

ČASOPIS  
STUDIA OECOLOGICA  
Ročník VI  
Číslo 1/2012

**Redakční rada:**

doc. Ing. Pavel Janoš, CSc. – šéfredaktor  
doc. Ing. Miroslav Farský, CSc. – výkonný redaktor  
prof. RNDr. Olga Kontrišová, CSc.  
doc. RNDr. Juraj Lesný, Ph.D.  
Ing. Martin Neruda, Ph.D.  
doc. MVDr. Pavel Novák, CSc.  
prof. Ing. Miloslav Šoch, CSc.

**Technický redaktor:**

Mgr. Ing. Petr Novák

**Recenzenti:**

doc. RNDr. Peter Andráš, CSc., Univerzita Mateja Bela, Slovensko  
RNDr. František Eichler, Ph.D., Liberec  
Mgr. Ladislava Filipová, Ph.D., Ústí nad Labem  
doc. RNDr. Jaromír Hajer, CSc., PřF UJEP, Ústí nad Labem  
prom. biolog Jiří Heteša, CSc., Agronomická fakulta Mendelovy univerzity, Brno  
Ing. Jana Hubáčková, Výzkumný ústav vodohospodářský T. G. Masaryka, Praha  
Ing. Radoslav Kandrik, Ph.D., TU Zvolen, Slovensko  
RNDr. Ján Kliment, CSc., Botanická zahrada Univerzity Komenského v Blatnici, Slovensko  
Ing. Jan Leníček, Zdravotní ústav, Ústí nad Labem  
Ing. David Milde, Ph.D., PřF UPOL, Olomouc  
Mgr. Antonín Roušar, Chomutov  
doc. Ing. Josef Seják, CSc., FŽP UJEP, Ústí nad Labem  
RNDr. Lucia Šolcová, Ph.D., Univerzita Konštantína Filozofa v Nitre, Slovensko  
Ing. Stanislav Štýs, DrSc., Most

**Foto obálky**

doc. Ing. Jana Říhová Ambrožová, Ph.D.

Vydává: FŽP UJEP v Ústí nad Labem  
Tisk: Ofsetový tisk Miroslav Jedlička

Toto číslo bylo dáno do tisku v červenci 2012  
ISSN 1802-212X  
MK ČR E 17061

# IDENTIFIKACE ZDROJŮ KOVŮ, POLYCYKlickÝCH AROMATICKÝCH UHLOVODÍKU A PLYNNÝCH ŠKODLIVIN METODOU PMF

## IDENTIFICATION OF SOURCE OF METALS, POLYCYCLIC AROMATIC HYDROCARBONS AND GASEOUS POLLUTANTS BY PMF

Jan POPELKA

Univerzita J. E. Purkyně, Fakulta životního prostředí, Králova výšina 7, Ústí nad Labem, 400 96, Česká republika, jan.popelka@ujep.cz

### Abstrakt

Metoda PMF (Positive Matrix Factorization) umožňuje identifikaci zdrojů znečištění ovzduší z hodnot imisních koncentrací v místě měření. Znalost profilů konkrétních zdrojů není nutná. Podmínkou použití PMF je dostatečný počet měření s dostatečným počtem polutantů a znalost nejistot měření. Vedle samotné identifikace zdrojů lze navíc kvantifikovat příspěvek zdrojů k imisní zátěži v místě měření a sledovat jeho změny v čase. Metoda PMF byla v tomto článku použita k identifikaci a kvantifikaci zdrojů znečištění kovy v Litoměřicích a kovy, plynnými polutanty a polycyklickými uhlovodíky (PAH) v Praze-Libuši. Pro volbu počtu faktorů (zdrojů) je popsáno několik postupů, z nichž byly aplikovány metoda hlavních komponent (PCA) a zpětná validace modelu PMF. Oba postupy nevedly vždy ke stejným závěrům o vhodném počtu faktorů.

### Abstract

PMF (Positive Matrix Factorization) enables the identification of sources of air pollution from airborne particulate data. Knowledge of specific profiles is not necessary. PMF requires sufficient number of data points with sufficient number of pollutants and measurement uncertainties. Besides the source identification, contribution of individual sources to air pollution can be quantified and tracked over time. PMF is in this article used to identify and quantify sources of metals in Litoměřice and sources of metals, gaseous pollutants and polycyclic aromatic hydrocarbons (PAH) in Praha-Libuš. To select a number of factors (sources), several techniques were described of which principal components analysis (PCA) and bootstrap validation of PMF model were applied. Both procedures didn't always let to similar conclusions about the appropriate number of factors.

**Klíčová slova:** znečištění ovzduší, kovy, organické látky, identifikace zdrojů, PMF

**Key words:** air pollution, metals, organic compounds, source apportionment, PMF

### Úvod

Díky svému dopadu na zdravotní stav obyvatelstva patří znečištění ovzduší dlouhodobě k výrazným problémům ochrany životního prostředí v České republice. Vedle velkých průmyslových zdrojů (elektrárny, povrchové doly, spalovny, teplárny, průmyslové podniky apod.) působí na kvalitu ovzduší stále více automobilová doprava a emise z lokálních topenišť obytných domů, jejichž podíl je v některých městech a vesnicích dominantní. Kvantifikací příspěvku jednotlivých zdrojů k celkové úrovni znečištění lze získat důležité informace využitelné k napravení stavu ovzduší v problematických oblastech. Řada stanic Českého hydrometeorologického ústavu a Zdravotních ústavů poskytuje data použitelná k tomuto účelu. Stačí jen aplikovat vhodné statistické metody a postupy k jejich zpracování.

## Identifikace zdrojů znečištění

### Receptorové modelování

Receptorové modelování nabízí statistické nástroje pro kvantifikaci podílu zdrojů na imisích v místě měření. Receptorové modelování je statistická metoda založená na znalosti profilu zdroje, tj. charakteristickém poměru analytů v konkrétním zdroji. Úřad pro ochranu životního prostředí Spojených států amerických (US EPA) vyvinul speciální software Chemical Mass Balance (EPA-CMB), který byl využit k analýze koncentrací PAH z měřicí stanice Zdravotního ústavu (ZÚ) v Ústí nad Labem (Leníček a kol., 2011). Receptorovým modelováním na Ostravsku se zabýval Velíšek (2007). Studie receptorového modelování z jihovýchodní části USA publikovali Zheng a kol. (2002) a ovzduší v Kalifornii se věnovali Schauer a kol. (Schauer a kol., 1996); (Schauer a kol., 2000). Receptorové modelování je hojně využívaným nástrojem kvantifikace zdrojů znečištění. Profily konkrétních zdrojů však často nejsou známy a jejich výzkum stále probíhá. Bez konkrétní znalosti profilů zdrojů nelze metodu receptorového modelování úspěšně aplikovat.

### Vícerozměrné metody

Zdroje znečištění lze identifikovat i bez detailní znalosti profilů zdrojů. Statistická analýza vícerozměrných dat nabízí nástroje, které lze k identifikaci a kvantifikaci použít. Založeny jsou na myšlence, že se koncentrace polutantů v ovzduší v čase mění a tyto změny jsou z jisté části zapříčiněny proměnlivou intenzitou emisí jejich zdroje. Doprava má jinou intenzitu ve všedních dnech a o víkendech a reaguje na různá dopravní omezení, jako jsou nepříznivé podmínky, změny sjízdnosti, dopravní uzávěry a podobně. Lokální topeniště se provozují v závislosti na venkovní teplotě. Velké zdroje mohou vykazovat změny v provozu z důvodu odstávek, regulací apod. Jednorázově se mohou vyskytovat emise způsobené haváriemi nebo výjimečnými událostmi jako jsou v publikacích z USA často zmiňované ohňostroje (Norris a kol., 2009; Lingwall a Christensen, 2007). Výkyvy imisí v místě měření lze statisticky postihnout, pokud je k dispozici dostatečné množství měření (nejčastěji dostatečný počet dnů) a dostatečný počet analytů. Zdroje znečištění lze mimo jiné identifikovat metodou hlavních komponent (PCA), faktorovou analýzou (FA) nebo shlukovou analýzou. Více o těchto metodách nabízejí Hebák a kol. (2007) nebo Meloun a kol. (2005). Všechny zmíněné metody jsou schopny definovat skupiny analytů vykazujících ve zkoumaném období společný nebo podobný vývoj. Při analýze polycyklických aromatických uhlovodíků (dále jen PAH) v Ústí nad Labem byly faktorovou analýzou identifikovány dva zdroje (doprava a spalování fosilních paliv ve stacionárních a mobilních zdrojích). Model se dvěma faktory vysvětlil 81,7 % variability naměřených koncentrací PAH. Zdroje se stejným složením byly identifikovány i hierarchickou klastrovou analýzou (Leníček a kol., 2011).

### Metoda PMF

Vhodným nástrojem identifikace zdrojů znečištění je metoda PMF (Positive Matrix Factorization). Matici koncentrací analytů  $\mathbf{X}$  (prvek  $x_{ij}$  představuje koncentraci  $i$ -tého analytu v  $j$ -tém vzorku) lze metodou PMF rozložit na matice  $\mathbf{F}$  a  $\mathbf{G}$  (Hopke, 2000; Xie a kol., 1999) dle rovnice:

$$x_{ij} = \sum_{k=1}^p f_{ik} \cdot g_{kj}$$

Matice  $\mathbf{F}$  popisuje skladbu analytů v jednotlivých zdrojích ( $f_{ik}$  je koncentrace  $i$ -tého analytu v  $k$ -tém zdroji), tedy jde o odhady profilů zdrojů. V matici  $\mathbf{G}$  popisuje každý prvek  $g_{kj}$  příspěvek  $k$ -tého zdroje v  $j$ -tém vzorku měření. Při řešení PMF je stanoveno několik omezení nutných pro úspěšný rozklad a získání smysluplných výsledků: a) naměřená data musejí být popsitelná modelem, model musí být schopen vysvětlit zjištěná pozorování; b) odhadovaný profil zdroje musí být vyjádřen nezápornými hodnotami, ve výsledku nesmí obsahovat záporná procenta analytů; c) příspěvek zdroje k imisnímu zatížení nesmí být záporný, zdroj se nemůže na znečištění ovzduší v místě odběru podílet záporným podílem; d) suma příspěvků všech zdrojů k imisnímu zatížení musí být menší nebo nejvýše stejná

jako je samotné imisní zatížení, součet příspěvků nemůže být větší než celkové imisní zatížení v místě měření (Hopke, 2000).

Řešení získané metodami PCA a FA jednak nemusí vždy splňovat výše uvedená omezení a je formováno pouze hierarchicky. Zvýšením dimenze úlohy o další faktor nedojde ke změně složení faktorů předchozích. V PMF se faktory zcela změní, pokud dojde ke změně jejich počtu (Xie a kol., 1999). Dalším problémem metod PCA a FA je v přístupu k nejistotám měření. Obě metody používají souhrnnou nejistotu vypočtenou z celého sloupce nebo řádku dat. Pattero a Tapper (Paatero a Tapper, 1993) navrhli použít nejistoty měření definované pro každý datový bod zvlášť. Data jsou přesnější a větší vliv na řešení mají datové body s nízkými nejistotami. Data spolu s nejistotami však nelze analyzovat pomocí PCA ani FA, metodou PMF však ano. (Hopke, 2000). Nejistoty jsou v modelu využity jako bodové váhy jednotlivých měření (point-by-point estimates). Model je odhadován minimalizací objektové funkce:

$$Q = \sum_{i=1}^m \sum_{j=1}^n \left( \frac{e_{ij}}{\sigma_{ij}} \right)^2, \quad e_{ij} = x_{ij} - \sum_{k=1}^p f_{ik} \cdot g_{kj}$$

při dodržení Hopkem uvedených omezení  $f_{ik} \geq 0$ ,  $g_{kj} \geq 0$ , kde  $\sigma$  je známá matice nejistot (Hopke, 2000; Xie a kol., 1999). Forma faktorizace je zcela odlišná od metod PCA a FA. Jeden z možných algoritmů řešení úlohy je založen na simultánních změnách obou matic  $\mathbf{F}$  a  $\mathbf{G}$  v každém iteračním kroku. Detailně je algoritmus popsán v pracích (Paatero a Tapper, 1994; Paatero, 1997). Algoritmus je implementován v programu EPA-PMF vytvořeným Úřadem pro ochranu životního prostředí Spojených států amerických (US EPA).

Ze zadání PMF plyne, že jako datový zdroj slouží matice koncentrací analytů  $\mathbf{X}$ , kde každý sloupec je tvořen jedním analytem, řádky jsou jednotlivé vzorky (nejčastěji dny měření). Nutné jsou i nejistoty měření jednotlivých hodnot  $\sigma_{ij}$ , ty mohou být dodány měřicí laboratoří nebo agenturou. Pokud nejsou k dispozici je nutno nejistoty odhadnout. Způsoby odhadu jsou uvedeny v pracích (Reff a kol., 2007; Pollissar a kol., 1998).

## Metodika

### *Optimální počet faktorů*

Při volbě počtu faktorů v modelu lze postupovat několika způsoby. Počet faktorů lze zvolit na základě znalosti počtu zdrojů znečištění. Tato informace však nemusí vždy být k dispozici, navíc vzniklý model neumožňuje identifikovat případný další, neznámý nebo opomenutý zdroj. Pokud je k dispozici jiný statistický software, který má v nabídce metodu PCA, stanoví se počet postupů doporučenými pro PCA (např. vlastní čísla  $> 1$  nebo pomocí Cattelova indexového grafu úpatí vlastních čísel) (Meloun a kol., 2005; Hebák a kol., 2007). Samotný program EPA-PMF nenabízí konkrétní kvantitativní kritérium, využít lze hodnotu objektové funkce  $Q(\text{True})$ . Za vhodný model lze považovat takový, jehož objektová funkce je přibližně rovná počtu stupňů volnosti úlohy, nebo se přibližně rovná celkovému počtu bodů datového pole. Takový postup může být zavádějící, pokud nejsou správně stanoveny nejistoty měření. Vhodný model lze určit i zkoumáním změn objektové funkce při postupném přidávání faktorů. Pokud přidáním dalšího faktoru nedojde k významnému snížení hodnoty objektové funkce, není model s přidáním faktorem vhodný (Hopke, 2000). Jde tedy o podobný postup jako u PCA, jen místo vlastních čísel jsou posuzovány hodnoty objektové funkce. Postup je však početně náročný, protože je potřeba odhadovat stále složitější modely s vyšším počtem faktorů.

Modely s nižším než vhodným počtem faktorů vykazují velké množství reziduí mimo meze  $\pm 3$  (Norris a kol., 2009). S rostoucím počtem faktorů se počet vybočujících reziduí snižuje.

V tomto článku bylo použito jednak analýzy PCA vypočtené programem STATGRAPHICS Centurion XVI a dále metody zpětné validace modelu přímo v programu EPA-PMF. Program EPA-PMF nabízí metodu bootstrap, založenou na generování zadaného počtu náhodných, vzájemně se nepřekrývajících podvýběrů, které jsou základem odhadu dílčích modelů a zpětného přiřazování analytů

do odhadnutých faktorů. Grafický výstup v podobě krabicových diagramů jednotlivých analytů sestavených pro každý z faktorů umožňuje definovat, zda model správně zařazuje analyty do vypočtených faktorů. Nevhodný model se v grafu projeví širokými krabičkami a velkým počtem odlehlých hodnot. Ukázka nevhodného modelu byla zpracována pro data kovů z Litoměřic (obr. 1).

### Nastavení programu

K definování vah hodnot  $x_{ij}$  a nejistot měření  $s_{ij}$  vstupních proměnných se v programu EPA-PMF používá poměr signálu a šumu  $S/N$  (*signal-to-noise ratio*):

$$(S/N)_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - s_{ij})^2}{\sum_{i=1}^n s_{ij}}}.$$

Analyty s poměrem  $S/N < 0,2$  jsou označeny jako „Bad“ a do další analýzy nejsou zahrnuty. Jedná se nejčastěji o analyty s výraznými extrémy a/nebo s velkým množstvím hodnot pod detekčními limity. Analyty s poměrem  $0,2 < S/N < 2$  se označují jako „Weak“, v analýze zůstávají, ale jejich nejistota je stanovena na trojnásobek původní nejistoty. Jako „Weak“ se doporučuje označit i takové analyty, které ve zkušebním modelu vykazují rezidua mimo meze  $\pm 3$ . Akceptovatelné analyty se označují jako „Strong“ a platí pro ně nerovnost  $S/N > 2$  (Paatero a Hopke, 2003; Norris a kol., 2009).

Vliv odlehlých hodnot na kvalitu výsledného modelu lze vedle analýzy reziduí zjistit i porovnáním objektové funkce modelu  $Q(\text{True})$  s objektovou funkcí robustního odhadu  $Q(\text{robust})$ . Pokud je rozdíl mezi oběma funkcemi do 1 %, neovlivňují extrémy kvalitu odhadu. Je-li rozdíl do 10 %, je vliv extrémů slabý. V případě většího rozdílu je nutno přehodnotit váhy přiřazené analytům změnou jejich kvalifikace na „Weak“ nebo jejich vyřazením kvalifikací „Bad“. Autoři programu dále doporučují přidat do úlohy ještě dodatečnou nejistotu v rozmezí 5–25 % (Norris a kol., 2009).

Po vypočtení základního modelu s vhodným počtem faktorů je v některých případech nutno aplikovat ještě rotaci faktorů nástrojem Fpeak, která zajistí interpretovatelnost výsledků. Pro kvantifikaci podílu zdrojů na celkovém znečištění se jako „Total Variable“ nastavuje proměnná popisující celkovou zátěž, tzv. referenční polutant (Norris a kol., 2009). Při analýze kovů jde o koncentrace prachu  $PM_{10}$ ,  $PM_{2,5}$  apod. Elementární uhlík (EC) by mohl být vhodný pro kvantifikaci podílu zdrojů emitujících polycyklické uhlovodíky (PAH).

### Data

K analýze byla využita měření ZÚ se sídlem Ústí nad Labem a ČHmÚ. První soubor obsahuje hodnoty koncentrací kovů (Cd, Pb, As, Cr, Ni, Cu, Mn, Be, Fe, V) a prachu  $PM_{10}$  z měřicí stanice ZÚ v Litoměřicích. Interval měření je čtrnáct dní s počátkem 7.1.2007 a koncem 26.12.2010. Soubor obsahuje 11 řad analytů po 104 měřeních. Kromě koncentrací vyžaduje program EPA-PMF i soubor s detekčními limity a nejistotami měření vyjádřenými pomocí relativních směrodatných odchylek (RSD), ty jsou uvedeny v tabulce 1. Druhé měření je realizováno na stanici ČHmÚ v Praze-Libuši a obsahuje hodnoty PAH (BbF, BaA, Flu, Pyr, Chry, I123cdP, DBahA, BghiP, Fen, A, BaP, BkF), kovů (As, Ni, Cu, Cd, Pb, Mn), plyných polutantů ( $SO_2$ , NO,  $NO_2$ , CO,  $O_3$ ,  $SNO_3$ ,  $NO_x$ ) a prachu  $PM_{10}$  a  $PM_{2,5}$  z let 2006–2010. Interval měření kovů je ob jeden den, organických sloučenin ob dva dny a zbylé polutanty spolu s prachem byly měřeny každý den. Soubor z Prahy-Libuše je tak značně nesourodý, měření analytů ve většině případů nepochází ze stejného dne a obsahuje i mnoho chybějících hodnot z důvodu nefunkčnosti měřicí stanice. Soubor byl proto rozdělen na dvě skupiny: PAH (12 řad po 567 měřeních) a kovy+plynné polutanty (15 řad po 897 hodnotách) a tyto byly metodou PMF analyzovány odděleně. Pro analýzu byly vybrány pouze dny s kompletními daty, tj. dny, kdy byly známy koncentrace všech analytů. Detekční limity a nejistoty měření jsou uvedeny v tabulce 1.

Tabulka 1. Přehled analytů, detekčních limitů a nejistot měření

Místo: Litoměřice											
Analyt	Cd	Pb	As	Cr	Ni	Cu	Mn	Be	Fe	V	PM <sub>10</sub>
Jednotka	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	mg/m <sup>3</sup>
Det. limit	0,017	0,067	0,333	0,333	0,333	0,333	0,333	0,0167	3,333	0,333	0,005
Nejistota (v %)	12,5	12,5	12,5	12,5	12,5	12,5	12,5	12,5	12,5	12,5	9,5

Místo: Praha-Libuš Kovy + plynné polutanty															
Analyt	As	Ni	Cu	Cd	Pb	Mn	SO <sub>2</sub>	NO	NO <sub>2</sub>	CO	O <sub>3</sub>	NO <sub>x</sub>	PM <sub>10</sub>	PM <sub>2,5</sub>	SNO <sub>3</sub>
Jednotka	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>
Det. limit	0,054	0,126	0,099	0,003	0,045	0,031	2500	1500	2000	120000	2000	2000	2000	21000	7100
Nejistota (v %)	21	39	34	24	25	30	8	8	10	8	8	10	20	10	30

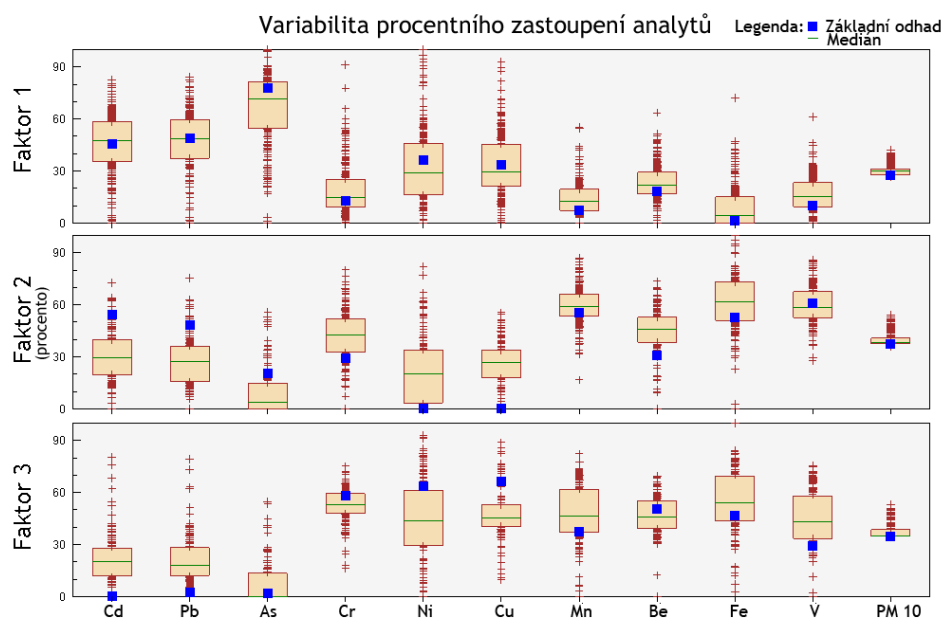
  

Místo: Praha-Libuš Organické sloučeniny												
Analyt	BbF	BaA	Flu	Pyr	Chry	I123cdP	DBahA	BghiP	Fen	A	BaP	BkF
Jednotka	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>	ng/m <sup>3</sup>
Det. limit	0,09	0,07	0,14	0,11	0,1	0,11	0,11	0,11	0,22	0,09	0,1	0,1
Nejistota (v %)	41,3	19,7	30,1	31,8	19,8	16,4	23,4	19,8	26,2	27,3	17,4	20,7

## Výsledky a diskuze

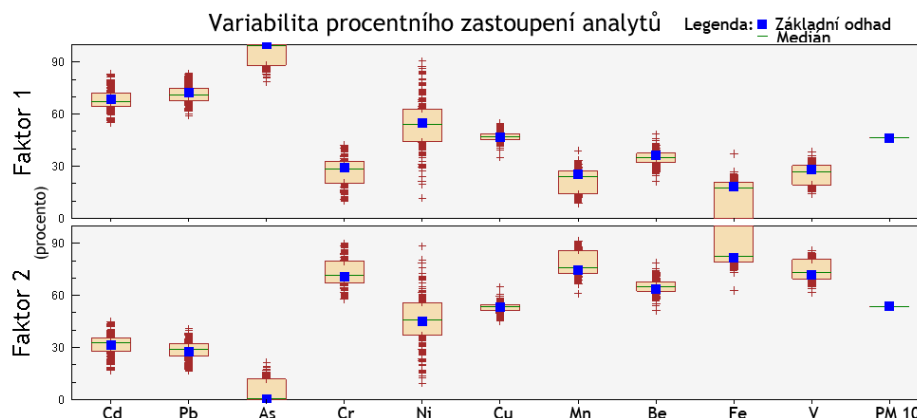
### Litoměřice

Počet faktorů byl metodou PCA stanoven na tři (vysvětleno 86,5 % variability; modelem se dvěma faktory vysvětleno 72,4 % variability). Odlišný počet faktorů přinesl postup PMF založený na porovnání výsledků zpětné validace. Byly porovnány dva modely, se dvěma resp. třemi faktory. Při zpětné validaci modelu se třemi faktory se ukázalo, že přidání třetího faktoru přináší nejednoznačné přiřazení všech kovů k jednotlivým faktorům (obr. 1).



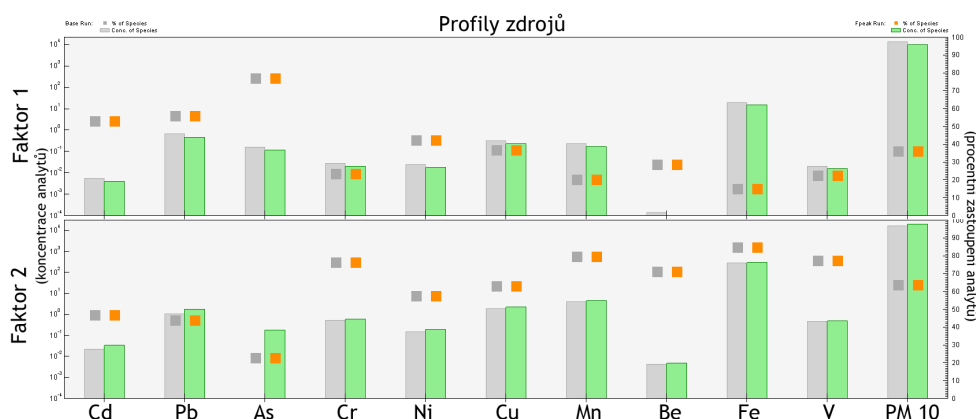
Obr. 1 Výsledky zpětné validace modelu se třemi faktory metodou bootstrap – 200 podvýběrů (ukázka nevhodného počtu faktorů)

Model se dvěma faktory se jeví jako vhodnější, pouze Ni nelze jednoznačně přiřadit ani do jednoho faktoru. Ani u jednoho kovu nepřekračují rezidua modelu se dvěma faktory hodnotu  $\pm 3$ . Determinační index  $R^2$  pro pozorované a modelem odhadnuté koncentrace je velice nízký u Cr (8 %) a Ni (15 %), u ostatních kovů je vyšší jak 60 %. Pomocí modelu se dvěma faktory lze tedy dobře odhadovat koncentrace všech kovů s výjimkou Cr a Ni (obr. 2). U obou kovů jsou problémy s hodnotami slepých pokusů a výsledky jsou silně ovlivněny zaokrouhlováním na celá čísla, čímž byla snížena nejistota měření. Z porovnání hodnot objektových funkcí  $Q(\text{True}) = 431$  a  $Q(\text{Robust}) = 431$  plyne, že model se dvěma faktory je robustní a není ovlivněn odlehlými hodnotami.



**Obr. 2** Výsledky zpětné validace modelu se dvěma faktory metodou bootstrap – 200 podvýběrů (ukázka vhodného počtu faktorů)

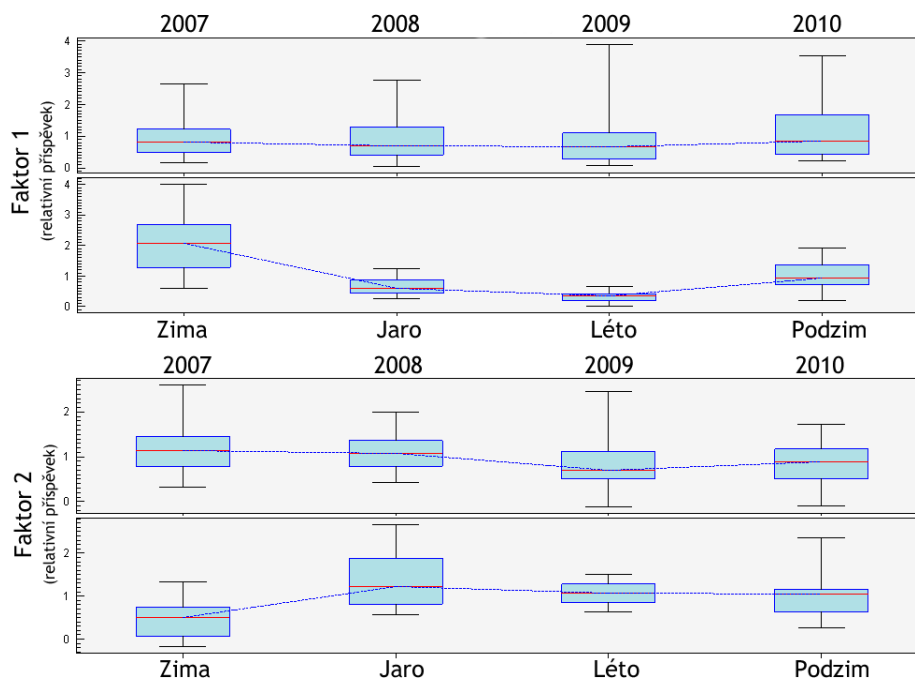
První zdroj emituje Cd (s relativním podílem 53 %), As (56 %) a Pb (77 %). Profil je uveden na obrázku 3. Je nejvyšší v zimě, nejnižší v létě (obr. 4). Zdrojem je spalování uhlí. Příspěvek zdroje mírně vzrostl v roce 2010. Druhý zdroj emituje Cu (s relativním podílem 63 %), Mn (80 %), Be (71 %), Fe (85 %) a V (77 %). Příspěvek zdroje je nejnižší v zimních měsících, vysoké příspěvky jsou od jara do podzimu. Mírně nižší byl v letech 2009 a 2010 (obr. 4). Interpretace zdroje není zcela jasná. Může se jednat o resuspendovaný prach a dopravu spolu s neznámým zdrojem.



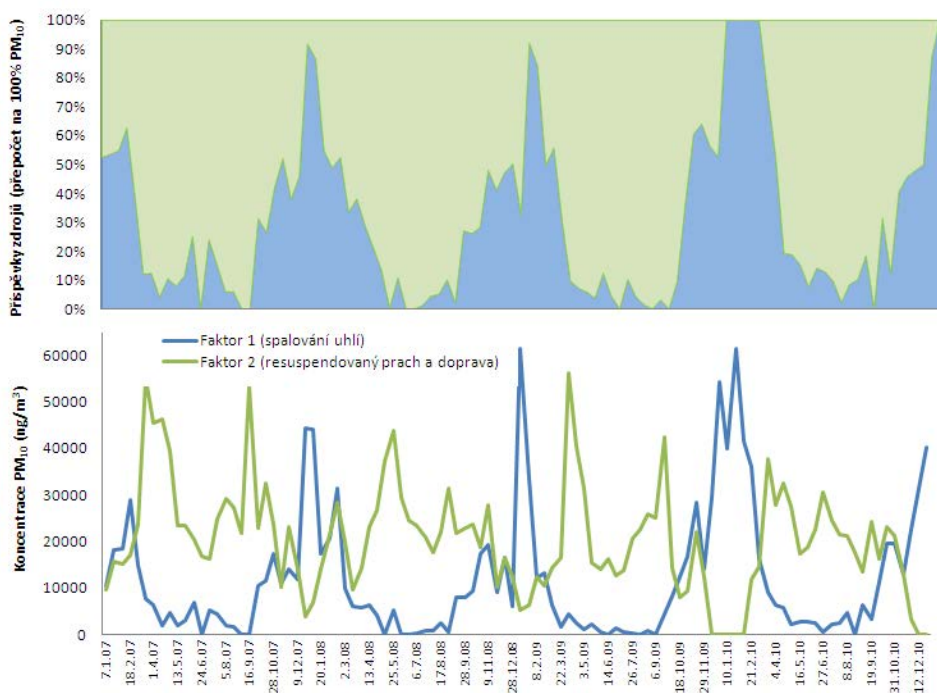
**Obr. 3** Profily zdrojů (faktor 1 - spalování uhlí, faktor 2 - resuspendovaný prach a doprava) před rotací faktorů (šedé sloupce a body) a po rotaci faktorů nástrojem Fpeak (zelené sloupce a žluté body). Profil je uveden v absolutních jednotkách (sloupce – logaritmické měřítko nalevo) a v relativních jednotkách (body – měřítko v procentech napravo).

Dominantní podíl na celkovém znečištění v místě měření má druhý zdroj (průměrný příspěvek za celé období je 67 %), který kulminuje od března do listopadu/prosince a v těchto měsících se příspěvek pohybuje mezi 80–100 %. Vývoj příspěvku v rámci jednotlivých let je velmi podobný a vykazuje sezónnost. Zejména v měsíci lednu a v roce 2010 i v únoru a březnu převládá znečištění

ze spalování uhlí, které se v těchto měsících na znečištění podílelo z 90–100 % (průměrný podíl za celé období je 33 %). Rozdíly příspěvku mezi jednotlivými roky jsou zřejmě způsobeny venkovní teplotou a počtem velmi chladných dnů v daných letech (obr. 5).



Obr. 4 Vývoj relativního příspěvku zdrojů (faktor 1 - spalování uhlí, faktor 2 - resuspendovaný prach a doprava) mezi roky 2007 a 2010 a v průběhu roku.



Obr. 5 Relativní (pře počet na 100 % emitovaného PM<sub>10</sub>) a absolutní příspěvek zdrojů (koncentrace emitovaného PM<sub>10</sub>).

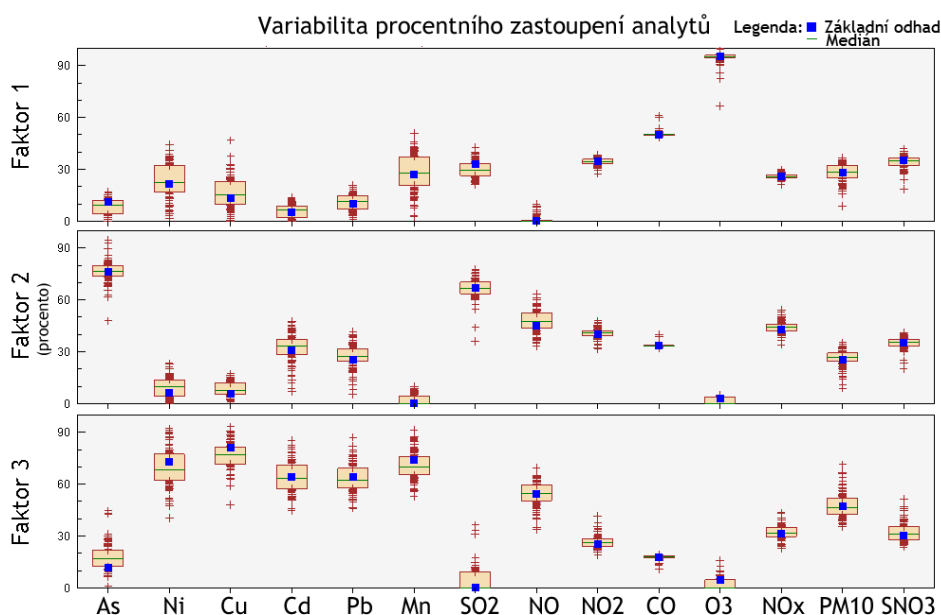
### Praha-Libuš (kovy)

Metodou PCA byl navržen model se třemi faktory (vysvětleno 74,2 % variability). Stejný počet zdrojů potvrdil i postup zpětnou validací modelů (obr. 6). Identifikován byl zdroj s vysokou koncentrací ozónu O<sub>3</sub> a CO, což lze přiřadit fotochemickým procesům v atmosféře – reakcím uhlovodíků a oxidů dusíku s emitovanými uhlovodíky z automobilové dopravy vlivem UV záření (obr. 9). Proto

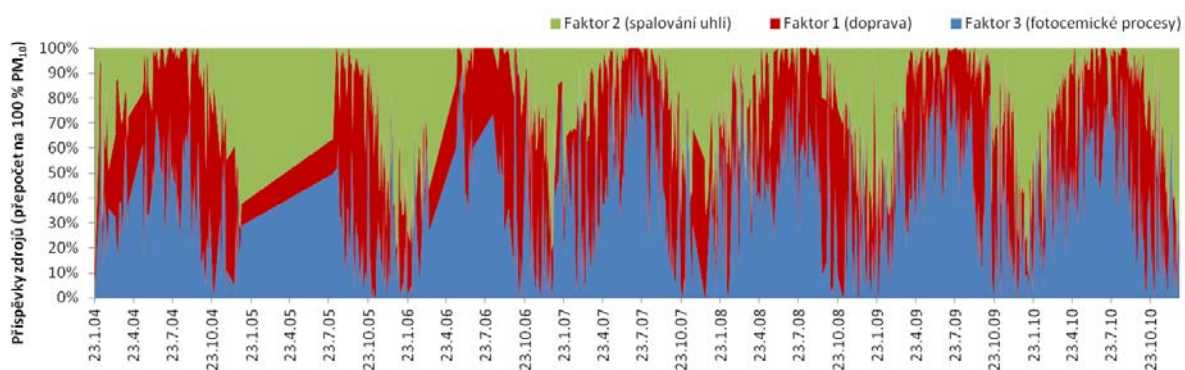


má zdroj největší příspěvek v letních měsících (obr. 10). Druhý zdroj emituje As a SO<sub>2</sub> a jde zcela jistě o spalování uhlí v zimě (obr. 9 a 10). Třetí zdroj emituje zbylé kovy (Ni, Cu, Cd, Pb, Mn) a jde také o dopravu s konstantním vývojem s mírným podzimním zvýšením koncentrace (obr. 9 a 10). První a třetí zdroj lze považovat za totožné. Jejich rozdělení je způsobeno odlišnými podmínkami, zejména intenzitou slunečního záření, které způsobuje rozdílnou reaktivnost emitovaných polutantů během roku. Plynné polutanty NO<sub>2</sub>, NO<sub>x</sub> a SNO<sub>3</sub> nelze jednoznačně přiřadit do žádného zdroje. Podle modelu se na každém zdroji podílejí z 1/3. NO se dle modelu podílí 1/2 na druhém a třetím zdroji. CO nejvíce koreluje s vývojem ozonu a byl zařazen do prvního zdroje.

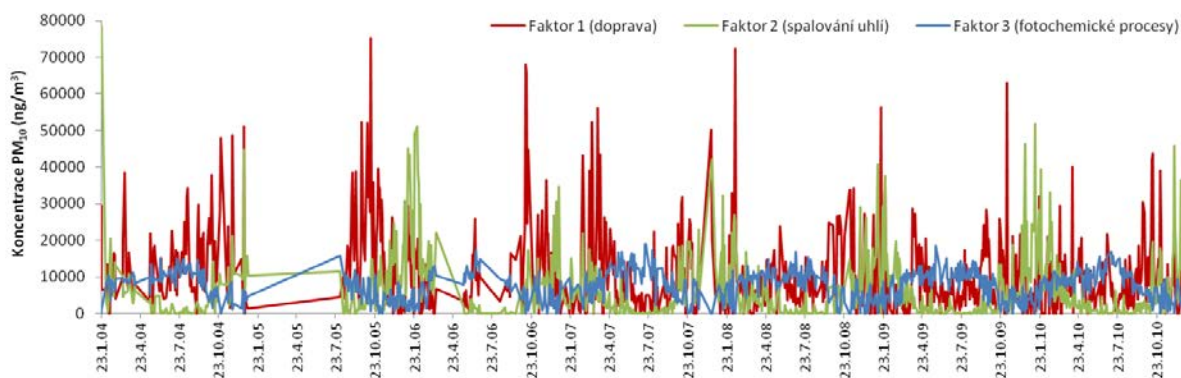
Vypočtené příspěvky zdrojů na znečištění vedou k následujícím závěrům: od března do října se na celkovém znečištění podílejí fotochemické procesy emisí z dopravy (průměrná koncentrace 48 % s maximy v létě až 90 %), spalování uhlí (průměrná koncentrace 40 %), průměrný podíl emisí kovů z dopravy je 12 %. Až čtyřnásobně vyšší celkové znečištění lze pozorovat od listopadu do února, kdy dominuje spalování uhlí s průměrným podílem 45 % s maximy až 80 % (obr. 7 a 8).



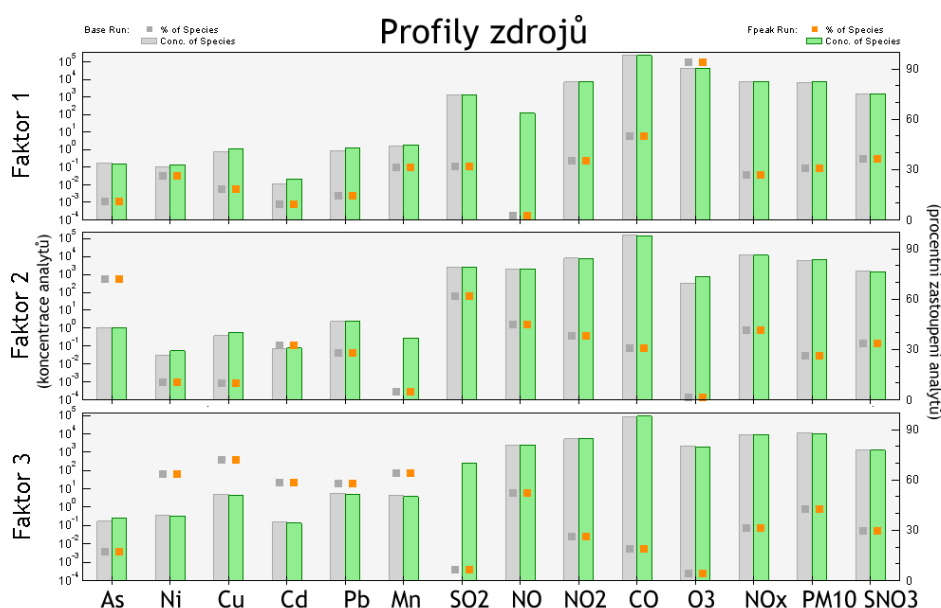
Obr. 6 Výsledky zpětné validace modelu se třemi faktory metodou bootstrap – 200 podvýběrů



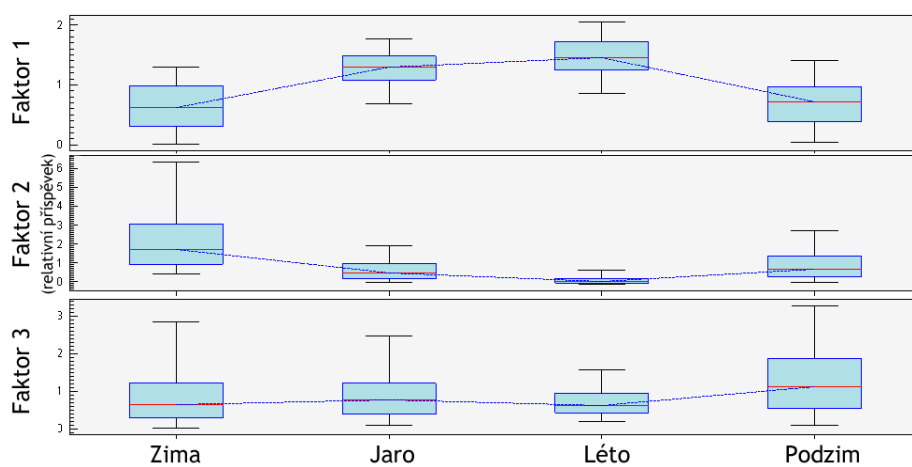
Obr. 7 Relativní příspěvek zdrojů (pře počet na 100 % emitovaného PM<sub>10</sub>)



**Obr. 8** Absolutní příspěvek zdrojů (koncentrace emitovaného PM<sub>10</sub>)



**Obr. 9** Profily zdrojů (faktor 1 - fotochemické procesy emisí z dopravy, faktor 2 - spalování uhlí, faktor 3 - automobilová doprava) před rotací faktorů (šedé sloupce a body) a po rotaci faktorů nástrojem Fpeak (zelené sloupce a žluté body). Profil je uveden v absolutních jednotkách (sloupce – logaritmické měřítko nalevo) a v relativních jednotkách (body – měřítko v procentech napravo).



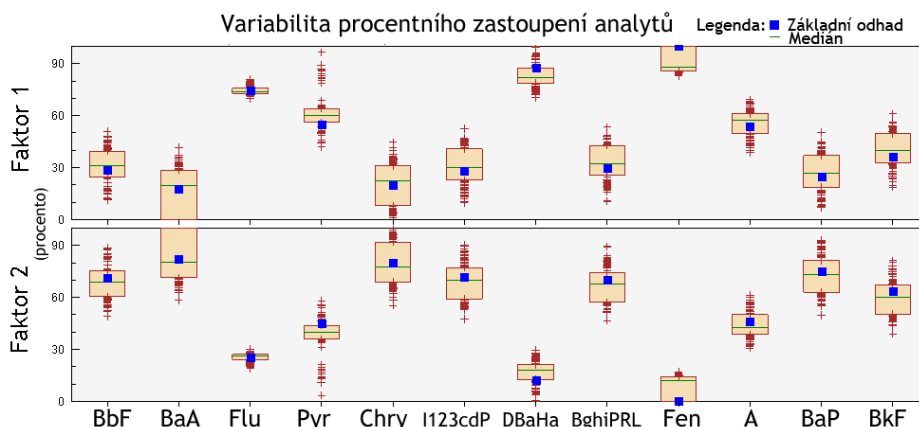
**Obr. 10** Vývoj relativního příspěvku zdrojů (faktor 1 - fotochemické procesy emisí z dopravy, faktor 2 - spalování uhlí, faktor 3 - automobilová doprava) v průběhu roku.

### Praha-Libuš (PAH)

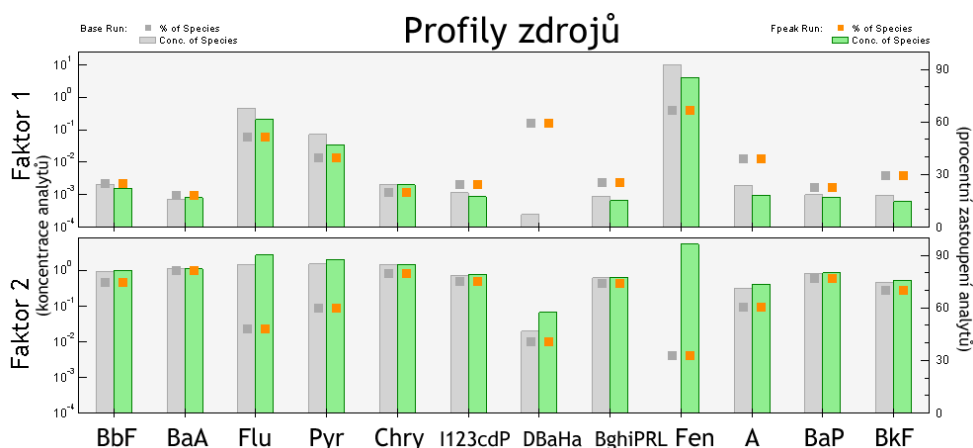
Při validaci modelu se dvěma faktory jsou všechny polutanty dobře přiřazovány do odpovídajících profilů, takže model se dvěma profily se zdá být vhodným pro analyzovaná data (obr. 11). Metodou PCA byl zvolen jen jeden faktor (vysvětleno 90 % variability). Z porovnání hodnot objektových funkcí  $Q(\text{True}) = 1438$  a  $Q(\text{Robust}) = 1438$  plyne, že model se dvěma faktory je robustní a není ovlivněn odlehkými hodnotami, které se objevují v koncentracích všech polutantů.

Polutanty lze pomocí PMF přiřadit ke dvěma samostatným zdrojům. Oba zdroje vykazují stejné změny koncentrací během roku: nízké koncentrace v létě a vysoké v zimě (obr. 13). Pyren nelze jednoznačně přiřadit do žádného zdroje, podle modelu se na jeho emisích podílí každý zdroj zhruba z 1/2. Zdroj 1 emituje fluoranten a fenantren, přiřazení DBaH a nelze považovat za směrodatné vzhledem k jeho velice nízkým koncentracím. Zdroj 1 nelze jednoznačně definovat. Možná se jedná o spalování plynu v lokálních plynových kotlích, které jsou v provozu převážně v zimních měsících. Pro potvrzení domněnky však chybí další podpurné analyty. Zbylé PAH byly zařazeny do druhého zdroje a zřejmě se jedná o spalování tuhých paliv (dřevo, uhlí) v rodinných domech (obr. 12 a 13).

U souboru PAH nelze stanovit relativní ani absolutní podíly profilů na celkovém znečištění, protože není znám referenční polutant potřebný pro výpočet. Takovým by mohl být organický uhlík (OC) nebo elementární uhlík (EC) použité mj. v Shrivastava a kol. (2007), Schauer a kol. (2000) nebo Zheng a kol. (2002).

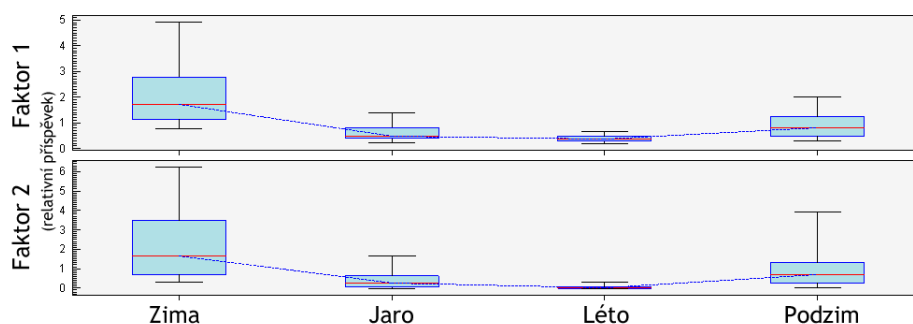


Obr. 11 Výsledky zpětné validace modelu se dvěma faktory metodou bootstrap – 200 podvýběrů



Obr. 12 Profily zdrojů (faktor 1, faktor 2 - spalování tuhých paliv) před rotací faktorů (šedé sloupce a body) a po rotaci faktorů nástrojem Fpeak (zelené sloupce a žluté body). Profil je uveden

v absolutních jednotkách (sloupce – logaritmické měřítko nalevo) a v relativních jednotkách (body – měřítko v procentech napravo).



**Obr. 13** Vývoj relativního příspěvku zdrojů (faktor 1, faktor 2 - spalování tuhých paliv) v průběhu roku

## Závěr

Metoda PMF umožňuje úspěšně a srozumitelně identifikovat a kvantifikovat zdroje znečištění z imisních dat v místě měření, pokud je k dispozici dostatečný počet analytů a dostatečné množství měření ideálně za delší časové období, ve kterém se projeví výkyvy zdrojů emisí. Metoda vyžaduje i znalost nebo odhad nejistot měření a je robustní při výpočtech zešikmených dat. K volbě počtu faktorů ve výsledném modelu lze využít postup zpětné validace, kdy vhodný model správně přiřazuje analyty k odhadnutým faktorům a rezidua nejsou příliš vysoká. Pokud je známa koncentrace referenčního polutantu, lze pomocí PMF kvantifikovat příspěvek zdroje (relativní i absolutní) v každém okamžiku měření. Změny příspěvku v čase spolu s odhadnutým profilem pomáhají k identifikaci zdroje znečištění.

Ve zkoumaných oblastech (Litoměřice a Praha-Libuš) se jako hlavní zdroje znečištění podařilo identifikovat spalování tuhých paliv v zimních měsících a dopravu v létě.

## Poděkování

Studie vznikla díky vstřícnosti Zdravotního ústavu v Ústí nad Labem a pobočky Českého hydrometeorologického ústavu v Ústí nad Labem. Obě instituce poskytly data pro aplikaci PMF.

## Seznam literatury

- HEBÁK, P., HUSTOPECKÝ, J., PECÁKOVÁ, I., PRŮŠA, M., ŘEZANKOVÁ, H., SVOBODOVÁ, A. a P. VLACH. (2007) *Vícerozměrné statistické metody* [3]. Praha: Informatorium, 2007. 271 s. ISBN 978-80-7333-001-9.
- HOPKE, K. P. (2000) *A Guide to Positive Matrix Factorization*. [Online] 8. 2. 2000. [cit. 2012-02-20]. Dostupné z: <http://www.epa.gov/ttnamti1/files/ambient/pm25/workshop/laymen.pdf>.
- LENÍČEK, J., KOVÁČ, M., SEKYRA, M., SYNEK, V., PLACHÁ, H. a D. RICHTEROVÁ. (2011) Polycyklické aromatické uhlovodíky v městském ovzduší jako markery imisního zatížení ze stacionárních a mobilních zdrojů. *Ochrana ovzduší*. 2011, č. 1, s. 7–14.
- LINGWALL, J. W. A W. F. CHRISTENSEN. (2007) Pollution source apportionment using a priori information and positive matrix factorization. *Chemometrics and Intelligent Laboratory Systems*. 2007, vol. 87, s. 281–294.
- MELOUN, M., MILITKÝ, J. a M. HILL. (2005) *Počítačová analýza vícerozměrných dat v příkladech*. Praha: Academia, 2005. 449 s. ISBN 80-200-1335-0.
- NORRIS, G., VEDENTHAM, R., WADE, K., BROWN, S., PROUTY, J., BAI, S., DeWINTER, J. a CH. FOLEY. (2009) *EPA Positive Matrix Factorisation (PMF) 4.0 Fundamentals and User Guide. Draft Version*. Washington: U.S. Environmental Protection Agency, 2009. 90 s.

- PAATERO, P. (1997) Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems*. 1997, vol. 37, issue 1, s. 23–35.
- PAATERO, P. a K. P. HOPKE. (2003) Discarding or downweighting high-noise. *Analytica Chimica Acta*. 2003, vol. 490, s. 277–289.
- PAATERO, P. a U. TAPPER. (1993) Analysis of different modes of factor analysis as least squares fit problems. *Chemometrics and Intelligent Laboratory Systems*. 1993, vol. 18, issue 2, s. 183–194.
- PAATERO, P. a U. TAPPER. (1994) Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*. 1994, vol. 5, no. 2, s. 111–126.
- POLLISSAR, A. V., PAATERO, P., HOPKE, P. K., MALM, W. C. a J. F. SISLER, (1998) Atmospheric aerosol over Alaska 2. Elemental composition and sources. *Journal of Geophysical Research*. 1998, vol. 103, no. d15.
- REFF, A., EBERLY, I. S. a V. P. BHAVE. (2007) Receptor modeling of ambient particulate matter data using positive matrix factorization: review of existing methods. *Journal of the Air & Waste Management Association*. 2007, vol. 57, s. 146–154.
- SCHAUER, J. J. a G.R. CASS. (2000) Source Apportionment of Wintertime Gas-Phase and Particle-Phase Air Pollutants Using Organic Compounds as Tracers. *Environmental Science & Technology*. 2000, vol. 34, no. 9, s. 1821–1832.
- SCHAUER, J. J., ROGGE, W. F., HILDEMAN, L. M., MAZUREK, M. A. a G. R. CASS. (1996) Source apportionment of airborne particulate matter using organic compounds as tracers. *Atmospheric Environment*. 1996, vol. 30, no. 22, s. 3837–3855.
- SHRIVASTAVA, M. K., SUBRAMANIAN, R., ROGGE, W. F. a A. L. ROBINSON, (2007) Sources of organic aerosol: Positive matrix factorization of molecular marker data and comparison of results from different source apportionment models. *Atmospheric Environment*. 2007, vol. 41, issue 40, s. 9353–9369.
- VELÍŠEK, J. (2007) Receptorové modelování. *Příloha časopisu Ochrana ovzduší*. 2007, leden.
- XIE, Y.-L., HOPKE, P. K., PAATERO, P., BARRIE, L. A. a S.-M. LI. (1999) Identification of Source Nature and Seasonal Variations of Arctic Aerosol By Positive Matrix Factorisation. *Journal of the Atmospheric Sciences*. 1999, vol. 56, s. 249–260.
- ZHENG, M., CASS, R. G., SCHAUER, J. J. a S. E. EDGERTON. (2002) Source Apportionment of PM<sub>2.5</sub> in the Southeastern United States Using Solvent-Extractable Organic Compounds as Tracers. *Environmental Science & Technology*. 2002, vol. 36, no. 11, s. 2361–2371.