

Příloha 1/A

Vyhodnocení experimentálních dat – Severní Čechy

1 Úvod

Cílem této zprávy je vyhodnocení dat, která byly pořízena v letech 2001 – 2002 v severních Čechách. Jedná se dvě stanoviště, označené TEPLICE a MOST, na kterých byly provedeny dvě odběrové kampaně: v zimě 2001 a létě 2002. Jednalo se o 24 hodinové odběry v průběhu jednoho měsíce. Datové soubory, získané analýzou odebraných vzorků, obsahovaly údaje o koncentraci polyaromatických uhlovodíků (PAH) a některých prvků. Dále byly provedeny podpisy 6 významných energetických zdrojů z hlediska obsahů PAH a některých prvků.

Pro matematicko-statistické zpracování těchto souborů byly využity multivariační statistické metody a program US EPA CMB 8.2.

2 Teoretická část

2.1 Multivariační průzkumová analýza

Prvotní informaci o struktuře dat může poskytnout **metoda hlavních komponent (PCA)**.

Metoda hlavních komponent

Metoda hlavních komponent neboli PCA je metodou vícerozměrné popisné statistiky. Matematickým způsobem snižuje počet proměnných charakteristik tím, že vytváří teoretické ("umělé") proměnné neboli parametry, které v sobě však zahrnují informaci o všech původních proměnných. Tyto nové proměnné, které vytvářejí nový souřadnicový prostor, nazýváme hlavními komponentami. Umístění bodů v tomto novém prostoru nazýváme *score*, zastoupení původních proměnných v hlavních komponentách nazýváme zátěžemi nebo komponentními váhami (*loadings*). Podaří-li se touto metodou vyjádřit původní parametry souboru pomocí menšího počtu přímo neměřitelných teoretických proměnných, které je možno racionálně interpretovat, získáme úspornější a efektivnější popis souboru studovaných objektů i struktury závislostí mezi jejich parametry. Zároveň si můžeme utvořit představu o druhu a intenzitě působení procesů, které podmiňují pozorovanou variabilitu datových souborů. PCA analýza může tedy sloužit k několika účelům:

- a) k identifikaci teoretických proměnných,
- b) k transformaci dat do systému ortogonálně proměnných,
- c) ke klasifikaci objektů nebo proměnných
- d) k redukci rozměrů zkoumaného problému, tj. redukci rozměru prostoru nahrazením většího počtu původních proměnných čili parametrů novým souborem o menším počtu teoretických proměnných.

Při výpočtu komponentní analýzy jsou důležité výše uvedené komponentní zátěže (*loadings*). Jsou to hodnoty, které se získávají průmětem vektorů z prostoru na ortogonální osy hlavních komponent, čili projekční koeficienty jednotlivých vektorů. Bývají také někdy označovány jako korelační koeficienty původních charakteristik s jednotlivými komponentami. Při interpretaci je nutné vzít v úvahu hlavní komponenty s největšími vlastními čísly a zjišťovat, které charakteristiky s jednotlivými komponentami korelují. Vysoké kladné a vysoké záporné váhy pro prvky téže komponenty jednoduše odrážejí antagonistické chování těchto prvků. Tímto způsobem si můžeme učinit představu o tom, které konkrétní vlivy nebo procesy jsou příčinou pozorované variability charakteristik a závislosti mezi jejich hodnotami.

2.2 Sdružovací (klastrová) analýza

Základní principy sdružovací analýzy

Výchozím bodem většiny sdružovacích metod je stanovení míry podobnosti objektů. Většinou požadujeme, aby byla nulou vyjádřena maximální rozdílnost a jedničkou totožnost. Z praktických důvodů však často pracujeme s mírami nepodobnosti neboli se vzdálenostmi, které nabývají nulových hodnot v případě totožnosti a jejichž hodnoty rostou se zvětšující se rozdílností objektů. Hodnota "vzdálenosti"

představuje spíše míru nepodobnosti a její doplněk je výrazem podobnosti. Nejznámější je euklidovská vzdálenost mezi dvěma body (jedincem A a B), určeným ortogonálními vektory:

$$d_{AB} = \sqrt{\sum_{i=1}^n (X_{iA} - X_{iB})^2}, \text{ kde}$$

X_{iA} = hodnota znaku i příslušná objektu A

X_{iB} = hodnota znaku i příslušná objektu B

Společnou nevýhodou všech vzdáleností je závislost na jednotkách měření (vzdálenost nejvíce ovlivní charakteristiky s největšími hodnotami směrodatných odchylek). Nepříznivě se projevuje i vliv korelovatelnosti charakteristik.

Všechny techniky sdružovací analýzy vycházejí z matice pozorování, která není předem rozdělena ani podle jedinců, ani podle proměnných. Na základě matice pozorování příslušných jedinců se formuluje postup jak sdružit jedince (způsob Q) nebo proměnné (způsob R) do navzájem oddělených skupin na podkladě některého z koeficientů podobnosti.

2.3 Fuzzy sdružovací analýza

Konvenční shlukové procedury se snaží rozdělit zpracovávaný soubor do skupin. Výsledkem je rozdělení, ve kterém je každý objekt jednoznačně přiřazen vždy právě k jednomu shluku. Z toho vyplývá, že vzorky, které ve skutečnosti představují přechody mezi dvěma nebo více shluky, a vzorky, které nepatří k žádnému shluku (outsideri), ovlivňují sumární složení shluků do kterých jsou zařazeny, v mnohem větší míře, než by odpovídalo skutečnosti. Konvenční shlukové procedury tedy poskytují ostrou klasifikaci objektů. V mnoha oborech, např. v oblasti životního prostředí, v biologii a biochemii se s ostrým rozdělením a ostrými hranicemi mezi jevy setkáváme zřídka. Typické jsou právě objekty reprezentující směsi několika koncových členů nebo objekty, které se nacházejí v přechodné pozici mezi určitými extrémy. Aplikace konvenčních shlukových procedur v praxi bývá i z tohoto hlediska oprávněně kritizována.

Fuzzy klasifikační procedury používají odlišný postup. Podobnost mezi objektem a shlukem se dá vyjádřit pomocí funkce, která nabývá hodnot od nuly do jedné a vyjadřuje členství jednotlivých objektů ve všech shlucích. Pro každý objekt souboru, který se rozpadá do c shluků, je tedy určena míra členství zároveň ve všech těchto shlucích. Pokud má členství v určitém shluku rovné nule, znamená to, že do něj vůbec nepatří, nebo k němu nemá žádný podobnostní vztah. Naopak objekt, který má v určitém shluku členství rovno jedné, patří pouze do tohoto shluku. Hodnoty mezi jedničkou a nulou svědčí o různě velkém podobnostním vztahu vzorku k odpovídajícím shlukům. Přitom platí, že součet hodnot členství jednoho vzorku ve všech c shlucích se rovná jedné.

Členství vzorku v daném shluku závisí na vzdálenosti tohoto vzorku od tzv. prototypu shluku. Platí

$$d_{ik}^2 = (x_k - v_i)'A(x_k - v_i), \text{ kde}$$

x_k je vektor hodnot proměnných k-tého shluku, v_i je vektor reprezentující prototyp i-tého shluku, \mathbf{A} je váhová matice určující způsob výpočtu vzdálenosti. k se mění od 1 do N (N je počet vzorků), i se mění od 1 do c (c je počet shluků). \mathbf{A} je čtvercová symetrická matice typu $p \times p$ (p je počet proměnných, čili rozměr úlohy).

Vektor v_i (prototyp i-tého shluku) se počítá jako vážený průměr

$$v_i = \sum_{k=1}^N (u_{ik})^m x_k / \sum_{k=1}^N (u_{ik})^m$$

(pro jednotlivé proměnné, tj. pro $j = 1, \dots, p$) u_{ik} je hodnota členství k-tého vzorku v i-tém shluku. Důležitý je parametr m , který vyjadřuje míru rozmazání (fuzziness). Čím je hodnota m větší, tím je model "rozmazanější", naopak čím více se m blíží jedné, tím je systém "ostřejší". Pro $m = 1$ dostáváme ostré rozdělení. Každý vzorek je členem vždy jen jednoho shluku. V tomto případě jsou prototypy shluků určeny jako aritmetické průměry ze vzorků, které jsou do nich zařazeny (jejich těžiště). Naopak, pro velké hodnoty m dostáváme rozdělení, ve kterém všechny vzorky mají stejné členství ve všech shlucích:

$$u_{ik} = \text{konst.} = \frac{1}{c}$$

Obecná pravidla pro volbu parametru m neexistují, použitelné hodnoty leží mezi 1,25 a 2,00.

Velkým přínosem fuzzy shlukování je schopnost této metody rozeznávat v datovém souboru kontinuální struktury, identifikovat outsidersy a směsné vzorky (včetně poměrů složek, které je skládají). Přítomnost outsiderů v souboru se může projevit dvojnásobným způsobem. Buď vytvoří jednoprvkové shluky, ve kterých mají ostatní vzorky nulová nebo velmi nízká členství, nebo mají přibližně stejná, nízká členství ve všech shlucích.

Fuzzy shlukování má navíc další vlastnost, která je zvýhodňuje před konvenčními metodami zejména v určitých aplikacích: nevyžaduje normalitu zpracovávaných dat. Jediným požadavkem je počet shluků.

Za nejužitečnější vlastnost fuzzy shlukování bývá považována stabilita prototypu shluků. Tato stabilita je dána způsobem jejich výpočtu; jde o průměry vážené funkcí členství, umocněné parametrem m , takže vzorky typické pro daný shluk přispívají více k jeho složení, než vzorky přechodné a outsidersy. Je přitom v pořádku, že se ve složení shluků uplatňují svým dílem také přechodné vzorky. Vliv outsiderů je většinou nepatrný. Prototypy shluků tedy můžeme považovat za reprezentativní odhady polohy daných shluků. Svým způsobem doplňují sérii robustních a neparametrických odhadů.

C-průměrové fuzzy shlukování v řadě aplikací prokázalo značnou účinnost a rozlišovací schopnosti. Můžeme ho řadit mezi robustní metody, protože jeho výsledky nejsou tolik ovlivněny odchylkami od normálního statistického rozdělení jako konvenční statistické metody. Aplikuje se s výhodou v těch situacích, kdy je složení studovaných vzorků určeno větším počtem faktorů, jejichž složení a počet většinou nejsou předem známy. Výsledky fuzzy shlukování se dobře shodují s výsledky sdružovací analýzy.

2.4 Lineární diskriminační analýza

Metody se používá pro řešení otázek třídění jedinců do určitých, předem definovaných skupin. Lze to doložit na příkladu dvou nebo více souborů, které jsou charakterizovány systémem znaků (proměnných, parametrů). O daném jedinci máme rozhodnout, do kterého ze souborů patří. Diskriminační funkce vymezuje kritéria pro oddělení jednotlivých souborů a tím i zařazení jedince.

Diskriminační analýza se liší od sdružovacích metod tím, že určuje příslušnost jedinců do předem vymezených skupin, zatímco sdružovací analýzy hledají skupiny přítomných jedinců a vytvářejí jejich systém, aniž by zpočátku byl určen charakter nebo počet skupin.

Diskriminační analýzou lze řešit v podstatě dvě skupiny klasifikačních problémů:

1. Otázky samostatnosti skupin jedinců vydělených na základě souboru znaků (proměnných). Tento problém lze stručně formulovat takto: Máme-li dva nebo více souborů představovaných jednoznačně roztríděnými jedinci, vyjádříme pomocí souboru znaků (proměnných) „vzdálenost“ mezi těmito soubory a tak získáme představu míry nezávislosti mezi jednotlivými soubory navzájem. Je-li vzdálenost mezi určitou dvojicí souborů menší než zvolené kritérium významnosti rozdílu mezi skutečně samostatnými skupinami, jsme oprávněni revidovat správnost původní zvolené klasifikace a hledat příčiny a chyby.
2. Otázky zařazení sporných jedinců do předem definovaných skupin na základě typických souborů. Každý z problematických souborů je zařazen do určité skupiny podle afinity svých znaků k té či oné skupině. Lze určit pravděpodobnost příslušnosti jedince k určeným skupinám.

Významnou součástí diskriminační analýzy je výpočet diskriminační funkce a testování „vzdálenosti“ mezi soubory.

2.5 Receptorové modelování

Receptorové modelování používá chemické a fyzikální vlastnosti plynů a aerosolových částic, které jsou změřeny pro emisní zdroje a receptory (v českém prostředí používáme slovo imisní charakteristiky místa), pro identifikaci a kvantifikaci příspěvku jednotlivých zdrojů na jednotlivých imisních (odběrových) místech. Z matematického hlediska zpracování emisních a imisních dat zahrnuje:

- a) chemickou hmotnostní bilanci
- b) faktorovou analýzu datové matice
- c) vícenásobnou lineární regresi

Základním požadavkem pro receptorové modelování je:

- a) stabilita emisních zdrojů v monitorovaném období
- b) nalezení všech významných emisních zdrojů, které přispívají k charakterizaci imisních míst

V práci byl použit programový soubor CMB 8.2, který byl poskytnut pracovníky US EPA v North Carolina.

3 Experimentální část

Odběrové vzorkovací plány, analýza odebraných vzorků a vytvoření datových matic pro oblast severní Čechy jsou předmětem dílčích zpráv za rok 2002 a 2003.

4 Vyhodnocení experimentálních dat a diskuse

Byly zpracovány 4 datové soubory imisních dat:

1. PAH, zima, 16. listopadu až 28. prosince 2001
2. Kovy, zima, 16. listopadu až 28. prosince 2001
3. PAH léto, 3. června až 29. června 2002
4. Kovy, léto, 3. června až 29. června 2002

Dále byly zpracovány dva soubory emisních dat, pro PAH a kovy z těchto velkých energetických zdrojů:

DEKONTA
LEDVICE_CEZ
CHZ LITVINOV
POCERADY_CEZ
TRMICE
KOMORANY

V souborech je použito pro označování jednotlivých PAH těchto zkratek:

4.1 Průzkumová analýza dat

Soubor PAH – zima

Soubor obsahoval 58 datových vektorů (TEPLICE, MOST). Každý z vektorů je charakterizován 13 koncentracemi PAH (ng/m³). χ^2 – test potvrdil vysoký stupeň korelace mezi jednotlivými proměnnými ($\chi^2 = 2300$ (78), χ^2 (78) = 99). Základní statistické charakteristiky souboru jsou uvedeny v následující Tabulce I.

Tabulka I Základní statistické charakteristiky souboru PAH – zima

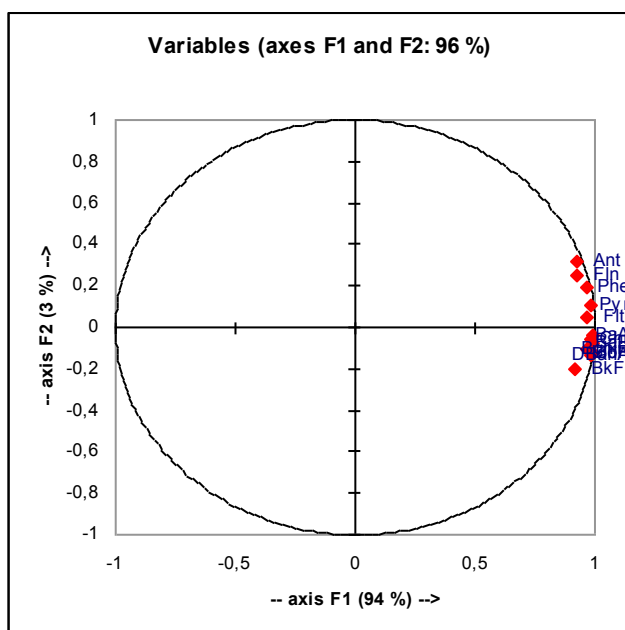
PAH	průměr	SD	RSD
Fln	70.8	50.6	71.4%
Phe	148.5	80.6	54.2%
Ant	15.0	14.5	96.5%
Flt	56.4	40.2	71.2%
Pyr	50.4	37.3	74.0%
BaA	13.3	11.5	86.1%
Chr	13.4	9.7	72.1%
BbF	10.4	8.1	78.1%
BkF	5.0	4.4	87.2%
BaP	8.3	7.5	91.1%
DBahA	1.0	0.9	88.8%
BghiP	7.7	6.4	82.3%
IcdP	8.2	6.3	77.2%

SD – výběrová směrodatná odchylka, RSD – rel. SD

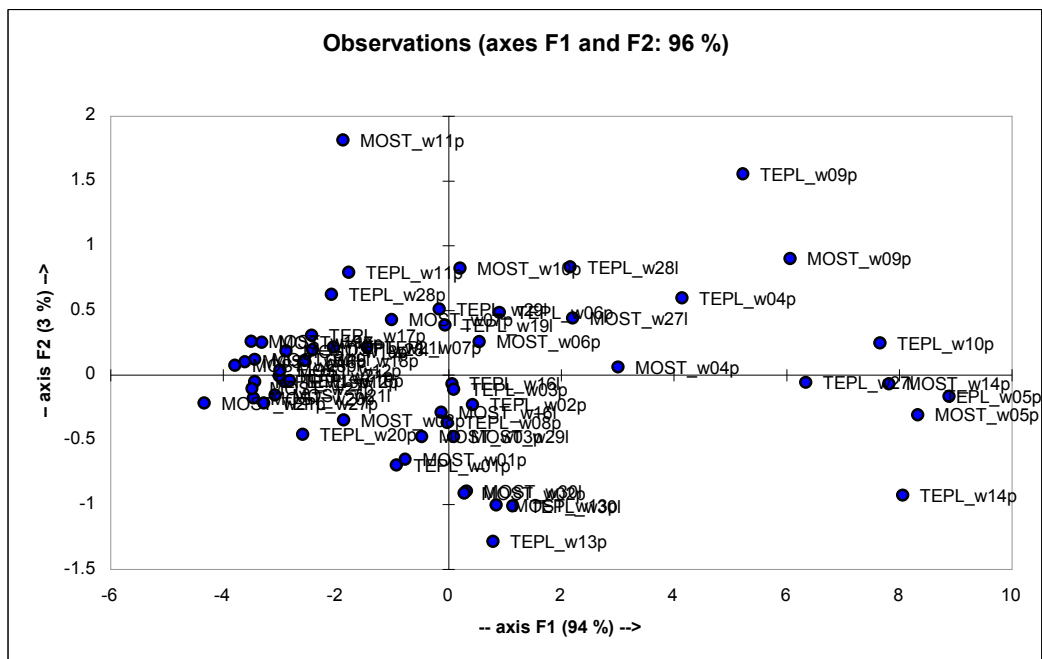
Vysoké hodnoty RSD svědčí o tom, že data nejsou homogenní, tj. nepocházejí z jednoho zdroje.

Na obrázku 1 je graf korelační kružnice, znázorňující korelaci původních proměnných (PAH) s prvními dvěma hlavními komponentami. První dvě hlavní komponenty nesou 96,2% veškeré informace o tomto souboru, jsou tedy postačující pro dokonalý popis původních proměnných. Všechny proměnné vykazují vysokou vzájemnou korelaci (na korelační kružnici jsou umístěny blízko sebe) a zároveň jsou z 94% popsány první hlavní komponentou.

Na obrázku 2 je graf skóre, který ukazuje rozdělení všech „vzorků“ (vektory pozorování) v prostoru prvních dvou hlavních komponent. Z obrázku je patrné, že vzorky netvoří pouze jeden shluk, pocházejí tedy z více emisních zdrojů.



Obr. 1 Korelační kružnice pro soubor PAH-zima



Obr. 2 Graf skóre pro soubor PAH - zima

Soubor PAH – léto

Soubor obsahoval 50 datových vektorů (TEPLICE, MOST). Každý z vektorů je charakterizován 12 koncentracemi PAH (ng/m³). χ^2 – test potvrdil vysoký stupeň korelace mezi jednotlivými proměnnými ($\chi^2 = 854$ (66), χ^2 (66) = 86). Základní statistické charakteristiky souboru jsou uvedeny v následující Tabulce II.

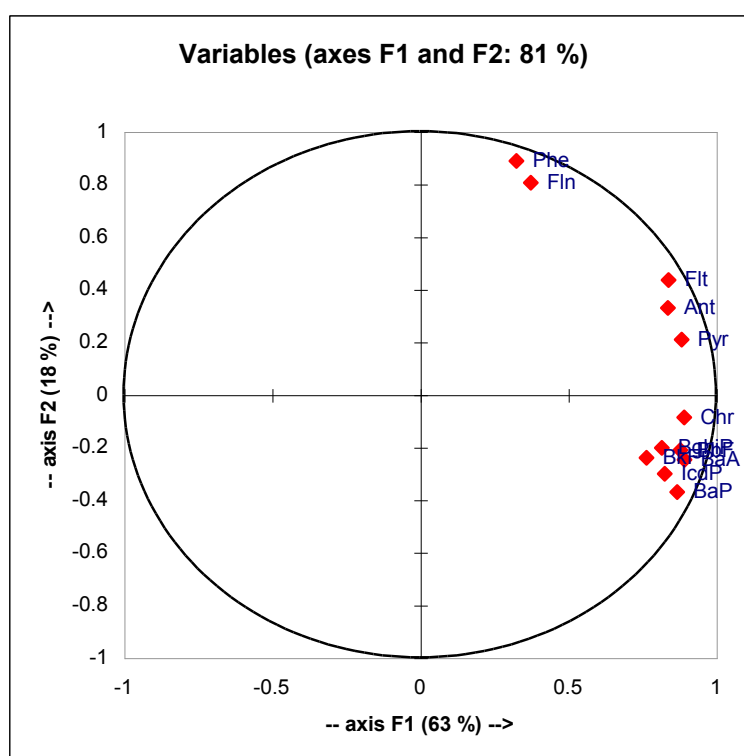
Tabulka II Základní statistické charakteristiky souboru PAH – léto

PAH	průměr	SD	RSD
Fln	7.9	5.3	67.2%
Phe	18.9	10.0	53.0%
Ant	1.0	0.7	70.8%
Flt	5.1	3.3	64.1%
Pyr	2.8	2.2	79.3%
BaA	0.4	0.4	101.4%
Chr	0.3	0.3	112.6%
BbF	0.2	0.3	110.9%
BkF	0.1	0.1	72.0%
BaP	0.2	0.2	91.1%
BghiP	0.2	0.2	91.8%
IcdP	0.2	0.2	84.7%

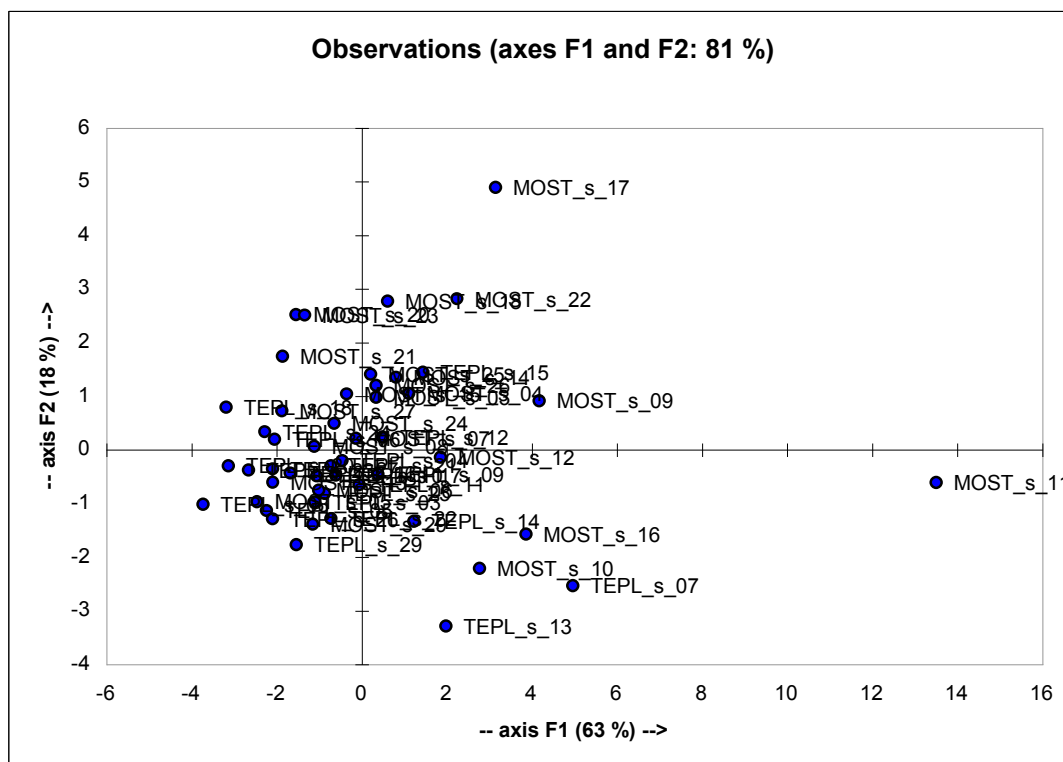
Vysoké hodnoty RSD opět, jako v předchozím případě, svědčí o značné nehomogenitě datové matice. Srovnáme-li hodnoty koncentrací PAH v létě a v zimě, vidíme řádově nižší hodnoty PAH v letním období. Je to zřejmě způsobeno tím, že velké energetické zdroje nejsou hlavním přispěvatelem PAH v životním prostředí.

Na obrázku 3 je graf korelační kružnice, znázorňující korelaci původních proměnných (PAH) s prvními dvěma hlavními komponentami. První dvě hlavní komponenty nesou 81% veškeré informace o tomto souboru, jsou tedy postačující pro dokonalý popis původních proměnných. Všechny proměnné vykazují vysokou vzájemnou korelaci, kromě Phe a Flt a zároveň jsou z 63% popsány první hlavní komponentou. Situace je tedy podstatně jiná než v zimě, což opět podporuje rozdílnost původu zdrojů v zimním a letním období.

Na obrázku 4 je graf skóre, který ukazuje rozdělení všech „vzorků“ (vektory pozorování) v prostoru prvních dvou hlavních komponent. Z obrázku je patrné, že vzorky netvoří pouze jeden shluk, pocházejí tedy z více emisních zdrojů. Rozptýlení je ovšem nižší než v případě souboru PAH – zima.



Obr. 3 Korelační kružnice pro soubor PAH-léto



Obr. 4 Graf skóre pro soubor PAH - léto

Soubor Kovy – zima

Soubor obsahoval 58 datových vektorů (TEPLICE, MOST). Každý z vektorů je charakterizován 11 koncentracemi kovů (ng/m^3). χ^2 – test potvrdil vysoký stupeň korelace mezi jednotlivými proměnnými ($\chi^2 = 695(55)$, $\chi^2(55) = 73$). Základní statistické charakteristiky souboru jsou uvedeny v následující Tabulce III.

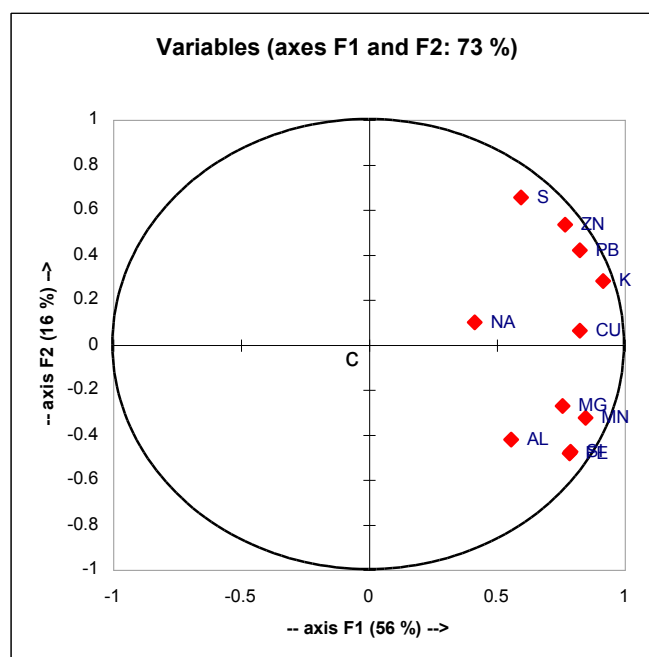
Tabulka III Základní statistické charakteristiky souboru Kovy – zima

Kovy	Mean	SD	RSD
Na	127.1	121.3	95.4%
Mg	23.3	28.3	121.9%
Al	113.3	176.7	155.9%
Si	203.1	241.7	119.0%
S	1366.2	1012.5	74.1%
K	125.9	102.2	81.2%
Mn	3.5	4.0	114.8%
Fe	122.7	122.6	99.9%
Cu	5.0	4.7	93.0%
Zn	43.1	35.1	81.4%
Pb	15.6	15.8	100.8%

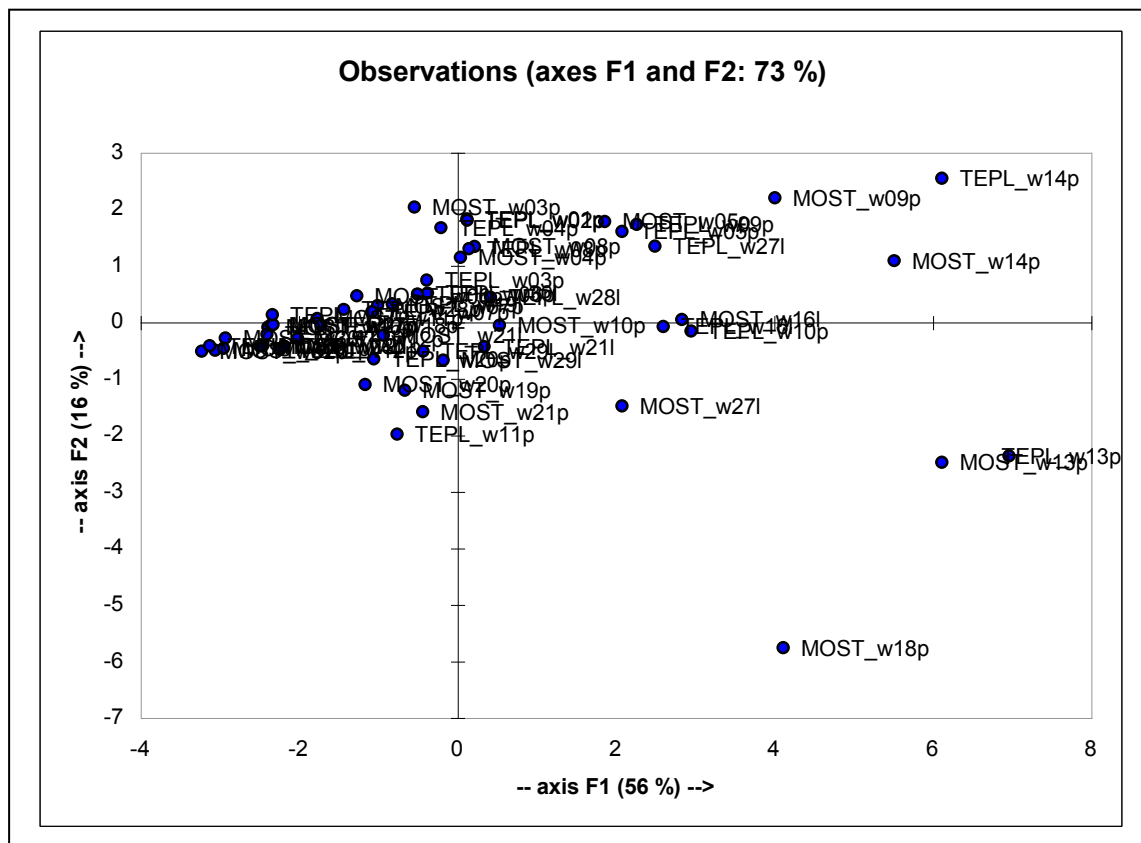
Soubor je charakterizován vysokými obsahy síry a křemíku a jednotlivé parametry jsou značně rozptýleny. Stejně jako v případě PAH to svědčí o tom, že k imisním koncentracím přispívá více zdrojů.

Na obrázku 5 je znázorněna korelační kružnice pro první dvě hlavní komponenty, které nesou 73% veškeré informace. Zajímavé jsou dvě skupiny kovů které spolu vysoce korelují. V první skupině to jsou S, Zn, Pb a K, ve druhé skupině pak Mg, Mn, Al Fe a Si. Cu stojí poněkud osamoceně i vzhledem k nízkým koncentracím ve všech vzorcích. Na vykazuje nízkou korelaci se všemi kovy.

Na obrázku 6 je ukázán graf skóre pro první dvě hlavní komponenty. Z grafu je patrné, že data tvoří dva velké shluky a několik osamocených bodů, které mohou svědčit o náhlých změnách poměrů při monitorování. Tyto vychýlené body jsou pak charakterizovány vysokými koncentracemi prakticky všech kovů, což bude potvrzeno v dalším zpracování.



Obr. 5 Korelační kružnice pro soubor Kovy - zima



Obr. 6 Graf skóre pro soubor Kovy - zima

Soubor Kovy - léto

Soubor obsahoval 50 datových vektorů (TEPLICE, MOST). Každý z vektorů je charakterizován 12 koncentracemi kovů (ng/m^3). χ^2 – test potvrdil vysoký stupeň korelace mezi jednotlivými proměnnými ($\chi^2 = 855(66)$, $\chi^2(66) = 86$). Základní statistické charakteristiky souboru jsou uvedeny v následující Tabulce IV.

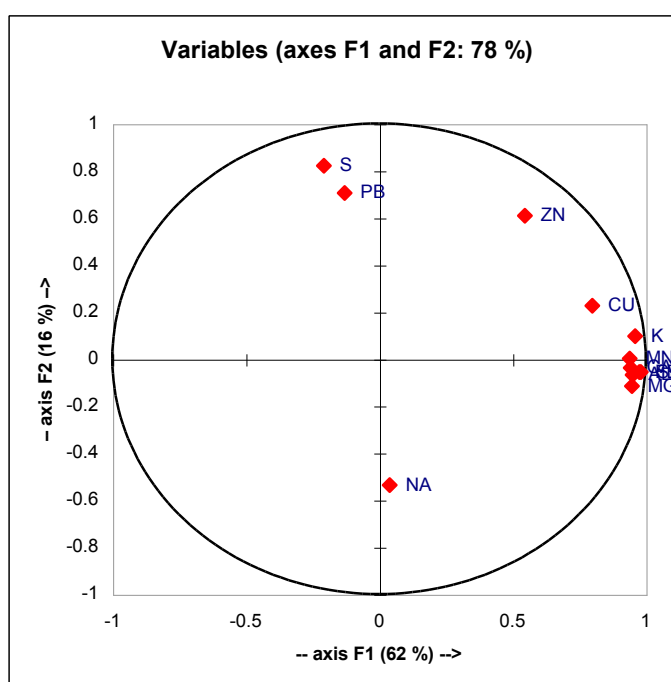
Tabulka IV Základní statistické charakteristiky souboru Kovy – léto

Kovy	Mean	SD	RSD
Na	76.5	57.0	74.5%
Mg	23.3	35.6	152.7%
Al	127.4	228.3	179.2%
Si	270.0	505.4	187.2%
S	1223.2	524.4	42.9%
K	74.8	70.6	94.3%
Ca	85.7	160.7	187.5%
Mn	3.9	7.3	187.0%
Fe	167.6	355.5	212.2%
Cu	3.3	2.2	69.0%
Zn	15.8	5.8	37.0%
Pb	5.8	5.2	89.6%

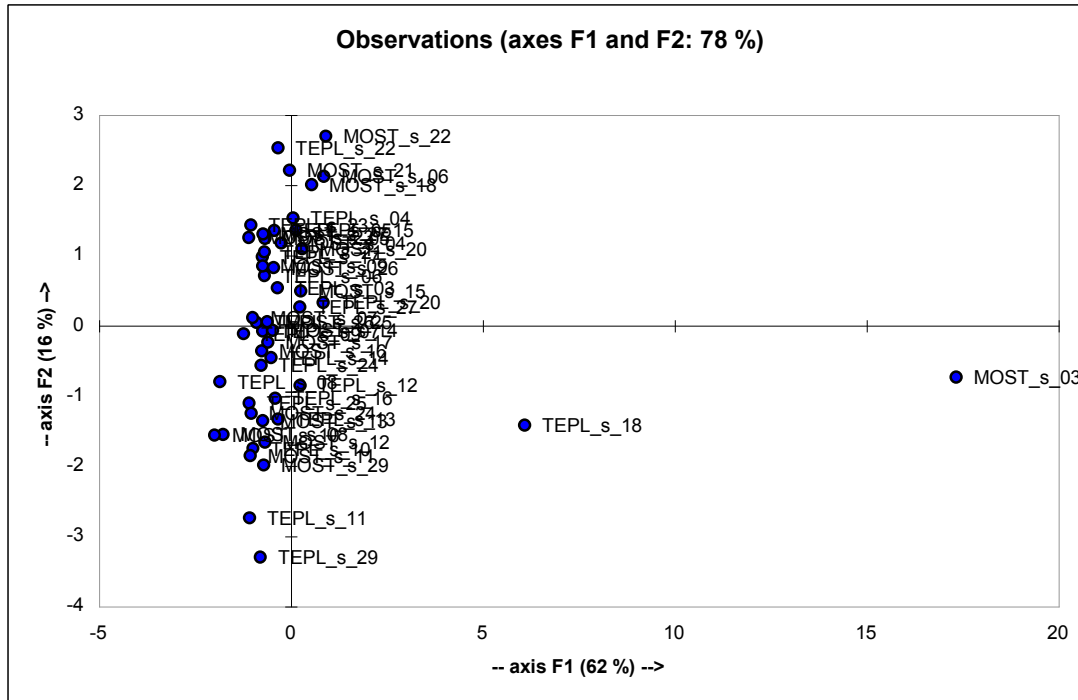
Charakteristika základních statistických parametrů je podobná situaci v souboru Kovy – zima s tím rozdílem, že v souboru Kovy – léto je koncentrace vápníku významně vyšší než v předchozím souboru. Je rovněž zajímavé, že letní soubor vykazuje nižší koncentraci Na a Pb.

Na obrázku 7 je graf korelační kružnice. Z obrázku je vidět, že S a Pb tvoří skupinu kovů, která stojí osamocně a zřejmě charakterizuje některé výjimečné epizody během monitorování.

Z obrázku 8 je zřejmé, že soubor se rozpadá na dva velké shluky a několik osamocených vzorkovacích míst. Zajímavá je také skutečnost, že první hlavní komponenta popisuje 62% variability a druhá 16%. To svědčí rovněž o rozdílech v letním a zimním období.



Obr. 7 Korelační kružnice pro soubor Kovy - léto



Obr. 8 Graf skóre pro soubor Kovy - léto

4.2 Klastrová analýza a lineární diskriminační analýza

Klastrová analýza byla použita pro zařazení jednotlivých vektorů pozorování do příslušných skupin. Lineární diskriminační analýza (LDA) totiž potřebuje, jako vstupní údaj, znalost předběžného rozdělení do jednotlivých skupin. Tyto vstupní informace byly v LDA dále iterovány, až se dospělo k nejlepšímu rozdělení. Charakteristiky jednotlivých skupin sloužili k odhadu charakteristik emisních zdrojů.

Soubor PAH – zima

Soubor byl klastrovou analýzou rozdělen do tří shluků. V Tabulce V je ukázáno nejlepší rozdělení, které vzniklo v procesu iterativního výpočtu LDA. V tabulce VI jsou uvedeny charakteristiky jednotlivých shluků (těžiště shluku). Těmito charakteristikám se někdy říká také prototypy.

Tabulka V Rozdělení souboru PAH – zima do shluků

25	9	24
TEPL_w16l	TEPL_w27l	TEPL_w21l
TEPL_w19l	TEPL_w04p	TEPL_w11p
TEPL_w28l	TEPL_w05p	TEPL_w12p
TEPL_w29l	TEPL_w09p	TEPL_w16p
TEPL_w30l	TEPL_w10p	TEPL_w17p
TEPL_w01p	TEPL_w14p	TEPL_w18p
TEPL_w02p	MOST_w05p	TEPL_w19p
TEPL_w03p	MOST_w09p	TEPL_w20p
TEPL_w06p	MOST_w14p	TEPL_w21p
TEPL_w07p		TEPL_w27p
TEPL_w08p		TEPL_w28p
TEPL_w13p		MOST_w19l
MOST_w16l		MOST_w21l
MOST_w27l		MOST_w28l
MOST_w29l		MOST_w08p
MOST_w30l		MOST_w12p
MOST_w01p		MOST_w16p
MOST_w02p		MOST_w17p
MOST_w03p		MOST_w18p
MOST_w04p		MOST_w19p
MOST_w06p		MOST_w20p
MOST_w07p		MOST_w21p
MOST_w10p		MOST_w27p
MOST_w11p		MOST_w28p
MOST_w13p		

Označení vzorků:

místo (TEPLICE, MOST)

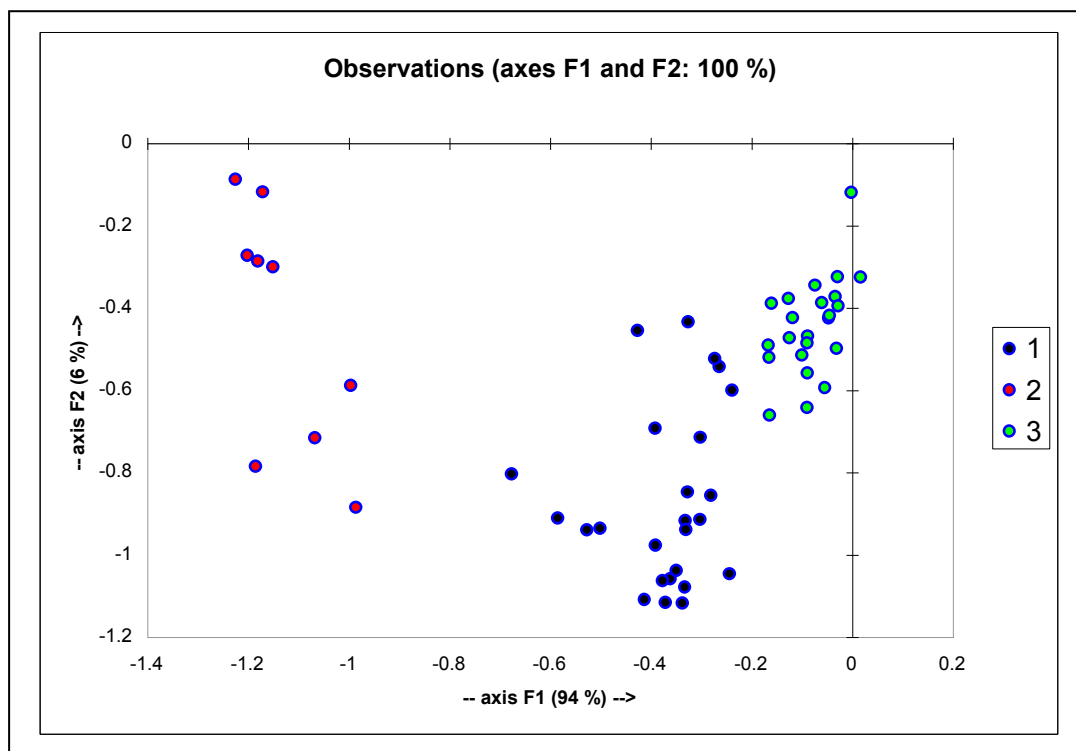
období w (zima)

datum odběru xxl,p (den, listopad nebo prosinec)

Tabulka VI Charakteristiky shluků (prototypů)

prototyp	Fln	Phe	Ant	Flt	Pyr	BaA	Chr
1	75.0	159.3	14.2	57.4	50.8	14.1	15.5
2	168.8	302.0	43.4	135.6	126.5	36.6	31.7
3	31.5	82.8	5.5	27.0	22.7	4.2	4.8
prototyp	BbF	BkF	BaP	DBahA	BghiP	lcdP	
1	11.8	5.9	9.0	1.1	8.5	9.2	
2	26.1	12.6	23.5	2.7	19.9	20.3	
3	3.4	1.5	2.2	0.3	2.6	2.8	

Na obrázku 9 je ukázán výsledek LDA, tj. rozdělení do jednotlivých shluků v prostoru prvních dvou hlavních komponent.



Obr. 9 Výsledek LDA pro soubor PAH – zima

Z výsledku LDA jasně vyplývá, že soubor PAH – zima (imisiční data) je tvořen třemi shluky a dá se tedy předpokládat, že situace na monitorovacích místech je ovlivňována minimálně třemi emisními zdroji.

Soubor PAH – léto

Soubor byl klastrovou analýzou rozdělen do čtyř shluků. V Tabulce VII je ukázáno nejlepší rozdělení, které vzniklo v procesu iterativního výpočtu LDA. V Tabulce VIII jsou uvedeny charakteristiky jednotlivých shluků

Tabulka VII Rozdělení souboru PAH – léto do shluků

11	24	12	3
TEPL_s_04	TEPL_s_03	TEPL_s_15	MOST_s_17
TEPL_s_12	TEPL_s_05	MOST_s_03	MOST_s_18
TEPL_s_17	TEPL_s_06	MOST_s_04	MOST_s_22
TEPL_s_18	TEPL_s_07	MOST_s_09	
TEPL_s_21	TEPL_s_08	MOST_s_11	
MOST_s_07	TEPL_s_09	MOST_s_14	
MOST_s_08	TEPL_s_10	MOST_s_15	
MOST_s_12	TEPL_s_11	MOST_s_20	
MOST_s_16	TEPL_s_13	MOST_s_21	
MOST_s_24	TEPL_s_14	MOST_s_23	
MOST_s_27	TEPL_s_16	MOST_s_25	
	TEPL_s_20	MOST_s_26	
	TEPL_s_22		
	TEPL_s_23		
	TEPL_s_24		
	TEPL_s_25		
	TEPL_s_26		
	TEPL_s_27		
	TEPL_s_29		
	MOST_s_05		
	MOST_s_06		
	MOST_s_10		
	MOST_s_13		
	MOST_s_29		

Označení vzorků:

místo (TEPLICE, MOST)

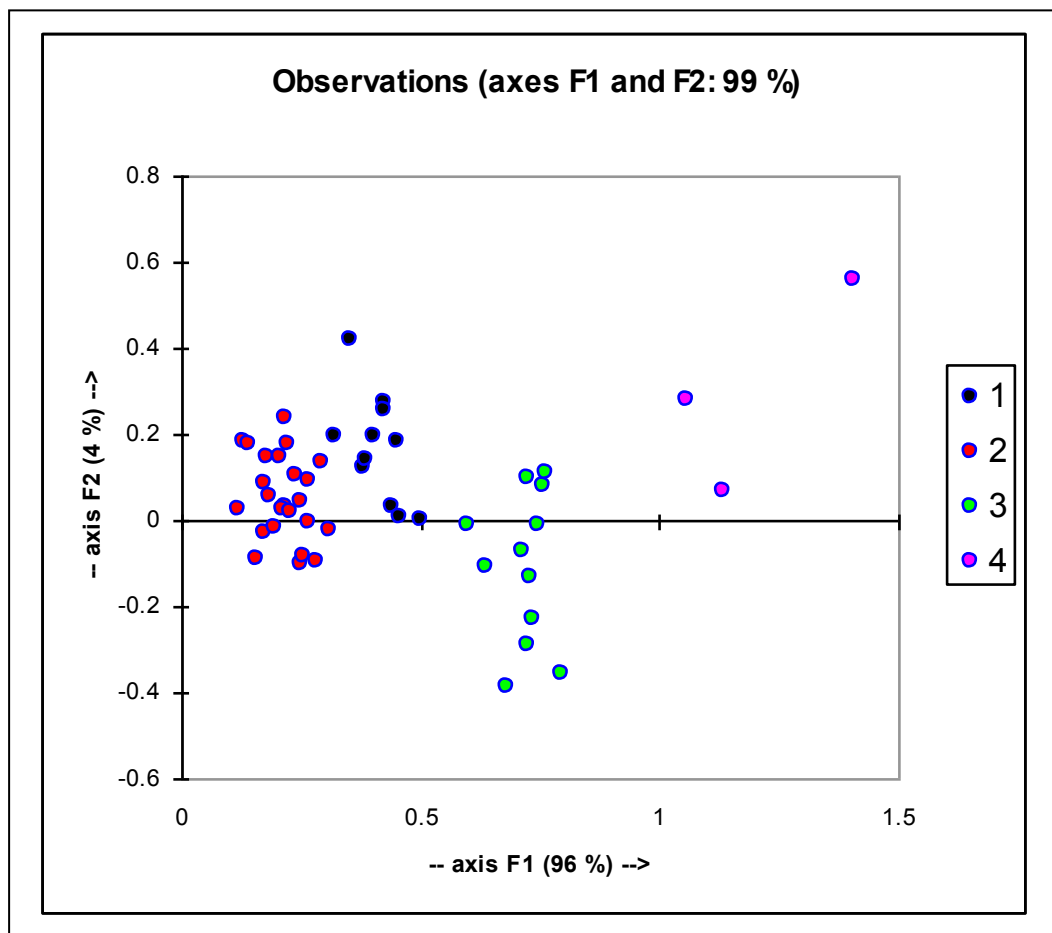
období s(léto)

datum odběru xx (den v červnu)

Tabulka VIII Charakteristiky shluků (prototypů)

prototyp	Fln	Phe	Ant	Flt	Pyr	BaA
1	7.89	19.70	0.78	4.73	2.63	0.41
2	4.17	11.15	0.66	3.25	1.91	0.38
3	11.52	27.43	1.60	8.25	4.20	0.60
4	23.60	44.63	1.54	9.01	4.39	0.26
	Chr	BbF	BkF	BaP	BghiP	IcdP
1	0.23	0.25	0.15	0.18	0.23	0.18
2	0.24	0.19	0.13	0.17	0.21	0.17
3	0.50	0.29	0.14	0.21	0.24	0.19
4	0.27	0.30	0.20	0.22	0.36	0.26

Na obr. 10 je ukázáno rozdělení do čtyř shluků jako výsledek LDA.



Obr. 10 Výsledek LDA pro soubor PAH – léto

Soubor Kovy – zima

Soubor byl klastrovou analýzou rozdělen do tří shluků. V Tabulce IX je ukázáno nejlepší rozdělení, které vzniklo v procesu iterativního výpočtu LDA. V Tabulce X jsou uvedeny charakteristiky jednotlivých shluků

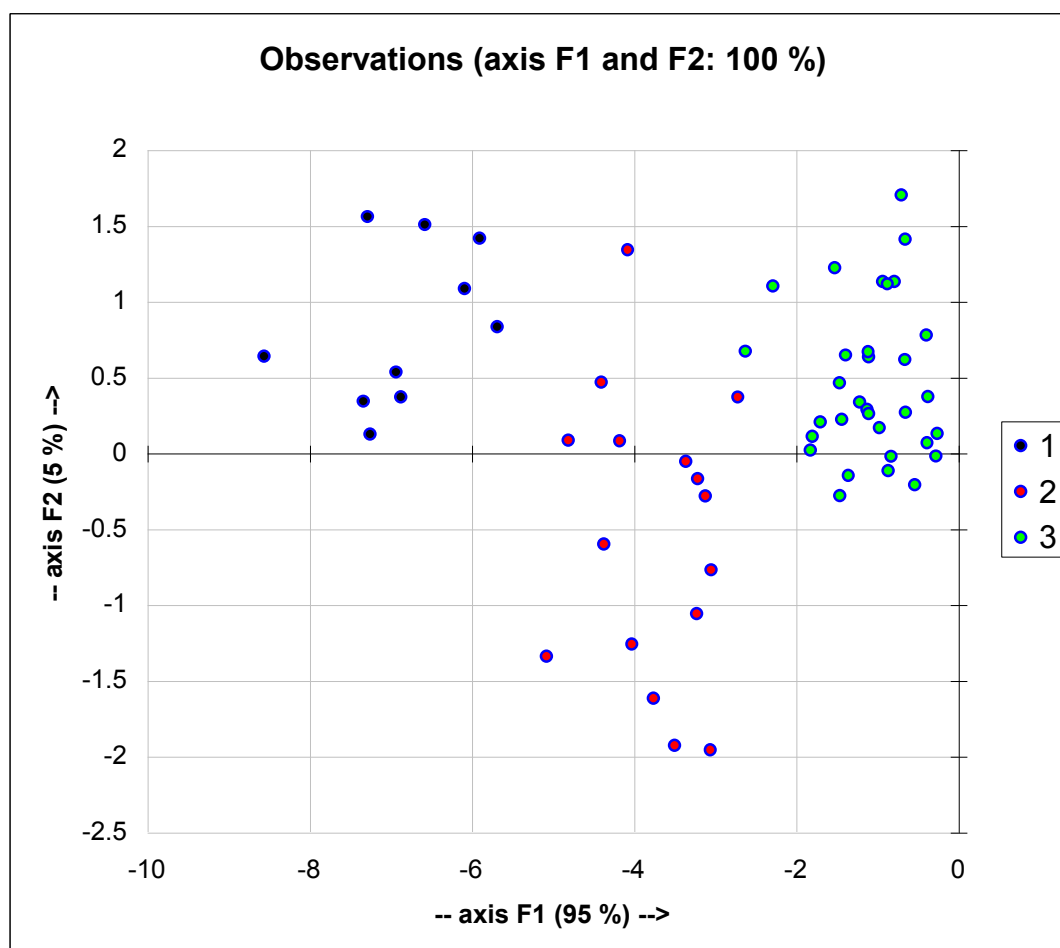
Tabulka IX Rozdělení souboru Kovy - zima do shluků

10	16	32
MOST_w03p	MOST_w16l	MOST_w19l
MOST_w05p	MOST_w27l	MOST_w21l
MOST_w09p	MOST_w04p	MOST_w28l
MOST_w14p	MOST_w06p	MOST_w29l
TEPL_w27l	MOST_w07p	MOST_w30l
TEPL_w01p	MOST_w08p	MOST_w01p
TEPL_w02p	MOST_w10p	MOST_w02p
TEPL_w05p	MOST_w13p	MOST_w11p
TEPL_w09p	TEPL_w21l	MOST_w12p
TEPL_w14p	TEPL_w30l	MOST_w16p
	TEPL_w03p	MOST_w17p
	TEPL_w04p	MOST_w18p
	TEPL_w07p	MOST_w19p
	TEPL_w08p	MOST_w20p
	TEPL_w10p	MOST_w21p
	TEPL_w13p	MOST_w27p
		MOST_w28p
		TEPL_w16l
		TEPL_w19l
		TEPL_w28l
		TEPL_w29l
		TEPL_w06p
		TEPL_w11p
		TEPL_w12p
		TEPL_w16p
		TEPL_w17p
		TEPL_w18p
		TEPL_w19p
		TEPL_w20p
		TEPL_w21p
		TEPL_w27p
		TEPL_w28p

Tabulka X Charakteristiky shluků (prototypů)

Prototyp	Na	Mg	Al	Si	S	K
1	150.0	24.2	121.8	267.7	3109.4	265.6
2	141.1	38.3	206.4	286.8	1757.8	166.9
3	112.3	14.7	60.6	136.7	604.1	59.3
Prototyp	Mn	Fe	Cu	Zn	Pb	
1	5.2	169.1	8.5	95.8	34.1	
2	4.5	143.2	6.9	50.1	23.5	
3	2.3	95.5	2.9	22.4	5.7	

Na obr. 11 je ukázán výsledek LDA a rozdělení do tří shluků.



Obr. 11 Výsledek LDA pro soubor Kovy - zima

Soubor Kovy – léto

Soubor byl klastrovou analýzou rozdělen do tří shluků. V Tabulce XI je ukázáno nejlepší rozdělení, které vzniklo v procesu iterativního výpočtu LDA. V Tabulce XII jsou uvedeny charakteristiky jednotlivých shluků

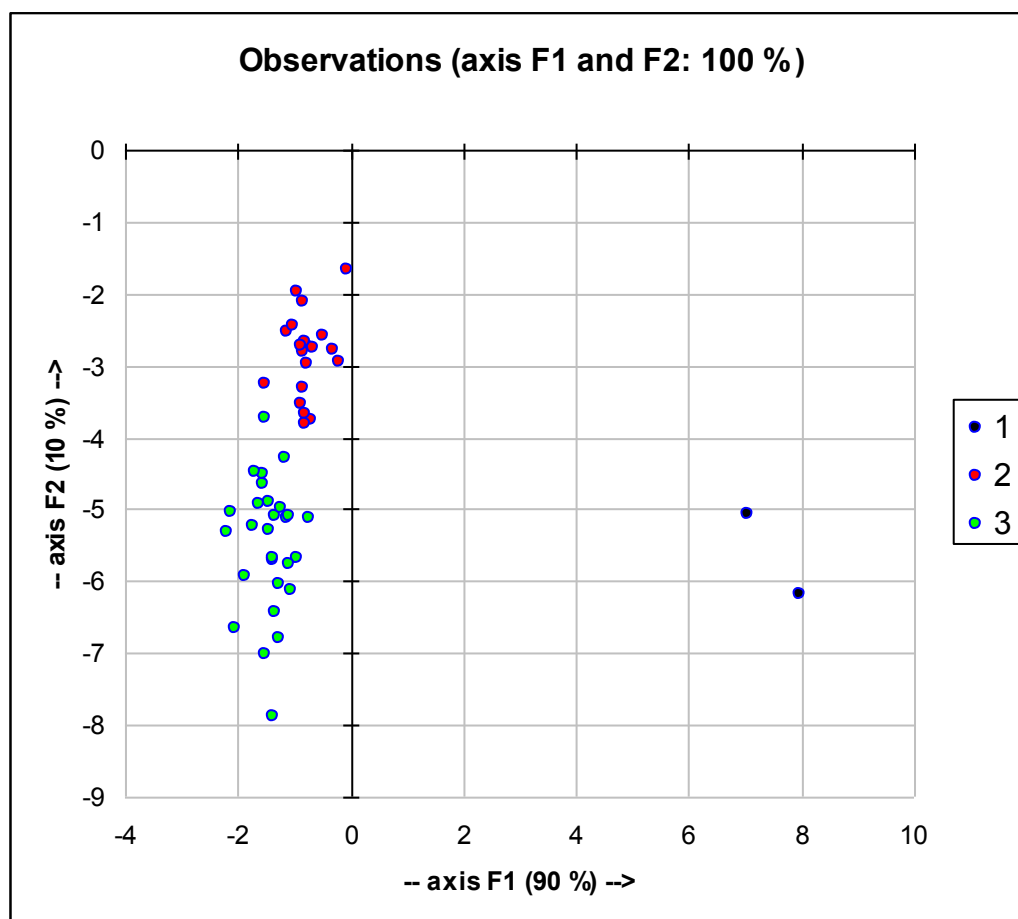
Tabulka XI Rozdělení souboru Kovy - léto do shluků

2	20	28
TEPL_s_18	TEPL_s_03	TEPL_s_04
MOST_s_03	TEPL_s_07	TEPL_s_05
	TEPL_s_08	TEPL_s_06
	TEPL_s_11	TEPL_s_09
	TEPL_s_12	TEPL_s_10
	TEPL_s_13	TEPL_s_15
	TEPL_s_14	TEPL_s_16
	TEPL_s_24	TEPL_s_17
	TEPL_s_25	TEPL_s_20
	TEPL_s_29	TEPL_s_21
	MOST_s_08	TEPL_s_22
	MOST_s_10	TEPL_s_23
	MOST_s_11	TEPL_s_26
	MOST_s_12	TEPL_s_27
	MOST_s_13	MOST_s_04
	MOST_s_14	MOST_s_05
	MOST_s_15	MOST_s_06
	MOST_s_24	MOST_s_07
	MOST_s_25	MOST_s_09
	MOST_s_29	MOST_s_16
		MOST_s_17
		MOST_s_18
		MOST_s_20
		MOST_s_21
		MOST_s_22
		MOST_s_23
		MOST_s_26
		MOST_s_27

Tabulka XII Charakteristiky shluků (prototypů)

Prototyp	Na	Mg	Al	Si	S	K
1	82.7	170.9	1066.5	2398.3	415.7	352.2
2	87.2	17.2	68.2	159.9	773.9	48.7
3	68.5	17.2	102.6	196.7	1601.8	73.7
Prototyp	Ca	Mn	Fe	Cu	Zn	Pb
1	803.6	37.7	1606.7	10.2	24.6	0.0
2	41.2	2.1	95.9	2.6	11.7	3.9
3	66.2	2.8	115.9	3.2	18.1	7.6

Na obr. 12 je ukázán výsledek LDA pro soubor Kovy – léto

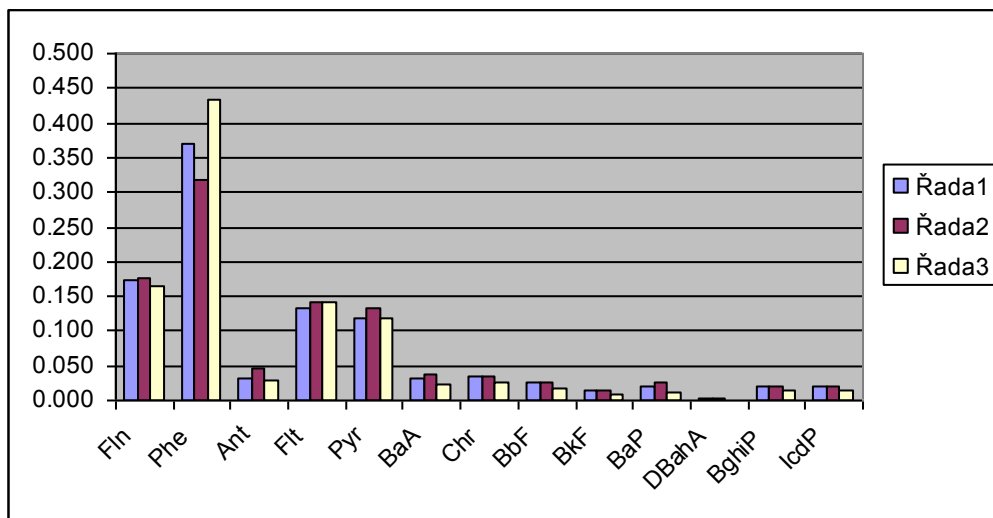


Obr. 12 Výsledek LDA pro soubor Kovy – léto

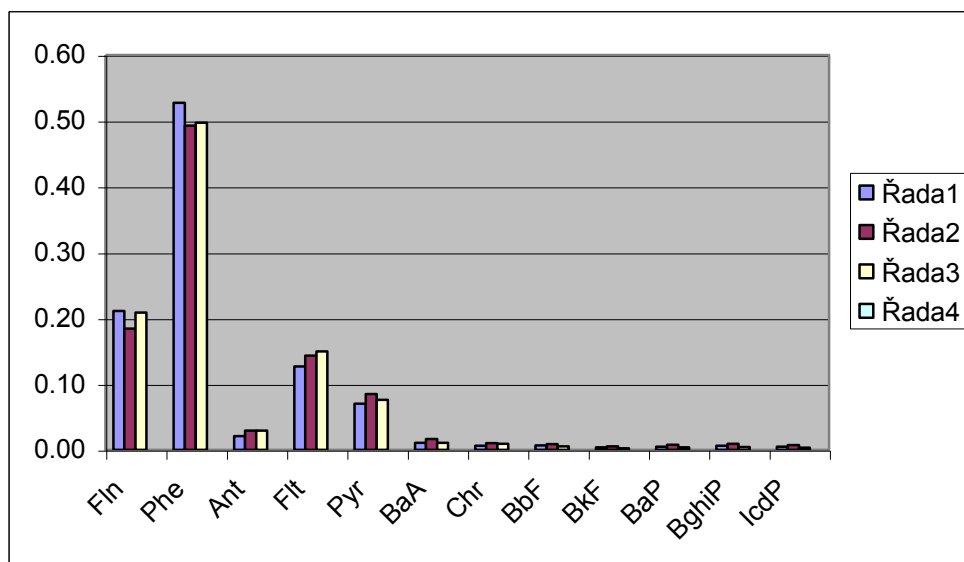
Srovnání prototypů jednotlivých souborů s podpisem zdroje

Na obr. 13 – 15 je uvedeno srovnání prototypů vypočtených LDA a podpisem zdroje. Ve všech případech bylo použito poměrné zastoupení jednotlivých složek jak v prototypech, tak v podpisech zdrojů.

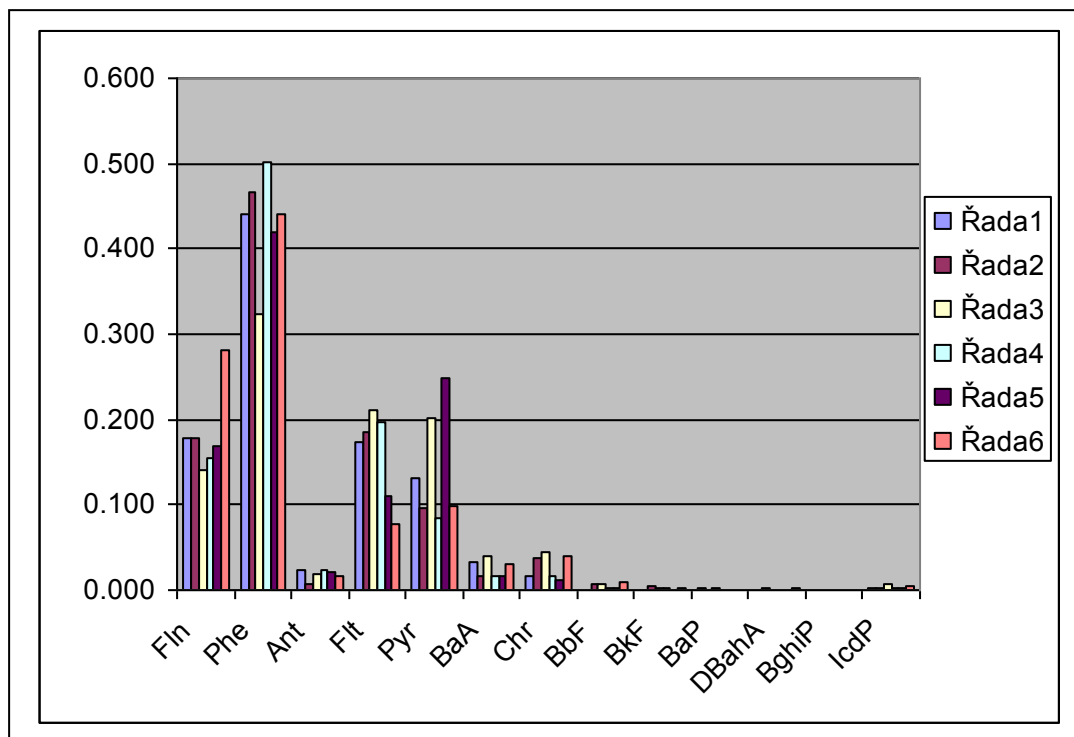
Soubory pro PAH



Obr. 13 Poměrné zastoupení PAH v souboru PAH - zima



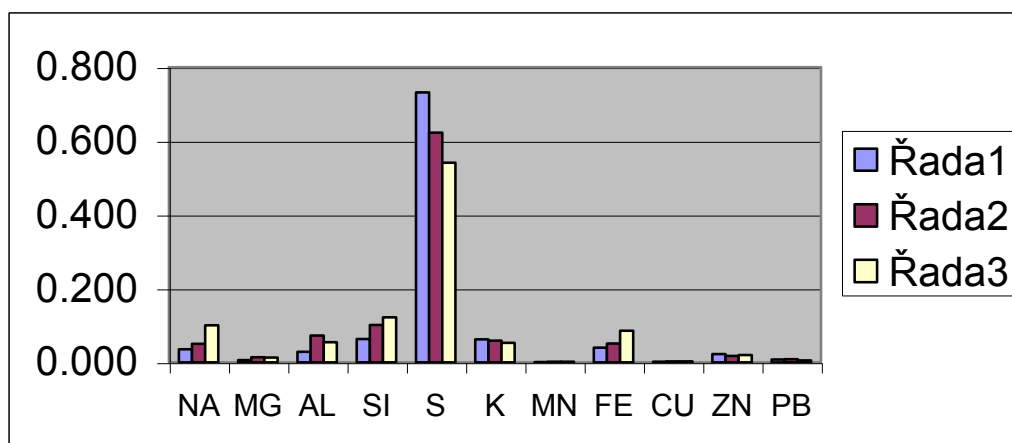
Obr. 14 Poměrné zastoupení PAH v souboru PAH – léto



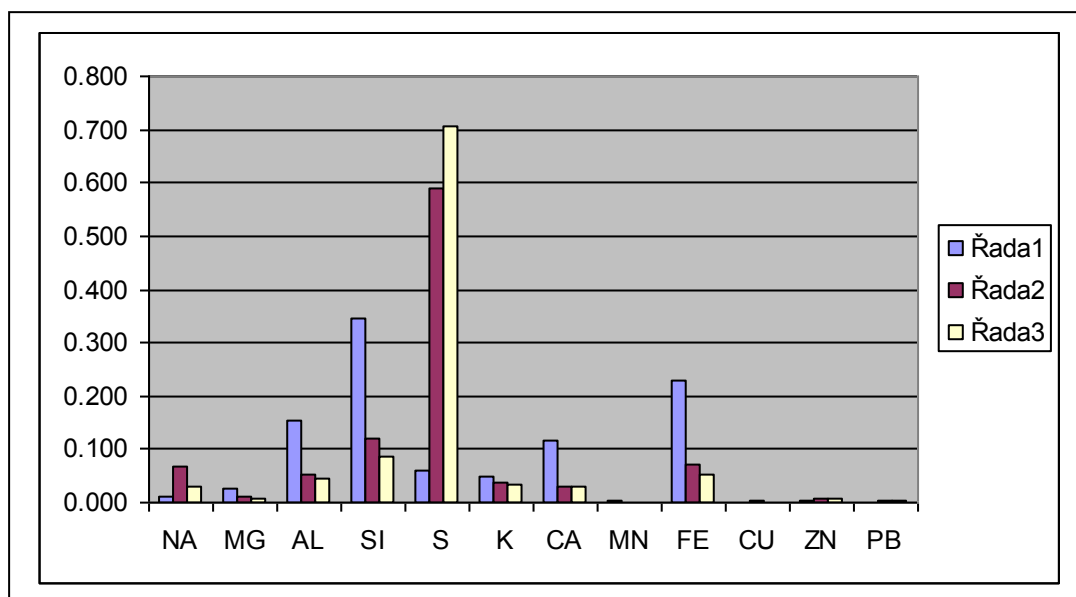
Obr. 15 Poměrné zastoupení PAH ve zdrojích emisí

1 – Dekonta; 2 - Ledvice_CEZ; 3 - CHZ Litvínov; 4 - Počeradý_CEZ;
5 – Trmice; 6 - Komořany

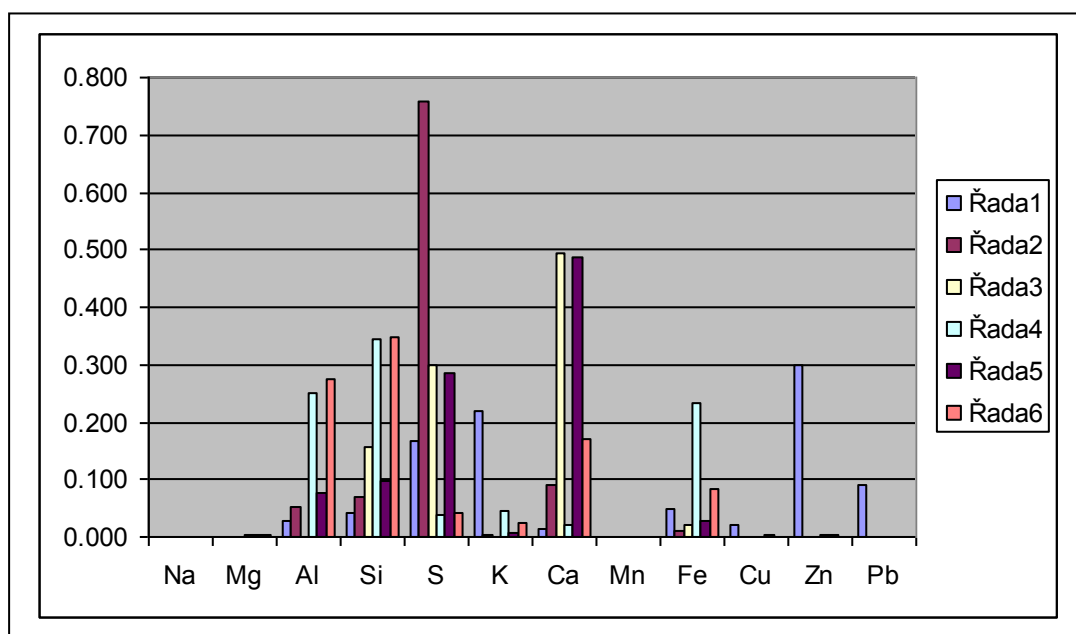
Soubory pro kovy



Obr. 16 Poměrné zastoupení kovů v souboru Kovy – zima



Obr. 17 Poměrné zastoupení kovů v souboru Kovy – léto



Obr. 18 Poměrné zastoupení kovů v emisních zdrojích

1 – Dekonta; 2 - Ledvice_CEZ; 3 - CHZ Litvínov; 4 - Počeradý_CEZ;
5 – Trmice; 6 - Komořany

Charakteristické indexy

Pro odhad původu zdroje emisí PAH se někdy používají tzv. indexy. Jsou to vlastně poměry koncentrací některých polyaromatických uhlovodíků. Nejčastěji se používají poměry **BaP/BghiP** a **PyR/BaP**. Z dřívějších měření je známo, že hodnoty poměru **BaP/BghiP** pro zdroje z dopravy nabývají hodnot 0,2 – 0,4, kdežto pro zdroje z lokálních topenišť jsou hodnoty větší než 0,8. Pro poměr **PyR/BaP** se hodnoty indexu pro dopravu pohybují v rozmezí 2 – 12, pro lokální topeniště 0,2 – 0,8. Vedle toho se dá použít index **Phe/Ant**, který indikuje např. zpracování dehtových frakcí. V následující tabulce XIII jsou vypočteny indexy pro jednotlivé prototypy v zimním a letním období.

Tabulka XIII Indexy pro jednotlivé prototypy souborů PAH – zima, léto

PAH - zima			
Prototyp	BaP/BghiP	PyR/BaP	Phe/Ant
1	1.1	5.6	11.2
2	1.2	5.4	7.0
3	0.8	10.3	15.1
PAH - léto			
Prototyp	BaP/BghiP	PyR/BaP	Phe/Ant
1	0.8	14.4	25.2
2	0.8	11.0	16.9
3	0.9	19.7	17.1
4	0.6	20.3	29.0
Zdroje		PyR/BaP	Phe/Ant
Dekonta		60.1	2.0
Ledvice_CEZ			6.6
CHZ Litvínov		128.2	1.6
Počerady_CEZ			2.1
Trmice			2.1
Komořany			2.5

I z této tabulky je zřejmé, že imisní data nemohou pocházet z velkých energetických zdrojů.

4.3 Receptorové modelování

Jako příklad receptorového modelování byla zvolena tzv. centrální pozorování (těžiště) jednotlivých klastrů. Výsledky modelování s využitím CMB 8.2 (US EPA) jsou uvedeny v přílohách.

Z výsledků je zřejmé, že ani v jednom případě se nedosáhlo dobré shody experimentu s modelem. Znamená to, jak ostatně ukázaly předchozí výpočty, že velké energetické zdroje nejsou příčinou emisí uhlovodíků i kovů.

5 Souhrn

Soubory, pocházející z měření v severních Čechách byly podrobeny zpracování multivariační analýzou. Bylo zjištěno, že emisní data pocházejí z minimálně tří emisních zdrojů, z nichž ani jeden nepatří mezi velké energetické zdroje. Další zpracování naměřených dat musí proto zahrnovat i tyto zdroje, jejichž podpis se bude muset získat z některých předchozích měření.