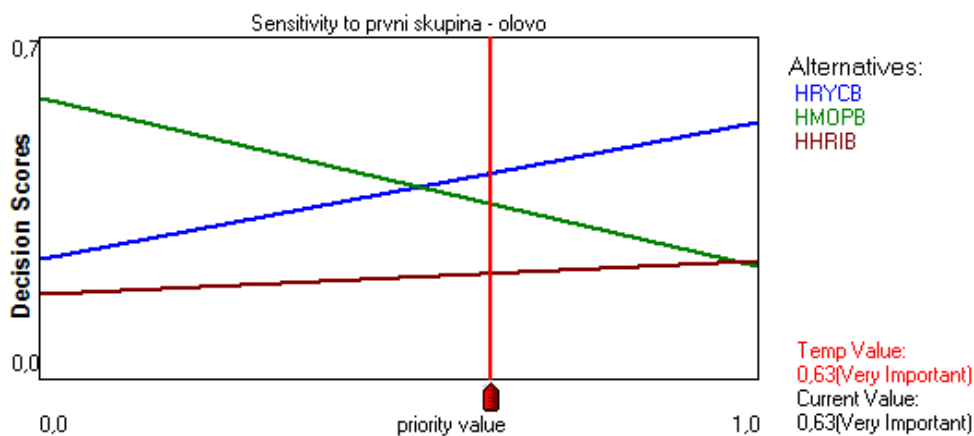
4.7. Dílčí závěr

Graf na obrázku (Obrázek 40) ukazuje, že na základě zvolených vah je pro rozhodovací problém nejlepší alternativa druhá. Rozhodovací proces nám doporučil jako lokalitu s nejnižším obsahem kovů v prašném spadu lokalitu Rýchory na 44.34 %. Modrý potok a Hřiběcí jsou velice vyrovnané doporučení na 32.55 % a 23.11 %.



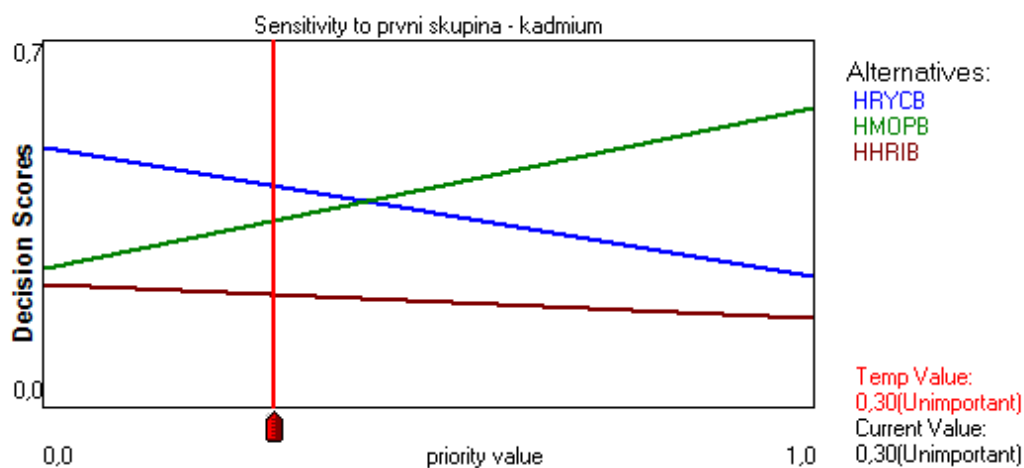
Obrázek 40 Výsledek Saatyho metody v prostředí MATLAB

Graf na obrázku (Obrázek 41) popisuje, jak by se změnil výsledek jednotlivých variant, kdyby se změnila váha kritéria obsahu olovo. Váha tohoto kritéria je nízká, kdyby se váha tohoto kritéria snížila pod 53 %, stala by se vítěznou lokalitou alternativa Modrý potok.

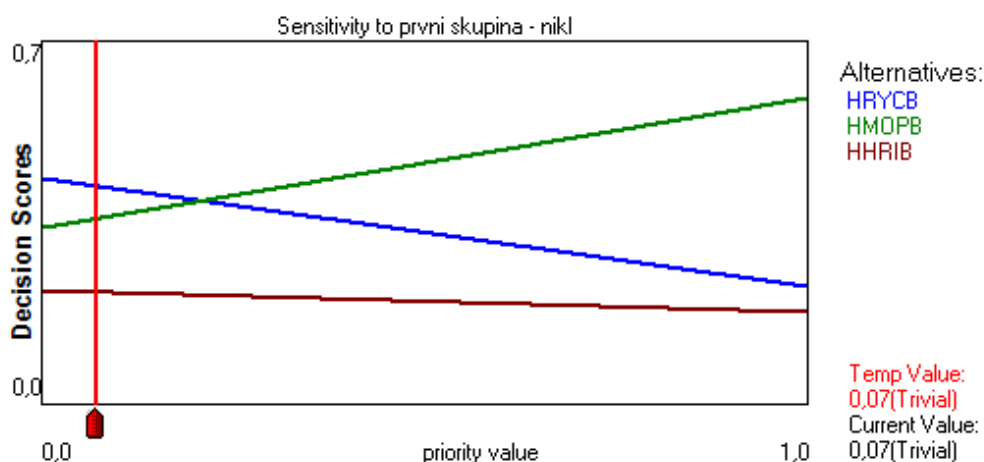


Obrázek 41 Výsledné rozhodnutí ovlivněné kritériem olovo

Tento graf na obrázku (Obrázek 42) znázorňuje, jak by se změnil výsledek jednotlivých variant, kdyby se změnila váha kritéria obsahu kadmia. Váha tohoto kritéria je nízká, kdyby se váha tohoto kritéria zvedla nad 41 %, stala by se vítěznou lokalitou alternativa Modrý potok.

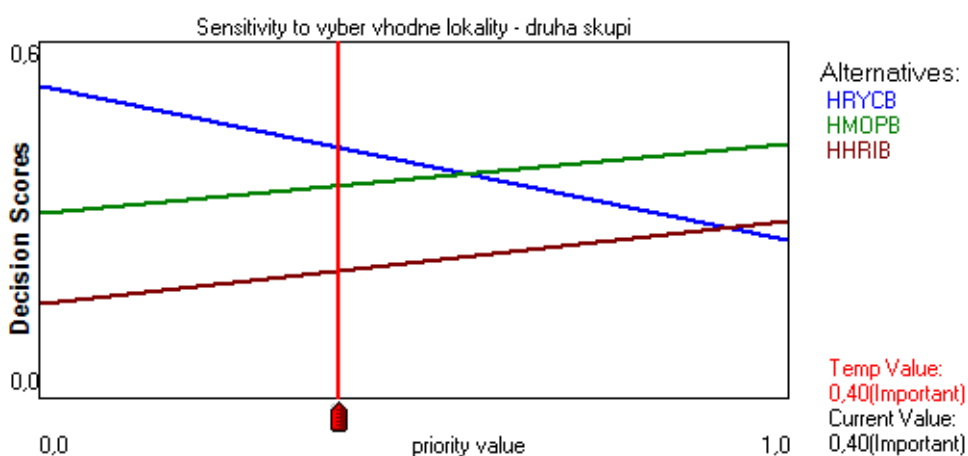


Obrázek 42 Výsledné rozhodnutí ovlivněné kritériem kadmium



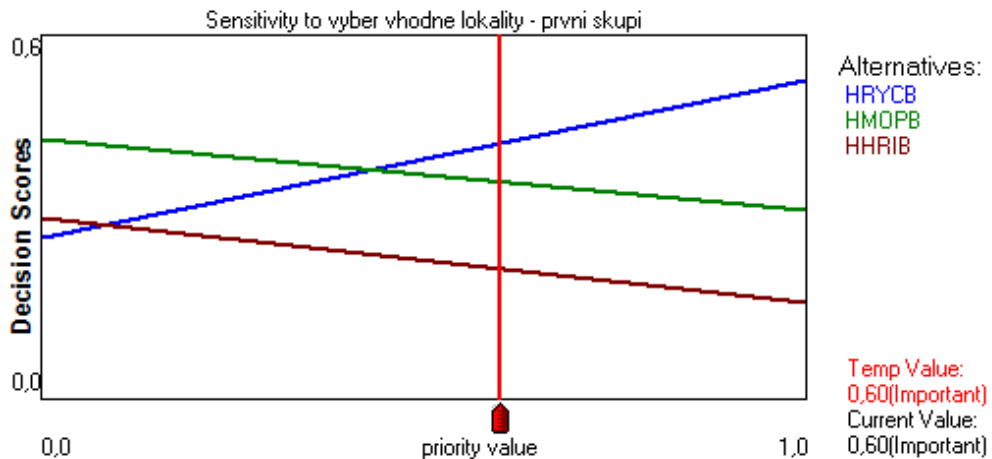
Obrázek 43 Výsledné rozhodnutí ovlivněné kritériem nickl

Graf na obrázku (Obrázek 43) popisuje, jak by se změnil výsledek jednotlivých variant, kdyby se změnila váha kritéria obsahu niklu. Váha tohoto kritéria je nízká, kdyby se váha tohoto kritéria zvedla nad 21 %, stala by se výslednou vítěznou lokalitou alternativa Modrý potok. Podle tohoto kritéria by se lokalita Hřiběcí nikdy nestala vítěznou alternativou.



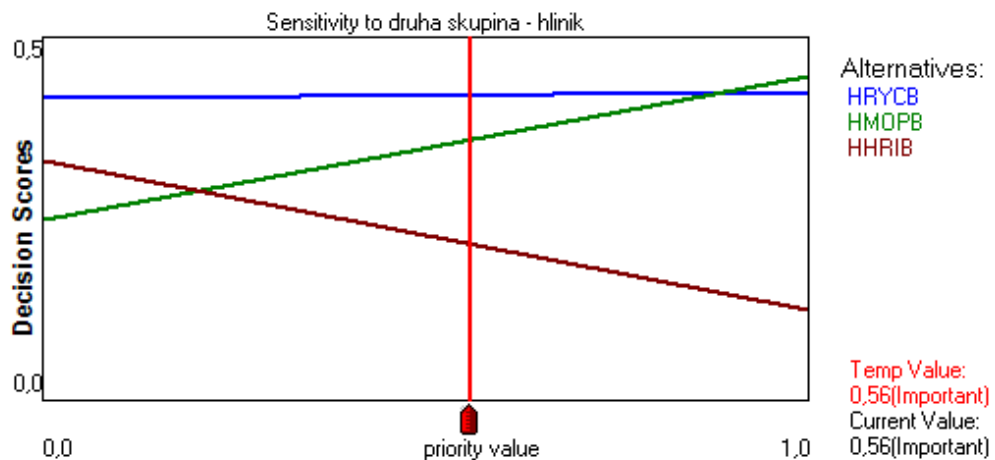
Obrázek 44 Výsledné rozhodnutí ovlivněné kritériem druhé skupiny

Graf na obrázku (Obrázek 44) popisuje, jak by se změnil výsledek jednotlivých variant, kdyby se změnila váha kritéria druhé skupiny. Váha tohoto kritéria je střední, kdyby se váha tohoto kritéria zvedla nad 56 %, stala by se vítěznou lokalitou alternativa Modrý potok.



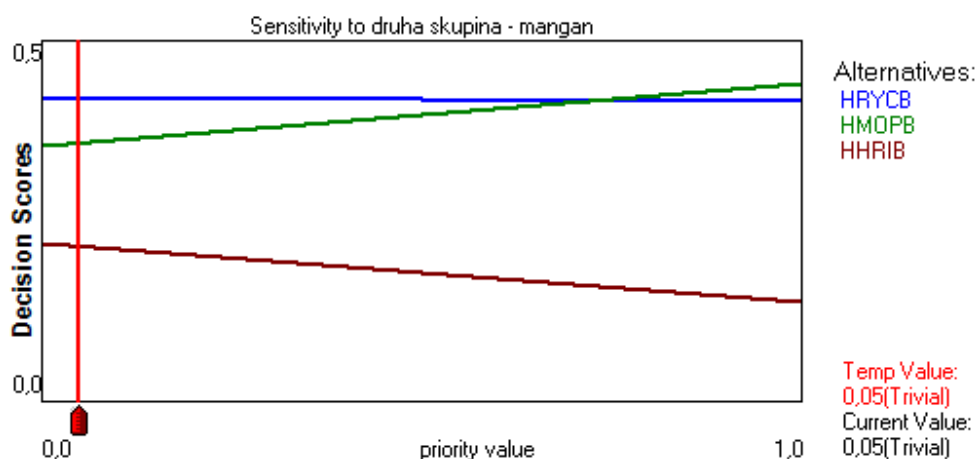
Obrázek 45 Výsledné rozhodnutí ovlivněné kritériem první skupiny

Graf na obrázku (Obrázek 45) názorně ukazuje, jak by se změnil výsledek jednotlivých variant, kdyby se změnila váha kritéria první skupiny. Váha tohoto kritéria je střední, kdyby se váha tohoto kritéria snížila pod 44 %, stala by se výslednou vítěznou lokalitou alternativa Modrý potok. Podle tohoto kritéria by se lokalita Hřibčící nikdy nestala vítěznou alternativou.



Obrázek 46 Výsledné rozhodnutí ovlivněné kritériem hlinik

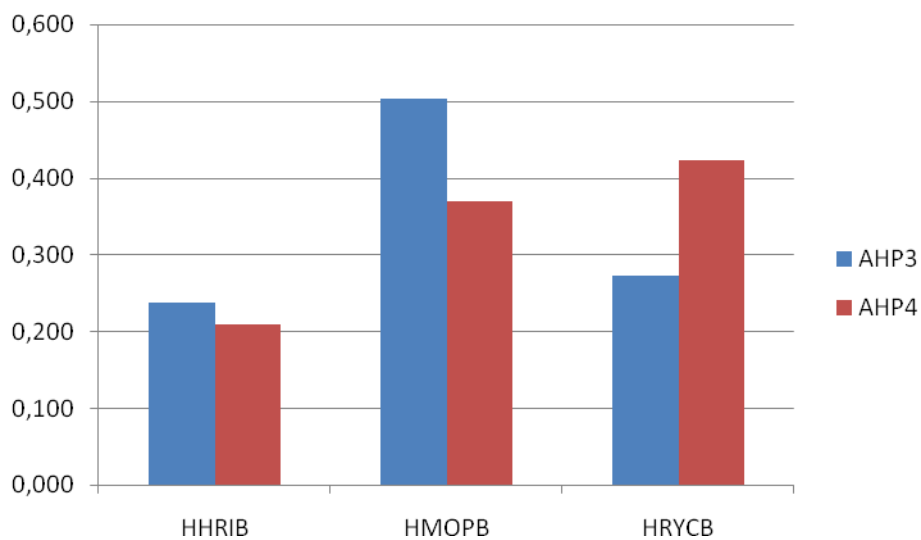
Graf na obrázku (Obrázek 46) popisuje, jak by se změnil výsledek jednotlivých variant, kdyby se změnila váha kritéria obsahu hliníku. Váha tohoto kritéria je střední, kdyby se váha tohoto kritéria zvedla nad 90 %, stala by se vítěznou lokalitou alternativa Modrý potok.



Obrázek 47 Výsledné rozhodnutí ovlivněné kritériem mangan

Graf na obrázku (Obrázek 47) popisuje, jak by se změnil výsledek jednotlivých variant, kdyby se změnila váha kritéria obsahu manganu. Váha tohoto kritéria je nízká, kdyby se váha tohoto kritéria zvedla nad 80 %, stala by se výslednou vítěznou lokalitou alternativa Modrý potok. Podle tohoto kritéria by se lokalita Hřibecí nikdy nestala vítěznou alternativou.

4.8. Závěr kapitoly



Obrázek 48 Rozhodnutí dle dvou metod

Použila jsem dva modely AHP. V jednom případě jsem použila tři a v druhém čtyři vrstvy. Jak názorně ukazuje následující obrázek, výsledky pro rozhodování podávají zcela různé. Na základě prvního modelu je optimální alternativa Modrý potok. Druhý model rozhodování by se přiklonil k optimální alternativě Rýchory. Výsledek rozhodnutí závisí na zvolené metodě, jak znázorňuje obrázek (Obrázek 48).

5. Závěr

Rozhodovací modely používáme k podpoře manažerských rozhodnutí. Systémy na podporu rozhodování slouží pouze jako doporučení pro jednoho nebo více rozhodovatelů. Konečné rozhodnutí provádí člověk. Oproti tomu klasifikační modely patří do strojového učení, jsou to naopak systémy, ve kterých se snažíme lidský faktor eliminovat. Pro oba typy modelů jsou potřeba různá vstupní data. Pro model rozhodovacího problému není nutné takové velké množství vstupních dat jako pro klasifikační modely, a je jedno, zdali jsou vstupní data kategorizovaná nebo číselná (spojitá). U rozhodovacích modelů jsou důležité váhy jednotlivých kritérií. Pro modely klasifikační jsou vstupní data ve větším rozsahu objektů i atributů, také je určeno. Vícekritériální rozhodovací metoda AHP a výhody rozhodovacího modelu:

- transparentnost rozhodování,
- malá vstupní matice,
- minimální požadavky na výpočty a vstupní informace,
- srozumitelnost výsledků.

Nevýhody rozhodovacího modelu:

- lidský faktor, rozhodovatel,
- vhodné určení kritérií,
- subjektivní pohled na daný rozhodovací problém,
- rozhodnutí se váže k datům a k určitému termínu.

Rozhodovací stromy a jejich výhody:

- nahrazení lidského faktoru,
- vytváření jednoduchých pravidel,
- srozumitelnost výsledků,
- použitelné na velký objem dat,
- minimální výpočtové požadavky,
- zpracovává spojitá i kategorizovaná vstupní data.

Nevýhody rozhodovacího stromu:

- nedají se použít při malém množství dat.

V úvodní kapitole jsem určila cíl své diplomové práce a navrhla postup řešení. V druhé kapitole jsem nastínila problematiku životního prostředí a jeho nenahraditelnost pro život člověka. Podkapitola 2.1 popisuje zvolená data. V podkapitole 2.2 popisují přípravu dat pro klasifikační model, provedení korelace, Dixonova testu pro odlehle hodnoty. V podkapitole 2.3 na obrázku (Obrázek 16) jsou vstupy pro rozhodovací stromy naměřené hodnoty prvků obsažených ve srážkách za období 2000 až 2007 ve třech lokalitách. Pro modelování jsou vstupní data upravena.

Algoritmus klasifikačního problému představuje obrázek (Obrázek 17). Model K-means jsem realizovala v kapitole 3.3 a výsledek jsem popsala v kapitole 3.4. Výstupem klasifikačního modelu je klasifikace shluků, vytvořených na základě obsahu kadmia a olova, podle lokality měření, naměřených srážek a pH srážek. Výsledek klasifikace jsem uvedla v kapitole 3.6 a 3.7. Model rozhodovacího stromu jsem popsala v kapitole 3.5 a znázornila v příloze (Příloha 4).

Pro analytický hierarchický proces jsou vstupy průměrné hodnoty obsahu kovů ve srážkách ve vybraných lokalitách za rok 2007. Výstupem rozhodovacího modelu je lokalita s nejnižším obsahem srážek. Výsledek třívrstvého modelu AHP jsem uvedla v kapitole 4.4, výsledek čtyřvrstvého modelu AHP jsem popsala v kapitole 4.7. Rozdílné výsledky a jejich vzájemné porovnání jsem realizovala v kapitole 4.8.

Jednotlivé části své diplomové práce jsem rozdělila do pěti fází na základě CRISP-DM viz [21], u kterých jsem znázornila časovou náročnost:

- části porozumění problému jsem věnovala 5 % času,
- porozumění datům a přípravu dat byly časově nejnáročnější, věnovala jsem tomu 60 % celkového času,
- modelování jsem věnovala 25 % času,
- na analýzu výsledků jsem vynaložila 10 % celkového času.

Ve své práci jsem se zabývala analýzou klasifikačních a rozhodovacích modelů. Navrhla a optimalizovala jsem dva rozhodovací modely pro podporu rozhodování a rozhodovací strom. Rozhodovací problémy jsem řešila pomocí AHP. Pro klasifikační problém jsem zvolila rozhodovací strom CRT. Cíl diplomové práce pro rozhodovací problém, který jsem si stanovila jako výběr lokality s nejnižším obsahem kovů, jsem splnila. Pro klasifikační problém jsem cíl upravila jako určení měření s nejnižším množstvím kadmia a olova dle lokality, množství srážek a pH.

Seznam literatury

- [1] BERKA, P., *Dobývání znalostí z databází* 1. vyd., Praha: Academia, 2003 ISBN: 80-200-1062-9
- [2] BERKA, P., *Tvorba znalostních systémů*, 1. vyd., Praha: Vysoká škola ekonomická, 1994 ISBN: 80-7079-676-6
- [3] BERRY, M. J. A., LINOFF, G. S., *Data mining techniques : for marketing, sales, and customer relationship management*, Indianapolis : Wiley, c2004, ISBN: 0-471-47064-3
- [4] CARLSSON, Ch., FULLÉR, R., *Fuzzy reasoning in decision making and optimization*, Heidelberg: Physica-Verlag, c2002, ISBN: 3-7908-1428-8
- [5] DUDORKIN, J., *Systémové inženýrství a rozhodování* 4. vyd. Praha: České vysoké učení technické v Praze, 2003, ISBN 80-01-02737-6
- [6] FOTR, J., DĚDINA, J., *Manažerské rozhodování*, 1. vyd., Praha: Ekopress, 1997, ISBN: 80-901991-7-8
- [7] FREITAS, A. A., *Data mining and knowledge discovery with evolutionary algorithms : with 74 figures and 10 tables*, Berlin Springer, 2002, ISBN: 3-540-43331-7
- [8] GREENWOOD N.N., EARNSHAW, A., *Chemie prvků*, svazek 1., 1. vyd., Praha: Informatorium, 1993, ISBN 80-85427-38-9
- [9] HILL, T. LEWICKI, P., *Methods and Applications : a Comprehensive Reference for Science, Industry, and Data Mining*, [online] [cit. 2009-3-25] dostupné z www: http://books.google.cz/books?id=TI2TGjeilMAC&pg=PA80&lpg=PA80&dq=chaid+cart+quest&source=bl&ots=PwgIzrHM6z&sig=CBOE0wUls8FIQVGIIXLyYKDZLMc&hl=cs&ei=hWDKSbvoL4uRsAbti_GVDg&sa=X&oi=book_result&resnum=1&ct=result#PPA95,M1
- [10] JIRKOVSKÝ, R., TRŽIL, J., MAŽÁRIOVÁ, G., *Abeceda chemických prvků*, 2. vyd., Bratislava: Alfa, 1985
- [11] KRÁL, V., *Malý encyklopedický slovník A – Ž*, 1. vyd., Praha : Academia, 1972
- [12] KUBANOVÁ, J., *Statistické metody pro ekonomickou a technickou praxi*, 2. vyd., Bratislava: Statis, 2004 ISBN 80-85659-37-9

- [13] KŘUPKA, J. *Teorie systémů I: pro kombinovanou formu studia*. 1. vyd., Pardubice: Univerzita Pardubice, 2007. ISBN 80-7194-923-X
- [14] MAŇAS, M., *Teorie her*, 1. vyd., Praha: Státní nakladatelství technické literatury, 1989
- [15] MAŇAS, M., FIALA, P., JABLONSKÝ, J. *Vícekritériální rozhodování*, 1. vyd., Praha: Vysoká škola ekonomická, 1994, ISBN: 80-7079-748-7
- [16] MAŇAS, M., HUŠEK, R., *Matematické modely v ekonomii*, 1. vyd., Praha: Státní nakladatelství technické literatury, 1989
- [17] MAŘÍK, V., ŠTĚPÁNKOVÁ, O., LAŽANSKÝ, J., *Umělá inteligence 1*, 1. vyd., Praha: Academia, 1997. ISBN 80-200-0496-3
- [18] MAŘÍK, V., ŠTĚPÁNKOVÁ, O., LAŽANSKÝ, J., *Umělá inteligence 2*, 1. vyd. Praha: Academia, 1993. ISBN 80-200-05004-8
- [19] MELOUN, M., MILITKÝ, J., *Statistická analýza experimentálních dat*, 2. vyd., Praha: Academia, 2004. ISBN 80-200-1254-0
- [20] LUKASOVÁ, A., ŠARMANOVÁ, J., *Metody shlukové analýzy*, 1. vyd. Praha: SNTL, 1985
- [21] PETR, P., *Data Mining : Díl I*. Pardubice: Univerzita Pardubice, 2008. ISBN 80-7194-886-1.
- [22] POPL M., Fährnich J., *Analytická chemie životního prostředí*, 2.vyd., Prha: VŠCHT 1995, ISBN 80-7080.165-4
- [23] RAMÍK, J., *Vícekritériální rozhodování – Analytický hierarchický proces (AHP)*. 1. vyd., Opava : Slezská univerzita, 1999 ISBN 80-7248-047-2
- [24] SAATY, T., *The Analytic Hierarchy Process*, 1.title., University of Pennsylvania: 1980, ISBN: 0-054371-2.
- [25] ZAPLATÍLEK, K., DOŇAR, B., *MATLAB : tvorba uživatelských aplikací*, 1. vyd., Praha : BEN - technická literatura, 2004, ISBN: 80-7300-133-0

Seznam elektronických zdrojů

- [26] Atmosféra, [online] [cit. 2009-1-14] dostupné z www:
<http://knc.czu.cz/~vachm/che_pro/ch_atmo.html#vstupy>
- [27] Consistency Index and Consistency Ratio [online] [cit. 2009-3-24] dostupné z www:
<<http://people.revoledu.com/kardi/tutorial/AHP/Consistency.htm>>
- [28] Koloběh vody na zemi, [online] [cit. 2009-1-14] dostupné z www:
<http://geologie.vsb.cz/geologie/KAPITOLY/10_zaklady_hydro/10_zaklady_hydrgeologie_soubory/image003.jpg>
- [29] Komentář k souhrnnému ročnímu tabelárnímu přehledu, [online] [cit. 2009-1-15] dostupné z www: <http://www.chmu.cz/uoco/isko/tab_roc/2007_enh/cze/pdf/kom.pdf>
- [30] Ministerstvo životního prostředí České republiky, [online] [cit. 2009-2-13] dostupné z www:
<http://www.env.cz/cz/ovzdusi_11.2.2008>
- [31] Statistická ročenka životního prostředí České republiky 2007, [online] [cit. 2009-1-14] dostupné z www: <[http://www.cenia.cz/web/www/web-pub2.nsf/\\$pid/CENMSFMVTMNS](http://www.cenia.cz/web/www/web-pub2.nsf/$pid/CENMSFMVTMNS)>
- [32] Znečištění ovzduší na území České republiky v roce 2007 Český hydrometeorologický ústav Úsek ochrany čistoty ovzduší, [online] [cit. 2009-1-14] dostupné z www:
<http://www.chmu.cz/uoco/isko/tab_roc/2007_enh/cze/index.html>

Seznam zkratek

Zkratka	Název
AHP	Analytický hierarchický proces
AMS	Automatizovaná monitorovací stanice
C5.0	Rozhodovací strom C5.0
CART	Classification And Regression Tree
CI	Index konzistence
CR	Poměr konzistence
ČGS	Česká geologická služba
ČHMÚ	Český hydrometeorologický ústav
HHRI	Hřiběcí
HMOP	Modrý potok
HRYC	Rýchory
CHAID	Chisquared Automatic Interaction Detector
ID3	Iterative Dichotomizer
NPSE	Národní program snižování emisí ČR
QUEST	Quick Unbiased and Efficient Statistical Tree
RI	Poměrný index
TDIDT	Top-Down Induction of Decision Trees
VÚV	Výzkumný ústav vodohospodářský T.G.M.
ŽP	Životní prostředí

Seznam obrázků

Obrázek 1 Koloběh vody [28]	9
Obrázek 2 Látkové toky v atmosféře	10
Obrázek 3 Staniční síť v okrese Trutnov [32]	11
Obrázek 4 Naměřené srážky a obsažené kovy za jednotlivá měření v roce 2006 v lokalitě HMOPB...	12
Obrázek 5 Naměřené srážky a obsažené kovy za jednotlivá měření v roce 2006 v lokalitě HHRIB	13
Obrázek 6 Naměřené srážky a obsažené kovy za jednotlivá měření v roce 2006 v lokalitě HRYCB...	14
Obrázek 7 Naměřené srážky v mm za rok 2007 ve vybraných lokalitách	14
Obrázek 8 Naměřené srážky v mm za rok 2006 ve vybraných lokalitách	15
Obrázek 9 Naměřené srážky v mm za rok 2003 ve vybraných lokalitách	16
Obrázek 10 Naměřené srážky v mm za rok 2001 ve vybraných lokalitách	16
Obrázek 11 Odlehlé hodnoty kadmia.....	19
Obrázek 12 Upravené hodnoty kadmia.....	20
Obrázek 13 Porovnání změny rozdělení pravděpodobností výskytu jednotlivých vlastností	21
Obrázek 14 Návrh rozhodovacího modelu.....	23
Obrázek 15 Návrh klasifikačního modelu.....	24
Obrázek 16 Model systému	25
Obrázek 17 Algoritmus klasifikace	28
Obrázek 18 Vstupy do uzlu K-means	31
Obrázek 19 Porovnání shlukování.....	31
Obrázek 20 Grafické znázornění shlukování pomocí dvou kovů kadmia a olova.....	32
Obrázek 21 Rozložení testovacích a trénovacích dat mezi shluky – upravená data	33
Obrázek 22 Vstupy do uzlu pro vytvoření rozhodovacího stromu	35
Obrázek 23 Grafické znázornění shlukování a lokalit	36
Obrázek 24 Grafické znázornění shlukování a srážek	37
Obrázek 25 Grafické znázornění shlukování a pH	37
Obrázek 26 Přehled správnosti klasifikace rozhodovacího stromu na upravených datech	38
Obrázek 27 Přehled správnosti klasifikace rozhodovacího stromu na upravených datech	39

Obrázek 28 Porovnání klasifikování lokalit do shluků podle obsahu kadmia a olova.....	40
Obrázek 29 Data a znalosti v procesu rozhodování [18].....	41
Obrázek 30 Algoritmus AHP	47
Obrázek 31 Systém hierarchického třívrstvého modelu.....	48
Obrázek 32 Saatyho metoda v prostředí MATLAB	51
Obrázek 33 Váhy kritérií jednotlivých alternativ pomocí Saatyho metody	52
Obrázek 34 Normované váhy kritérií pro vybrané metody	53
Obrázek 35 Výsledné rozhodnutí ovlivněné kritériem olovo	53
Obrázek 36 Výsledné rozhodnutí ovlivněné kritériem zinek	54
Obrázek 37 Výsledné rozhodnutí ovlivněné kritériem železo.....	54
Obrázek 38 Systém hierarchický čtyřvrstvého modelu.....	55
Obrázek 39 Normované váhy kritérií.....	57
Obrázek 40 Výsledek Saatyho metody v prostředí MATLAB	57
Obrázek 41 Výsledné rozhodnutí ovlivněné kritériem olovo.....	58
Obrázek 42 Výsledné rozhodnutí ovlivněné kritériem kadmium	58
Obrázek 43 Výsledné rozhodnutí ovlivněné kritériem nikl.....	59
Obrázek 44 Výsledné rozhodnutí ovlivněné kritériem druhé skupiny	59
Obrázek 45 Výsledné rozhodnutí ovlivněné kritériem první skupiny	60
Obrázek 46 Výsledné rozhodnutí ovlivněné kritériem hliník.....	60
Obrázek 47 Výsledné rozhodnutí ovlivněné kritériem mangan	61
Obrázek 48 Rozhodnutí dle dvou metod.....	61

Seznam tabulek

Tabulka 1 Měřené látky	18
Tabulka 2 Průměrné hodnoty naměřených kovů ve srážkách v letech 2000 - 2007	19
Tabulka 3 Směrodatná odchylka neměřených kovů ve srážkách v letech 2000 - 2007	19
Tabulka 4 Průměry a směrodatné odchylky pro neupravená a upravená data.....	22
Tabulka 5 Průměrný index [27]	45
Tabulka 6 Průměrné hodnoty roku 2007 pro vybrané lokality	46
Tabulka 7 Rozhodovací kritéria a jejich priority	47
Tabulka 8 Fullerův trojúhelník	49
Tabulka 9 Saatyho matice	50
Tabulka 10 Váhy kritérií w_m pro jednotlivé metody	50
Tabulka 11 Výsledky jednotlivých metod.....	52
Tabulka 12 Saatyho matice prvního řáda	56

Seznam vzorců

- (1) Dixonův test odlehlých hodnot
- (2) Entropie uzlu
- (3) Entropie celková
- (4) Fullerova metoda pro určení kritéria bez vyrovnání nulových vah
- (5) Fullerova metoda pro určení kritéria s vyrovnáním nulových vah
- (6) Pravidla Saatyho matice
- (7) Saatyho metoda
- (8) Index konzistence
- (9) Poměr konzistence
- (10) Aritmetický průměr
- (11) Geometrický průměr

Seznam příloh

Příloha 1: Zdrojový kód třívrstvého modelu AHP	74
Příloha 2: Zdrojový kód čtyřvrstvého modelu AHP	76
Příloha 3 Saatyho matice druhého řádu pro oba modely AHP	78
Příloha 4 Rozhodovací strom CRT pro vlastnosti (lokalita, srážky, pH)	80
Příloha 5 Další rozhodovací stromy	81
Příloha 6 Ukázka zdrojových dat	81
Příloha 7 Korelace	84
Příloha 8 Statistika	85

Příloha 1: Zdrojový kód třívrstvého modelu AHP

```
clc % vymazat okno
K=7; % pocet kriterii
A=3; % pocet alternativ
velka=[ 1.00 0.11 0.33 0.20 0.11 0.22 0.14
        9.00 1.00 8.00 5.00 3.00 5.00 3.00
        3.00 0.12 1.00 0.33 0.20 0.11 0.14
        5.00 0.13 0.33 1.00 0.20 0.33 0.20
        9.00 0.33 8.00 5.00 1.00 5.00 3.00
        5.00 0.20 5.00 3.00 0.20 1.00 0.20
        7.00 0.33 7.00 5.00 0.33 5.00 1.00];
[ro,lambda]=eig(velka); %funkce vlastního cisla a vektoru;
max_lambda0=norm(lambda) % vypise maximalni vlastni cislo;
max_ro0=ro(:,1) % vypise vlastni vektor u max vlastního cisla;
KI0=(max_lambda0-K)/(K-1) % konsistentni index;
s01=(max_ro0(1:7))/sum(max_ro0)

mala1=[ 1.00 0.14 0.20
        7.00 1.00 3.00
        5.00 0.33 1.00];
[ro,lambda]=eig(mala1);
max_lambda1=norm(lambda)
max_ro1=ro(:,1)
KI1=(max_lambda1-A)/(A-1)

mala2=[ 1.00 0.14 0.33
        7.00 1.00 3.00
        3.00 0.33 1.00];

[ro,lambda]=eig(mala2);
max_lambda2=norm(lambda)
max_ro2=ro(:,1)
KI2=(max_lambda2-A)/(A-1)

mala3=[ 1.00 5.00 3.00
        0.20 1.00 0.33
        0.33 3.00 1.00];

[ro,lambda]=eig(mala3);
max_lambda3=norm(lambda)
max_ro3=ro(:,1)
KI3=(max_lambda3-A)/(A-1)

mala4=[ 1.00 0.20 0.33
        5.00 1.00 3.00
        3.00 0.33 1.00];

[ro,lambda]=eig(mala4);
max_lambda4=norm(lambda)
max_ro4=ro(:,1)
```

```

KI4=(max_lambda4-A)/(A-1)

mala5=[ 1.00  0.20  0.33
        5.00  1.00  5.00
        3.00  0.20  1.00];

[ro,lambda]=eig(mala5);
max_lambda5=norm(lambda)
max_ro5=ro(:,1)

KI5=(max_lambda5-A)/(A-1)

mala6=[ 1.00  3.00  0.20
        0.33  1.00  0.20
        5.00  5.00  1.00];

[ro,lambda]=eig(mala6);
max_lambda6=norm(lambda)
max_ro6=ro(:,1)
KI6=(max_lambda6-A)/(A-1)

mala7=[1.00 7.00  3.00
        0.14 1.00  0.33
        0.33 3.00  1.00];

[ro,lambda]=eig(mala7);
max_lambda7=norm(lambda)
max_ro7=ro(:,1)
KI7=(max_lambda7-A)/(A-1)

wi=s01'
vi=[max_ro1';max_ro2';max_ro3';max_ro4';max_ro5';max_ro6';max_ro7']
alternativy=wi*vi

CI=(max_lambda0-K)/(K-1) %konsintencni index;
CR=CI/1.32 %konsistentni prumer;

figure('Name','vysledek AHP3')
a=alternativy'
bar(a)
xlabel('alternativy')
ylabel('hodnoty')

```

Příloha 2: Zdrojový kód čtyřvrstvého modelu AHP

```
clc % vymazat okno
B=3; % pocet kriterii skupiny 1
C=4; % pocet kriterii skupiny 2
A=3; % pocet alternativ

skup1=[ 1.00 3.00 0.20
        0.33 1.00 0.20
        5.00 5.00 1.00];

[ro,lambda]=eig(skup1); %funkce vlastního čísla a vektoru;
max_lambda01=norm(lambda) % vypise maximalni vlastni cislo;
max_ro01=ro(:,1) % vypise vlastni vektor u max vlastniho čísla;

s01=(max_ro01(1:3))/sum(max_ro01)

KI01=(max_lambda01-B)/(B-1) % konsistentní index;

skup2=[ 1.00 3.00 7.00 5.00
        0.33 1.00 5.00 3.00
        0.14 0.20 1.00 0.20
        0.20 0.33 5.00 1.00];

[ro,lambda]=eig(skup2); %funkce vlastního čísla a vektoru;
max_lambda02=norm(lambda) % vypise maximalni vlastni cislo;
max_ro02=ro(:,1) % vypise vlastni vektor u max vlastniho čísla;

s02=(max_ro02(1:4))/sum(max_ro02)

KI02=(max_lambda02-C)/(C-1) % konsistentní index;

skupina1=s01*0.6
skupina2=s02*0.4

vahy=[skupina1' skupina2']

mala1=[ 1.00 0.14 0.20
        7.00 1.00 3.00
        5.00 0.33 1.00];
[ro,lambda]=eig(mala1);
max_lambda1=norm(lambda)
max_ro1=ro(:,1)

s1=(max_ro1(1:3))/sum(max_ro1)
KI1=(max_lambda1-A)/(A-1)

mala2=[ 1.00 0.14 0.33
        7.00 1.00 3.00
        3.00 0.33 1.00];
```

```

[ro,lambda]=eig(mala2);
max_lambda2=norm(lambda)
max_ro2=ro(:,1)

s2=(max_ro2(1:3))/sum(max_ro2)

KI2=(max_lambda2-A)/(A-1)

mala3=[ 1.00 5.00 3.00
        0.20 1.00 0.33
        0.33 3.00 1.00];

[ro,lambda]=eig(mala3);
max_lambda3=norm(lambda)
max_ro3=ro(:,1)
s3=(max_ro3(1:3))/sum(max_ro3)
KI3=(max_lambda3-A)/(A-1)

mala4=[ 1.00 0.20 0.33
        5.00 1.00 3.00
        3.00 0.33 1.00];

[ro,lambda]=eig(mala4);
max_lambda4=norm(lambda)
max_ro4=ro(:,1)
s4=(max_ro4(1:3))/sum(max_ro4)
KI4=(max_lambda4-A)/(A-1)

mala5=[ 1.00 0.20 0.33
        5.00 1.00 5.00
        3.00 0.20 1.00];

[ro,lambda]=eig(mala5);
max_lambda5=norm(lambda)
max_ro5=ro(:,1)
s5=(max_ro5(1:3))/sum(max_ro5)
KI5=(max_lambda5-A)/(A-1)

mala6=[ 1.00 3.00 0.20
        0.33 1.00 0.20
        5.00 5.00 1.00];

[ro,lambda]=eig(mala6);
max_lambda6=norm(lambda)
max_ro6=ro(:,1)
s6=(max_ro6(1:3))/sum(max_ro6)
KI6=(max_lambda6-A)/(A-1)

```

```

mala7=[1.00 7.00 3.00
        0.14 1.00 0.33
        0.33 3.00 1.00];

[ro,lambda]=eig(mala7);
max_lambda7=norm(lambda)
max_ro7=ro(:,1)
s7=(max_ro7(1:3))/sum(max_ro7)
KI7=(max_lambda7-A)/(A-1)

wm=vahy
vi=[s2';s5';s6';s1';s3';s4';s7']
alternativy=wm*vi

CI01=(max_lambda01-B)/(B-1) %konsintencetni index;
CR=CI01/0.7 %konsistentni prumer;
CI02=(max_lambda02-C)/(C-1) %konsintencetni index;
CR=CI02/0.9 %konsistentni prumer;

figure('Name','vysledek AHP4')
a=alternativy'
bar(a)
xlabel('alternativy')
ylabel('hodnoty')

```

Příloha 3 Saatyho matice druhého řádu pro oba modely AHP

K1	V1	V2	V3
V1	1,00	0,14	0,20
V2	7,00	1,00	3,00
V3	5,00	0,33	1,00

K2	V1	V2	V3
V1	1,00	0,14	0,33
V2	7,00	1,00	3,00
V3	3,00	0,33	1,00

K3	V1	V2	V3
V1	1,00	5,00	3,00
V2	0,20	1,00	0,33
V3	0,33	3,00	1,00

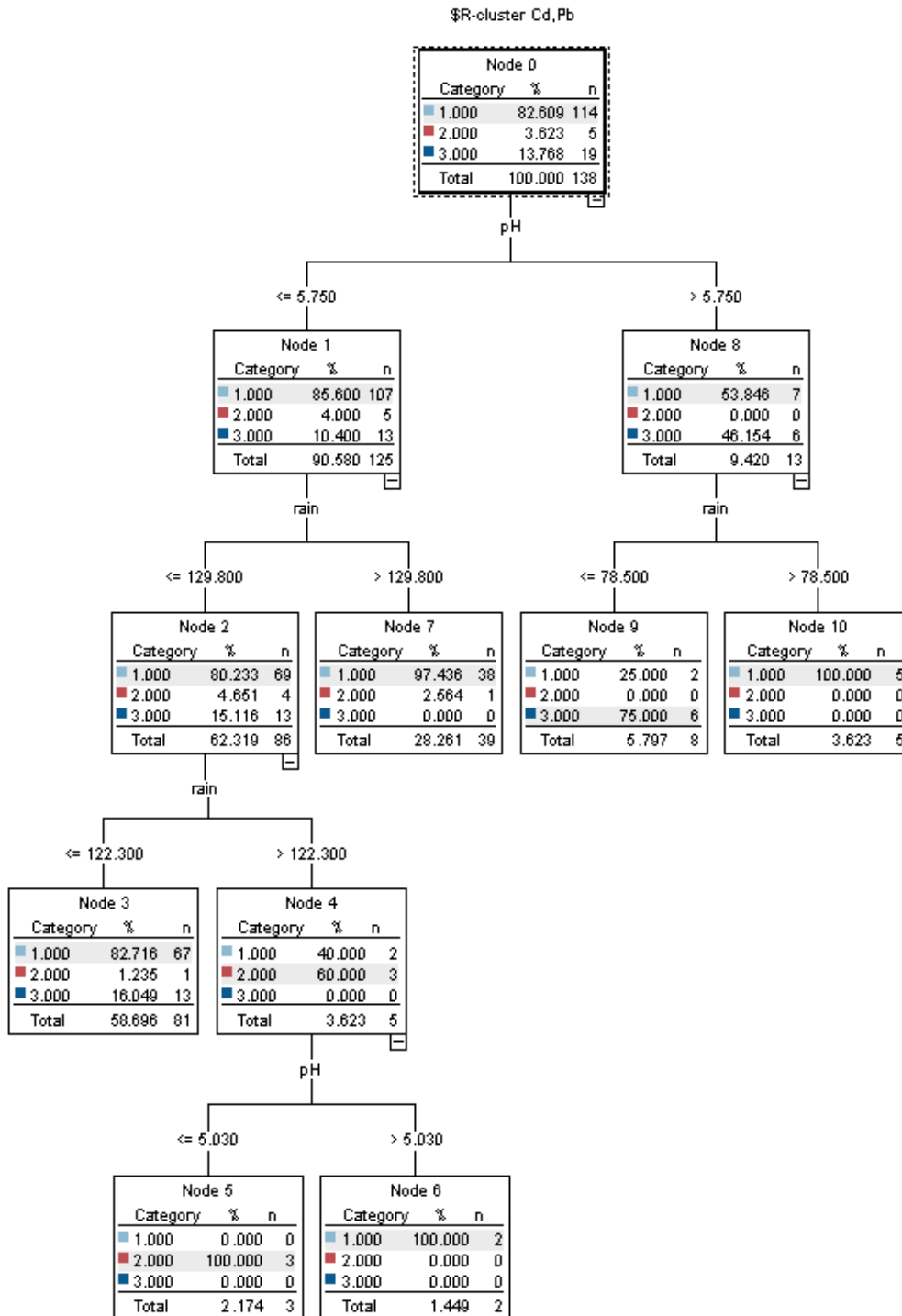
K4	V1	V2	V3
V1	1,00	0,20	0,33
V2	5,00	1,00	3,00
V3	3,00	0,33	1,00

K5	V1	V2	V3
V1	1,00	0,20	0,33
V2	5,00	1,00	5,00
V3	3,00	0,20	1,00

K6	V1	V2	V3
V1	1,00	3,00	0,20
V2	0,33	1,00	0,20
V3	5,00	5,00	1,00

K7	V1	V2	V3
V1	1,00	7,00	3,00
V2	0,14	1,00	0,33
V3	0,33	3,00	1,00

Příloha 4 Rozhodovací strom CRT pro vlastnosti (lokality, srážky, pH)



Příloha 5 Další rozhodovací stromy

Results for output field \$KM-K-Means kovy

Comparing \$R-\$KM-K-Means kovy with \$KM-K-Means kovy

'Partition'	1_Training		2_Testing	
Correct	95	68,84%	86	58,9%
Wrong	43	31,16%	60	41,1%
Total	138		146	

Coincidence Matrix for \$R-\$KM-K-Means kovy (rows show actuals)

'Partition' = 1_Training	cluster-1	cluster-3
cluster-1	63	2
cluster-2	12	2
cluster-3	27	32
'Partition' = 2_Testing	cluster-1	cluster-3
cluster-1	62	10
cluster-2	12	3
cluster-3	35	24

Rozhodovací strom - lokality, srážky, pH, shlukování – pěti kovů, upravená data.

Results for output field \$KM-K-Means kovy

Comparing \$R-\$KM-K-Means kovy with \$KM-K-Means kovy

'Partition'	1_Training		2_Testing	
Correct	120	86,96%	90	61,64%
Wrong	18	13,04%	56	38,36%
Total	138		146	

Coincidence Matrix for \$R-\$KM-K-Means kovy (rows show actuals)

'Partition' = 1_Training	cluster-1	cluster-2	cluster-3
cluster-1	62	0	3
cluster-2	8	6	0
cluster-3	6	1	52
'Partition' = 2_Testing	cluster-1	cluster-2	cluster-3
cluster-1	49	11	12
cluster-2	11	4	0
cluster-3	22	0	37

Rozhodovací strom – vlastnosti mimo kovů, shlukování – pomocí pěti kovů, upravená data.

Příloha 6 Ukázka zdrojových dat

metření	rain [mm]	cond [μS/cm]	pH	Na [ug/l]	K [ug/l]	Zn [ug/l]	Al [ug/l]	Fe [ug/l]	...	Cd [ug/l]	Ni [ug/l]	F [ug/l]	NO3 [ug/l]	SO4 [ug/l]	rok	měsíc	lokalita	organizace
1	280,0	11,6	4,8	176,0	53,0	15,9	24,0	25,0		0,0	0,3	10,0	1060,0	1160,0	2007	1	2	2
2	108,0	14,3	4,5	83,0	70,0	11,5	5,0	25,0		0,0	1,0	10,0	2010,0	1060,0	2007	2	2	2
3	96,0	10,6	4,8	45,0	61,0	11,1	25,0	25,0		0,1	0,3	26,0	1380,0	860,0	2007	3	2	2
4	100,0	19,6	5,7	142,0	513,0	5,0	5,0	25,0		0,1	0,9	10,0	2245,0	2550,0	2007	5	2	2
5	130,0	15,0	5,2	130,0	180,0	58,0	5,0	25,0		0,0	0,3	10,0	1640,0	1977,0	2007	6	2	2
6	158,0	19,4	4,2	90,0	30,0	45,0	15,0	25,0		0,2	2,1	10,0	1489,0	1735,0	2007	7	2	2
7	134,0	11,6	4,6	140,0	170,0	32,0	5,0	25,0		0,0	0,6	10,0	409,0	964,0	2007	8	2	2
8	320,0	44,1	4,1	540,0	2580,0	35,0	44,0	25,0		0,1	0,3	35,0	4832,0	3464,0	2007	9	2	2
9	61,0	21,9	4,4	260,0	130,0	230,0	27,2	80,0		0,3	1,1	10,0	2460,0	2080,0	2007	10	2	2
10	233,8	1,9	4,6	201,0	64,0	2,5	20,0	10,0		0,1	1,0	12,0	2040,0	1580,6	2007	1	1	1
11	113,4	2,0	4,7	100,0	60,0	2,5	20,0	10,0		0,1	1,0	10,0	2680,0	1580,6	2007	2	1	1
12	108,0	3,0	4,4	144,0	110,0	5,2	20,0	10,0		0,1	1,0	12,0	4770,0	1580,6	2007	3	1	1
13	2,5	23,1	4,9	629,0	1110,0	32,5	213,0	125,0		0,7	5,9	66,0	36300,0	1580,6	2007	4	1	1
14	83,9	7,2	4,1	122,0	320,0	7,4	20,0	10,0		0,1	1,0	15,0	16100,0	1580,6	2007	5	1	1
15	90,4	2,0	4,5	82,0	113,0	2,5	20,0	10,0		0,1	1,0	5,0	2730,0	1580,6	2007	6	1	1
16	200,6	1,7	6,0	127,0	62,0	2,5	20,0	10,0		0,1	1,0	14,0	1482,0	1580,6	2007	7	1	1
17	166,7	12,5	3,7	237,0	359,0	7,1	20,0	10,0		0,1	1,0	10,0	24090,0	1580,6	2007	9	1	1
18	35,2	12,2	4,9	73,0	57,0	5,5	20,0	32,0		0,1	1,0	5,0	16700,0	1580,6	2007	10	1	1
19	132,3	5,4	4,0	211,0	80,0	2,5	20,0	10,0		0,1	1,0	5,0	8561,0	1580,6	2007	11	1	1
20	59,5	1,2	4,8	249,0	52,0	6,0	20,0	10,0		0,1	1,0	5,0	1470,0	1580,6	2007	12	1	1
...																		
93	125,4	129,7	7,5	662,0	215,0	14,8	40,0	33,5		0,2	0,5	65,4	4740,0	14600,0	2005	1	3	1
94	119,5	17,3	4,6	831,0	229,0	34,1	40,0	25,9		0,2	1,5	19,7	1670,0	1750,0	2005	2	3	1
95	60,0	39,6	4,3	445,0	178,0	49,0	68,3	88,9		3,4	2,6	48,8	4690,0	4060,0	2005	3	3	1

96	41,3	20,2	5,2	564,0	331,0	17,8	40,0	20,0		0,2	0,5	170,0	2540,0	3240,0	2005	4	3	1
97	80,0	31,2	4,5	199,0	367,0	8,8	55,7	73,0		0,2	0,5	58,2	1150,0	2520,0	2005	5	3	1
98	49,9	12,3	5,6	93,0	96,0	19,5	60,7	98,0		0,2	0,6	12,0	1000,0	1630,0	2005	6	3	1
99	194,2	8,2	5,4	439,0	105,0	11,3	40,0	33,8		0,2	0,5	14,0	1000,0	1080,0	2005	7	3	1
100	81,7	97,6	4,4	81,0	95,0	8,8	40,0	48,4		0,2	0,5	11,0	1000,0	2500,0	2005	8	3	1
101	109,6	10,7	5,1	150,0	121,0	7,4	65,0	60,0		0,2	0,5	6,0	1000,0	1720,0	2005	9	3	1
102	14,6	46,5	4,2	480,0	430,0	55,7	213,0	170,0		0,7	1,0	39,0	1500,0	7960,0	2005	10	3	1
103	52,9	25,4	4,7	390,0	250,0	40,1	71,2	109,0		0,2	0,5	26,0	3010,0	3140,0	2005	11	3	1
104	94,2	18,0	4,7	490,0	210,0	24,6	50,0	20,0		0,2	0,9	3,0	2290,0	1470,0	2005	12	3	1
105	514,5	16,7	5,7	820,0	180,0	16,0	74,0	20,0		0,1	0,2	10,0	2040,0	1800,0	2004	1	2	2
106	207,0	25,7	4,5	410,0	70,0	34,0	53,0	20,0		0,2	1,3	10,0	3444,0	1737,0	2004	2	2	2
107	96,5	18,0	5,2	210,0	800,0	23,0	51,0	20,0		0,1	5,4	10,0	2050,0	2231,0	2004	3	2	2
108	114,0	30,5	6,7	180,0	50,0	9277,0	27,2	60,0		0,7	2,7	10,0	540,0	1129,0	2004	4	2	2
109	93,0	23,2	5,9	550,0	410,0	2189,0	27,2	20,0		0,7	2,7	10,0	2549,0	2840,0	2004	5	2	2
110	216,3	26,5	6,0	105,0	587,0	24,9	13,0	20,0		0,1	2,7	24,0	2710,0	3370,0	2004	6	2	2
111	139,5	12,4	5,3	84,0	24,0	40,5	11,0	20,0		0,2	2,7	24,0	1660,0	1600,0	2004	7	2	2
112	216,3	12,2	4,8	26,0	10,0	27,3	5,0	3,0		0,1	0,9	10,0	430,0	1220,0	2004	8	2	2
113	197,7	17,2	4,6	252,0	26,0	17,7	5,0	3,0		0,1	0,6	10,0	850,0	2380,0	2004	9	2	2
114	167,4	29,9	4,4	100,0	50,0	56,0	27,2	20,0		0,7	2,7	19,8	280,0	2550,0	2004	10	2	2
115	232,6	119,6	4,9	1350,0	9310,0	68,0	59,0	80,0		0,5	1,9	31,0	724,0	10809,0	2004	11	2	2
...																		
263	240,7	18,5	4,4	0,3	0,1	5,6	8,3	0,1		0,0	0,1	0,0	2,3	1,0	2000	3	1	1
264	20,1	57,4	4,1	0,7	0,7	59,8	14,9	0,3		0,2	0,2	0,0	11,7	6,1	2000	4	1	1
265	83,0	17,4	5,2	0,1	0,2	13,5	6,4	0,1		0,1	0,1	0,0	3,7	2,1	2000	5	1	1

266	45,7	22,9	4,5	0,3	0,1	22,1	10,1	0,3		0,1	0,0	0,0	3,6	1,9	2000	6	1	1
267	155,5	11,9	4,8	0,2	0,1	2,5	4,3	0,2		0,0	0,0	0,0	1,4	1,3	2000	7	1	1
268	64,7	41,2	3,9	0,1	0,3	17,6	13,6	0,1		0,2	0,1	0,0	6,7	3,0	2000	8	1	1
269	108,3	16,8	4,5	0,1	0,1	9,6	4,8	0,1		0,1	0,0	0,0	1,9	2,3	2000	9	1	1
270	55,0	23,2	4,4	0,2	0,2	21,1	11,5	0,2		0,1	0,1	0,0	3,1	3,7	2000	10	1	1
271	101,8	25,5	4,2	0,3	0,1	12,8	5,3	0,1		0,0	0,0	0,0	2,6	2,2	2000	11	1	1
272	117,6	16,6	4,7	0,4	0,1	17,6	7,6	0,1		0,0	0,0	0,0	1,8	1,8	2000	12	1	1
273	117,2	18,6	4,2	0,5	0,2	17,3	7,2	0,1		0,1	0,0	0,0	2,8	1,6	2000	1	3	1
274	131,5	19,9	4,6	0,4	0,2	37,5	18,0	0,2		0,3	0,1	0,0	3,2	1,7	2000	2	3	1
275	199,8	12,9	5,1	0,3	0,1	10,7	8,1	0,1		0,1	0,0	0,0	1,9	1,6	2000	3	3	1
276	13,1	37,8	4,1	0,2	0,3	52,6	21,2	0,3		0,3	0,2	0,0	9,6	5,3	2000	4	3	1
277	89,7	22,5	4,5	0,2	0,3	22,2	20,9	0,1		0,2	0,1	0,0	4,5	2,2	2000	5	3	1
278	58,9	37,5	4,1	0,4	0,2	17,3	8,8	0,1		0,1	0,1	0,0	5,9	2,6	2000	6	3	1
279	132,7	11,7	4,7	0,2	0,1	2,5	8,5	0,1		0,0	0,0	0,0	1,8	1,0	2000	7	3	1
280	57,3	32,7	4,2	0,2	0,4	23,5	8,6	0,1		0,1	0,1	0,0	5,8	4,6	2000	8	3	1
281	50,9	22,9	4,6	0,2	0,1	19,1	37,1	0,1		0,1	0,1	0,0	3,1	4,6	2000	9	3	1
282	54,4	33,2	4,4	0,3	0,1	33,0	20,5	0,3		0,1	0,1	0,0	5,1	5,2	2000	10	3	1
283	87,6	28,3	4,2	0,2	0,0	19,4	13,9	0,1		0,1	0,1	0,0	3,0	3,3	2000	11	3	1
284	82,1	21,3	4,7	0,6	0,2	20,3	7,0	0,1		0,1	0,0	0,0	2,5	2,8	2000	12	3	1
odchylka	85	25	1	276	649	566	33	40		2	6	31	4169	2099	2			
průměr	114	23	5	188	200	67	28	28		1	3	20	2135	1581	2003			

	rain	cond	pH	Na	K	NH4	Mg	Ca	Mn	Zn	Al	Fe	Pb	Cd	Ni	F	Cl	NO3	SO4	lokalita
rain	1																			
cond	-0,05	1																		
pH	0,03	0,16	1																	
Na	-0,03	0,15	0,20	1																
K	0,05	0,26	0,12	0,46	1															
NH4	-0,09	0,39	0,01	0,19	0,41	1														
Mg	-0,11	0,16	0,09	0,59	0,43	0,17	1													
Ca	-0,10	0,07	0,09	0,47	0,22	0,09	0,74	1												
Mn	-0,01	0,14	0,07	0,30	0,60	0,19	0,30	0,21	1											
Zn	-0,01	0,02	0,18	0,03	0,00	0,00	-0,01	0,00	0,00	1										
Al	-0,27	0,20	0,14	0,54	0,25	0,23	0,57	0,37	0,23	0,07	1									
Fe	-0,26	0,13	0,17	0,46	0,22	0,14	0,46	0,25	0,12	0,06	0,79	1								
Pb	-0,29	0,10	-0,01	0,38	0,09	0,04	0,40	0,25	0,05	0,07	0,66	0,69	1							
Cd	-0,02	-0,03	-0,05	-0,13	-0,06	-0,09	-0,09	-0,06	-0,08	-0,12	-0,15	-0,11	0,03	1						
Ni	-0,07	-0,02	-0,07	-0,15	-0,05	-0,07	-0,10	-0,06	-0,09	-0,15	-0,15	-0,13	0,02	0,73	1					
F	-0,20	0,09	0,21	0,32	0,17	0,12	0,33	0,22	0,03	-0,03	0,42	0,36	0,30	-0,07	-0,06	1				
Cl	-0,12	0,28	0,20	0,69	0,46	0,42	0,58	0,42	0,33	0,00	0,65	0,43	0,40	-0,12	-0,11	0,35	1			
NO3	-0,20	0,25	-0,16	0,31	0,16	0,50	0,39	0,33	0,11	-0,02	0,47	0,26	0,25	-0,10	-0,08	0,24	0,68	1		
SO4	-0,23	0,41	0,20	0,58	0,42	0,54	0,52	0,42	0,27	0,01	0,60	0,45	0,43	-0,14	-0,15	0,40	0,77	0,65	1	
lokalita	-0,06	0,01	-0,08	0,12	-0,01	-0,05	0,07	0,11	0,01	0,01	0,08	0,06	0,22	0,01	-0,02	0,06	-0,04	-0,12	0,12	1

	směrodatná odchylka	mezikvartilové rozpětí	minimum	maximum	rozpětí	součet	aritmetický průměr	modus	medián
rain	85,20	564,90	2,10	567,00	564,90	32439,30	114,22	60,00	95,55
cond	25,18	310,96	1,04	312,00	310,96	6503,63	23,14	4,00	17,50
pH	0,69	4,16	3,69	7,85	4,16	1321,50	4,91	4,51	4,72
Na	276,96	2269,98	0,02	2270,00	2269,98	53275,97	188,25	50,00	100,00
K	650,44	9309,98	0,02	9310,00	9309,98	56501,82	199,65	0,11	70,00
NH4	1679,35	13899,99	0,01	13900,00	13899,99	189609,48	674,77	0,01	210,00
Mg	146,07	1120,00	0,00	1120,00	1120,00	21886,98	77,34	0,05	30,00
Ca	1143,72	15399,97	0,03	15400,00	15399,97	119800,60	423,32	0,19	140,00
Mn	33,22	374,98	0,02	375,00	374,98	3269,18	11,63	2,50	5,00
Zn	566,55	9276,97	0,03	9277,00	9276,97	18821,37	66,51	0,03	16,00
Al	33,20	213,00	0,00	213,00	213,00	7567,66	27,22	40,00	16,50
Fe	40,84	270,00	0,00	270,00	270,00	7819,21	27,63	0,10	10,00
Pb	5,38	39,00	0,00	39,00	39,00	935,47	3,35	0,30	1,30
Cd	2,11	21,20	0,00	21,20	21,20	181,76	0,65	0,20	0,10
Ni	6,62	40,10	0,00	40,10	40,10	718,64	2,66	0,50	0,50
F	31,29	389,99	0,01	390,00	389,99	5554,92	19,77	10,00	10,00
Cl	632,43	6879,92	0,08	6880,00	6879,92	116766,09	412,60	0,10	279,00
NO3	4175,91	36299,87	0,13	36300,00	36299,87	604207,33	2135,01	0,30	1000,00
SO4	2145,29	16699,98	0,02	16700,00	16699,98	429914,52	1580,57	0,05	1166,50

Kód	Organizace	četnosti	procenta
3	ČGS	35	12,32
2	ČHMÚ	58	20,42
1	VÚV	191	67,25
		284	100

Kód	Lokalita		četnosti	procenta
1	HHRIB	HHRI2	95	33,45
2	HMOPB	HMOP2	93	32,75
3	HRYCB	HRYC2	96	33,80
			284	100