

Příloha 2/B

Positive Matrix Factorization 1.1

1. EPA PMF 1.1

Positive Matrix Factorization (dále jen PMF) je nástroj sloužící ke statistickému zpracování naměřených dat. PMF představuje nový přístup k provádění faktorové analýzy. Patří do skupiny analýz, které eliminují duplicity a zhušťují informace obsažené v původních proměnných do menšího počtu vzájemně nekorelovaných proměnných. PMF je využitelný speciálně při delších imisních měřicích kampaních, kdy jsou znečišťující látky na síti monitorovacích stanic sledovány v průběhu např. několika let a tento rozsáhlý soubor dat je pro další zpracování potřeba nějakým způsobem vyhodnotit.

EPA PMF je v podstatě grafické rozhraní modelu Positive Matrix Factorization, pracující na základě multivariačních nástrojů implementovaných v programu ME-2. EPA PMF 1.1 řeší hlavní problémy receptorového modelování použitím omezení počtu dat, vážení (průměrů) a metody nejmenších čtverců.

Model obecně předpokládá p zdrojů nebo zdrojových typů působících na receptor (příjemce) a lineární kombinace působení těchto p zdrojů způsobuje na receptoru pozorované koncentrace různých znečišťujících látek. Matematicky vyjádřeno:

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

kde:

- x_{ij} je koncentrace znečišťující látky zjištěná u příjemce (a tedy známá)
- i představuje i-tý den
- j představuje j-tou znečišťující látku
- g_{ik} je příspěvek k-tého faktoru k receptoru pro i-tý den
- f_{kj} je příslušná frakce k-tého faktoru a j-té znečišťující látky a
- e_{ij} je zbytek (rozdíl) pro i-tý den a j-tou znečišťující látku

V EPA PMF je předpoklad, že pouze x_{ij} je známo, přičemž cílem analýzy je odhadnout příspěvek g_{ik} a frakci f_{kj} . Dalším předpokladem je, že příspěvky jednotlivých zdrojů jsou „nezáporné“. Dodatečně EPA PMF dovoluje uživateli říci, jak velká nejistota je přidělena každému x_{ij} . Údajům s velkými nejistotami není dovoleno ovlivnit odhad příspěvku k danému profilu tak výrazně, jako údajům s menší nejistotou.

Úkolem EPA PMF je tedy minimalizovat sumu mocnin

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

kde:

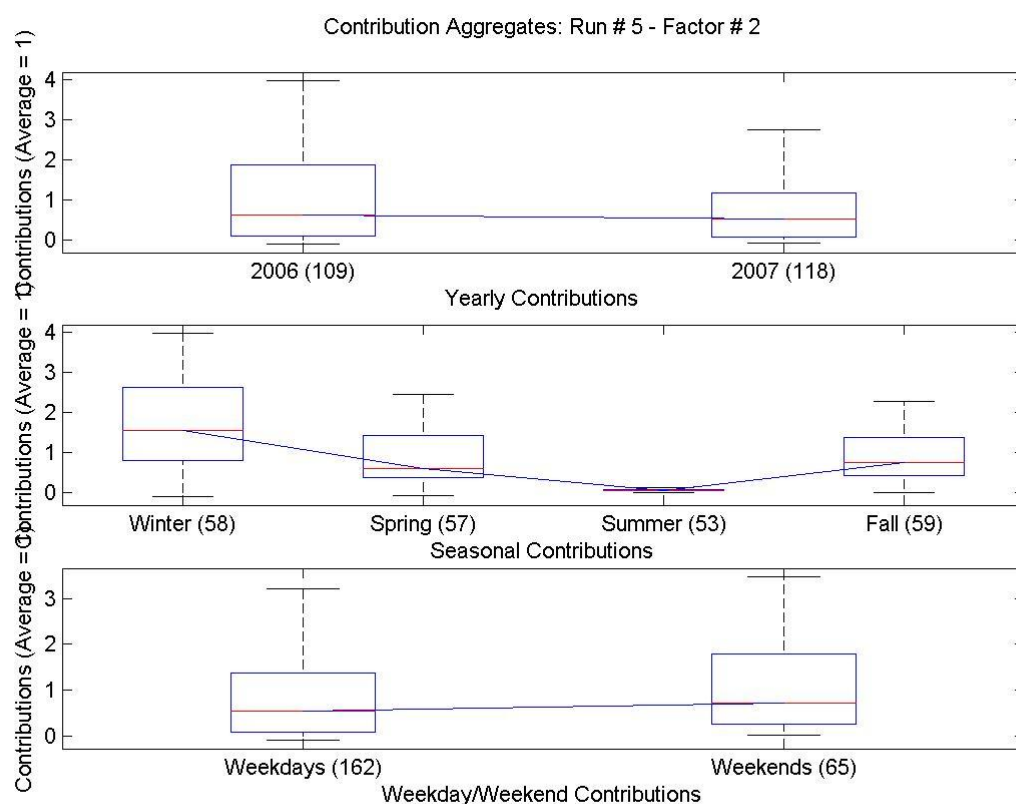
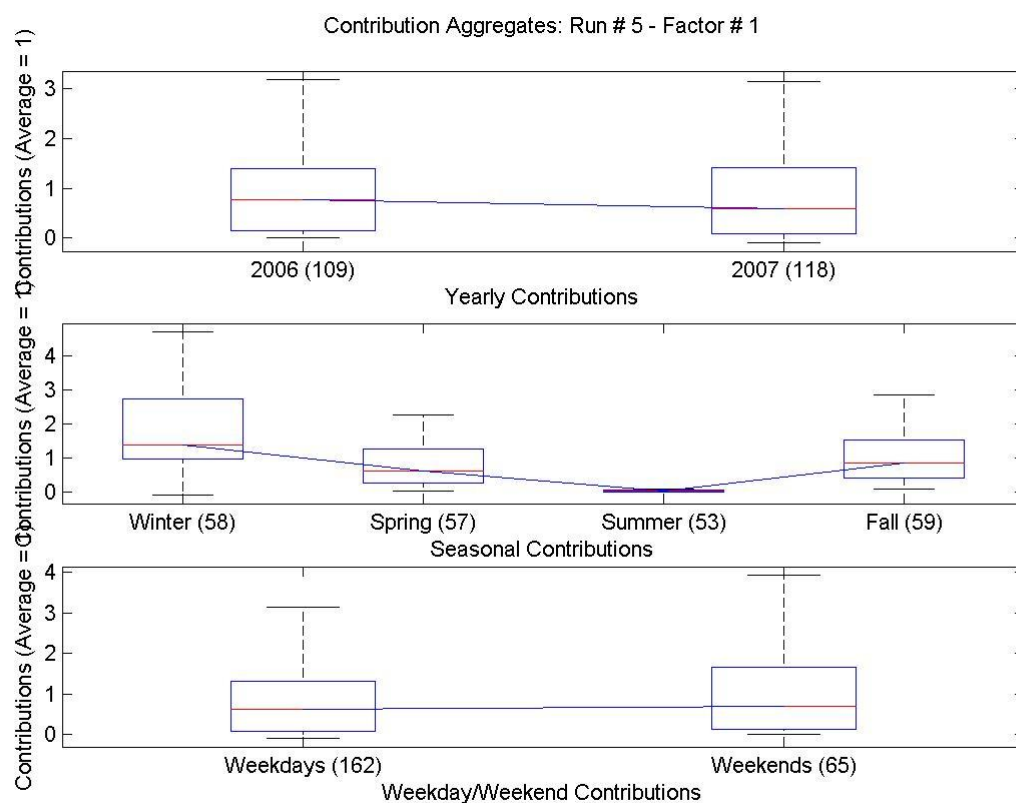
s_{ij} je nejistota příslušící j-té znečišťující látce pro i-tý den

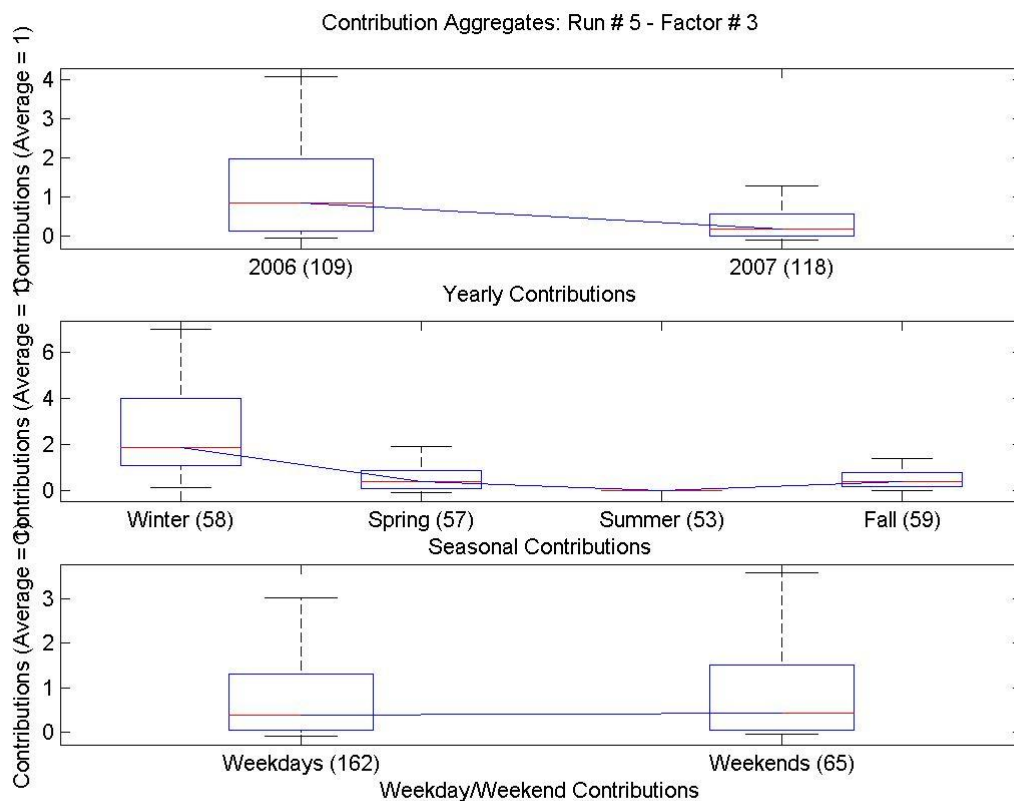
Positive Matrix Factorization může být použit k určení zdrojových profilů založených na imisních datech.

- PMF používá Metodu nejmenších čtverců vhodnou pro data s normálním rozdělením a Metodu maximální věrohodnosti odhadu pro data s logaritmicko normálním rozdělením
- PMF normalizuje datové body jejich analytickými nejistotami
- PMF omezuje faktorovou zátěž a faktorové skóre na nezáporné hodnoty a proto minimalizuje nejasnost způsobenou rotačními faktory. Toto je jeden z hlavních rozdílů mezi PMF a analýzou hlavních komponent (PCA)
- PMF vyjadřuje faktorovou zátěž v koncentračních jednotkách což dovoluje použít faktory přímo jako zdrojové profily (podpisy)
- PMF uvádí nejistoty pro faktorovou zátěž a faktorové skóre, které dělají zátěže a skóre jednodušší k použití v kvantitativních procedurách jako je Chemical Mass Balance

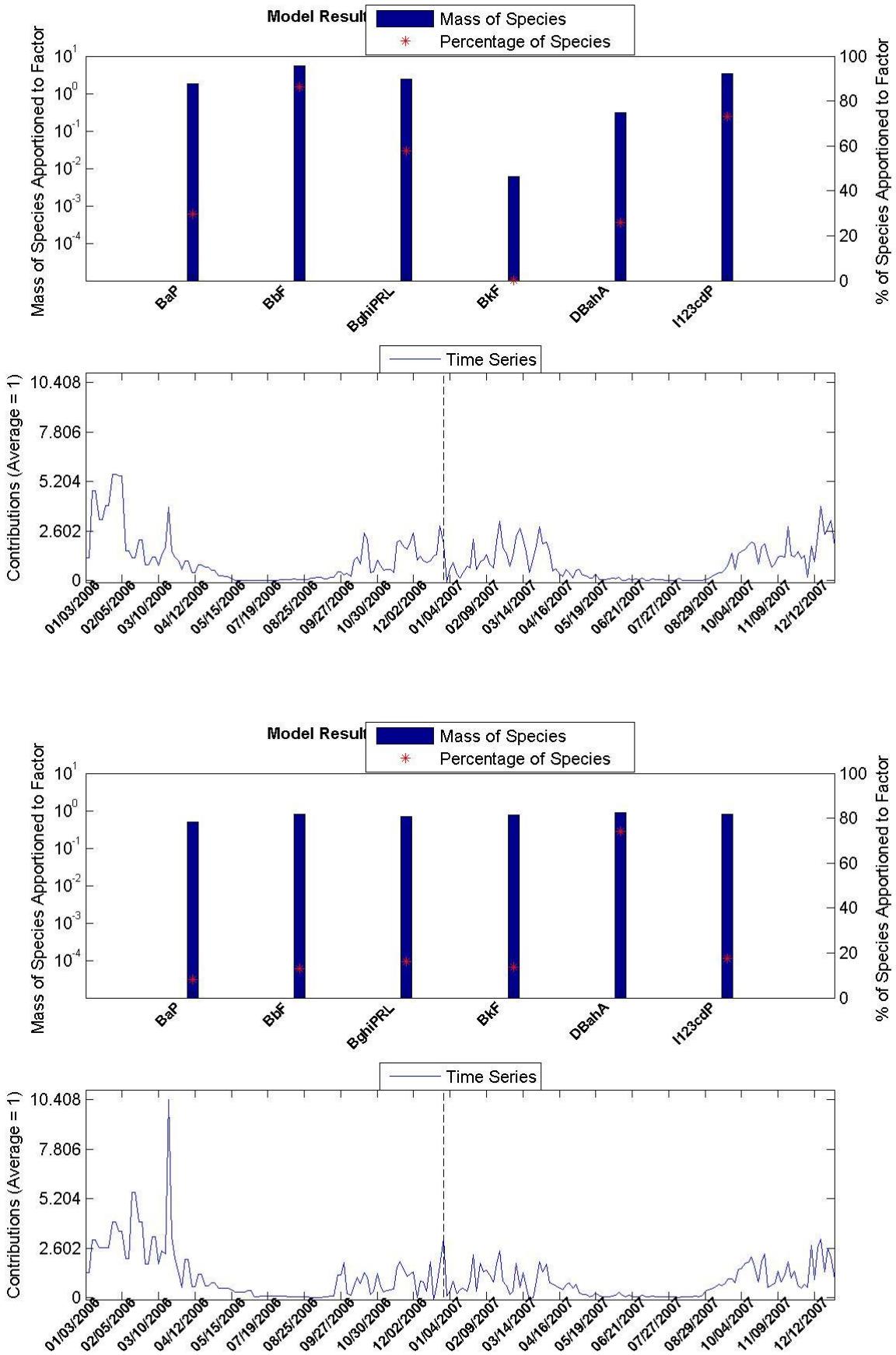
V následujícím textu pro názornost uvádíme aplikaci PMF na data z ISKO poskytnutá Českým hydrometeorologickým ústavem. K uvedeným výsledkům je však potřeba přistupovat **pouze** jako k rešeršní ukázce, získané faktory nemohou být a nebudou dále při modelování využity. Obdobnému vyhodnocení budou podrobena data získaná v rámci imisních měřících kampaní projektu na vybraných stanicích AIM:

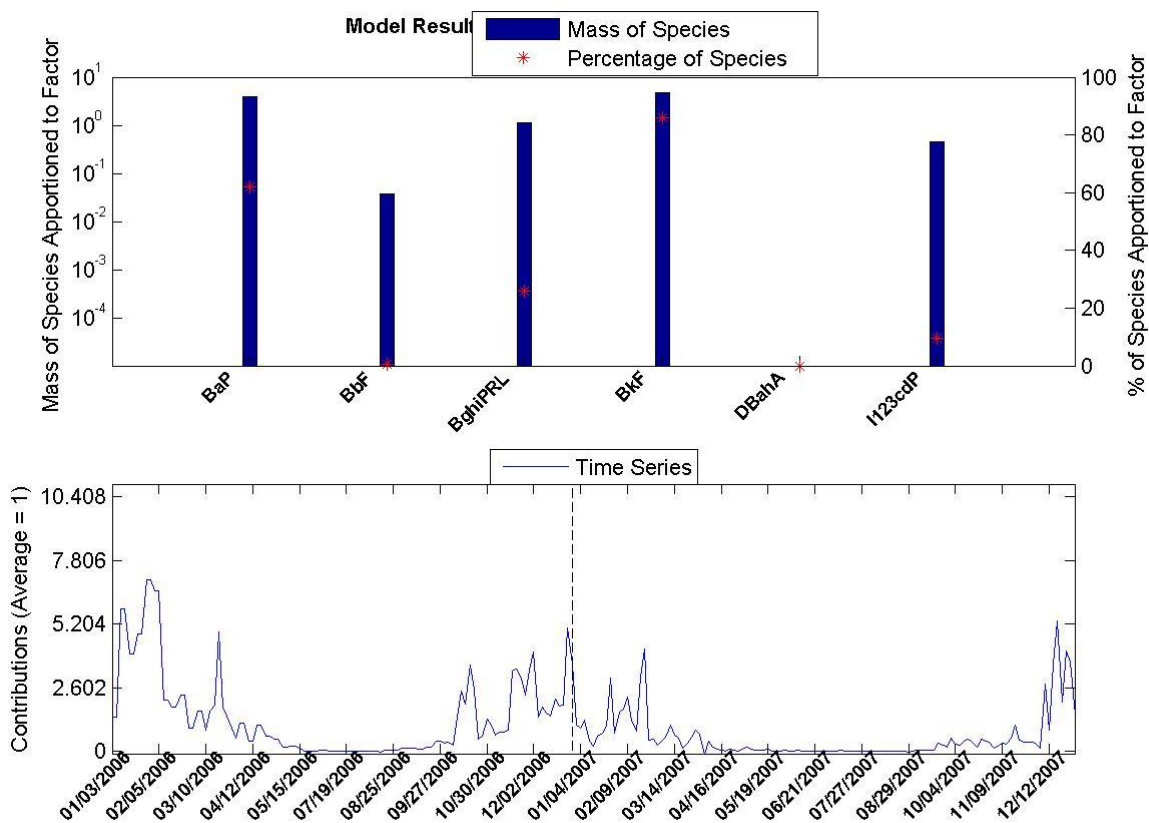
- Beroun
- Brandýs nad Labem
- Kladno - Švermov

1.1. Vyhodnocení – polycyklické aromatické uhlovodíky – AIM Kladno - Švermov



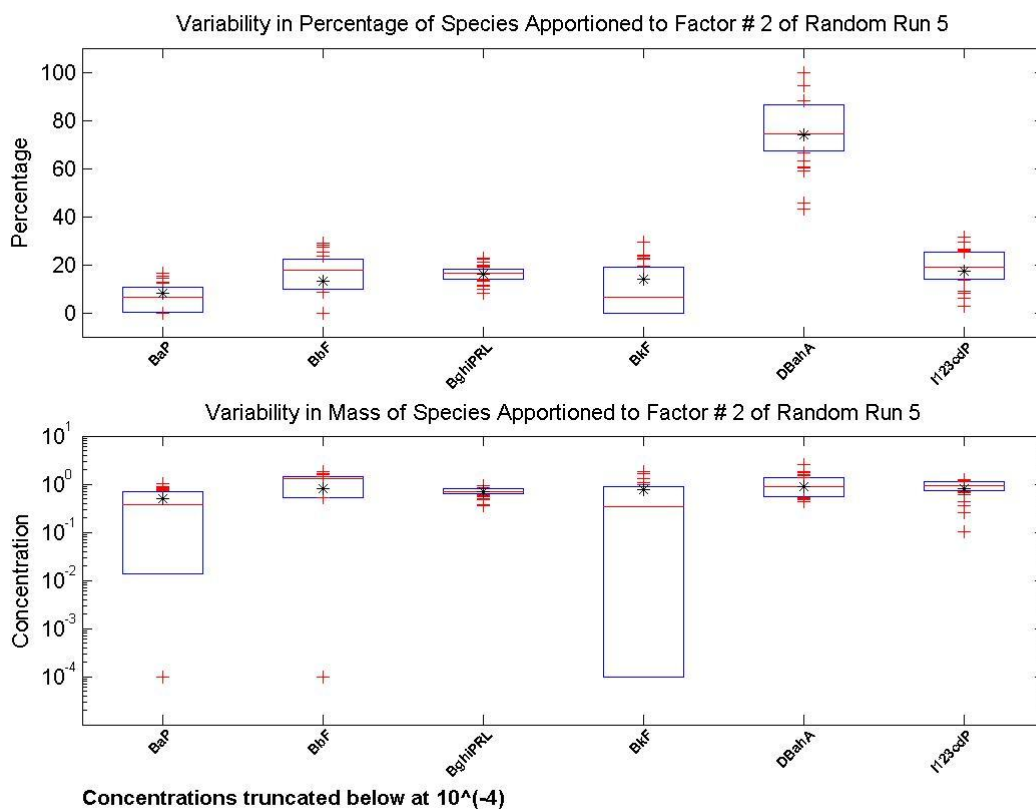
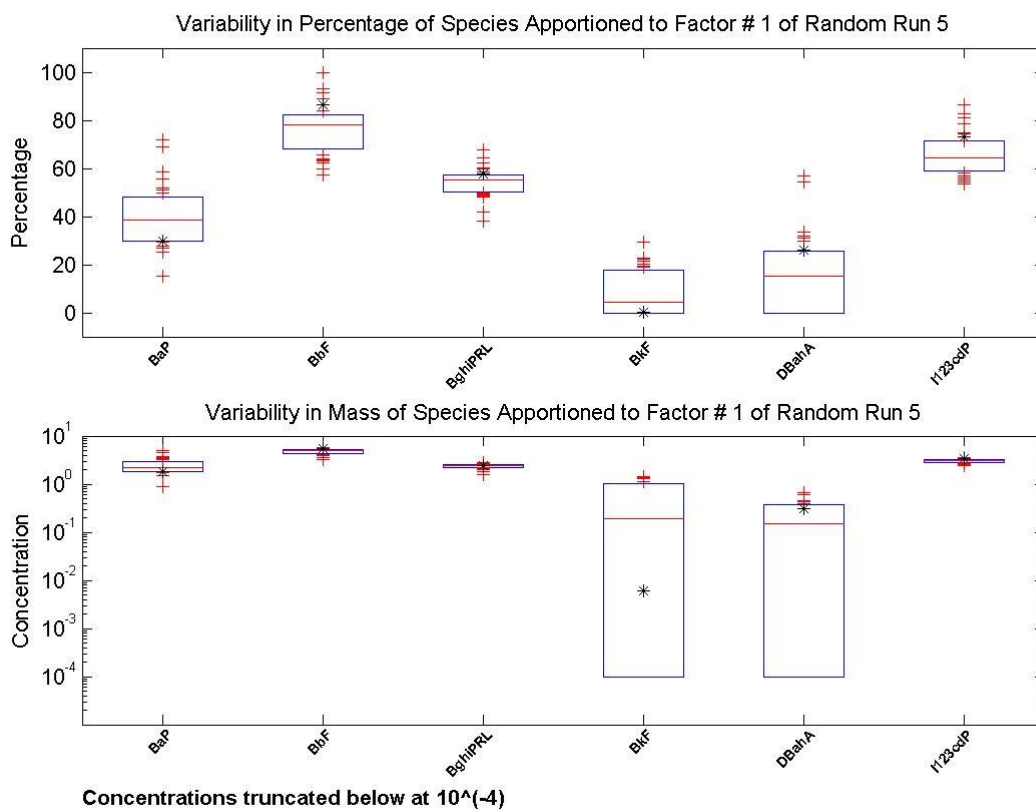
Předcházející typ grafu zobrazuje sumarizaci časové řady (viz. dále) pro daný faktor. Graf zobrazuje variabilitu příspěvku v roce, ročním období a při dělení pracovní den/víkend. Červená čára v diagramu představuje medián, krabicový graf má rozsah od 25 do 75 percentilu a černě označená oblast vymezuje 10 až 90 percentil. Extrémy (méně než 10 a více než 90 percentil) nejsou zobrazeny. Číslo v závorce u popisu osy x udává počet vzorků v každé kategorii.

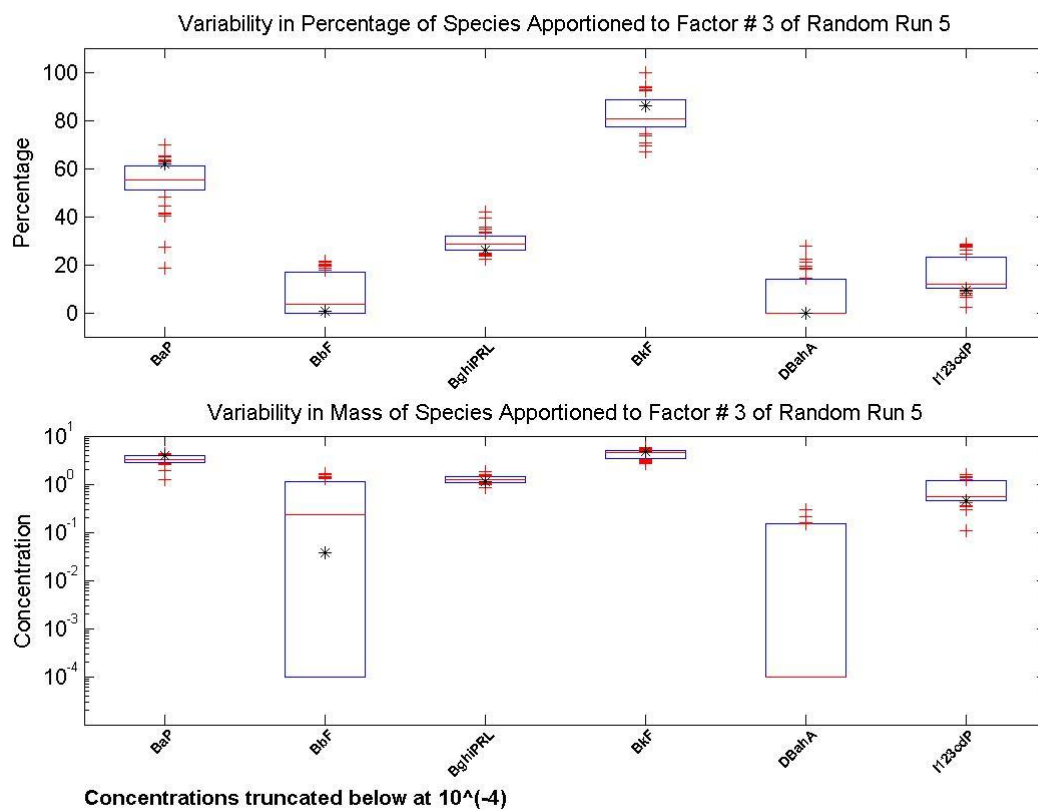




Tento typ grafu zobrazuje modelované faktory. Horní část zobrazuje samotný profil, spodní část pak časovou řadu. Profil je zobrazen za použití dvou měřítek. Sloupce zobrazují množství každé znečišťující látky „přidělené“ k faktoru, kde jednotky jsou stejné jako vyjádření koncentrace ve vstupních souborech ($\mu\text{g}/\text{m}^3$).

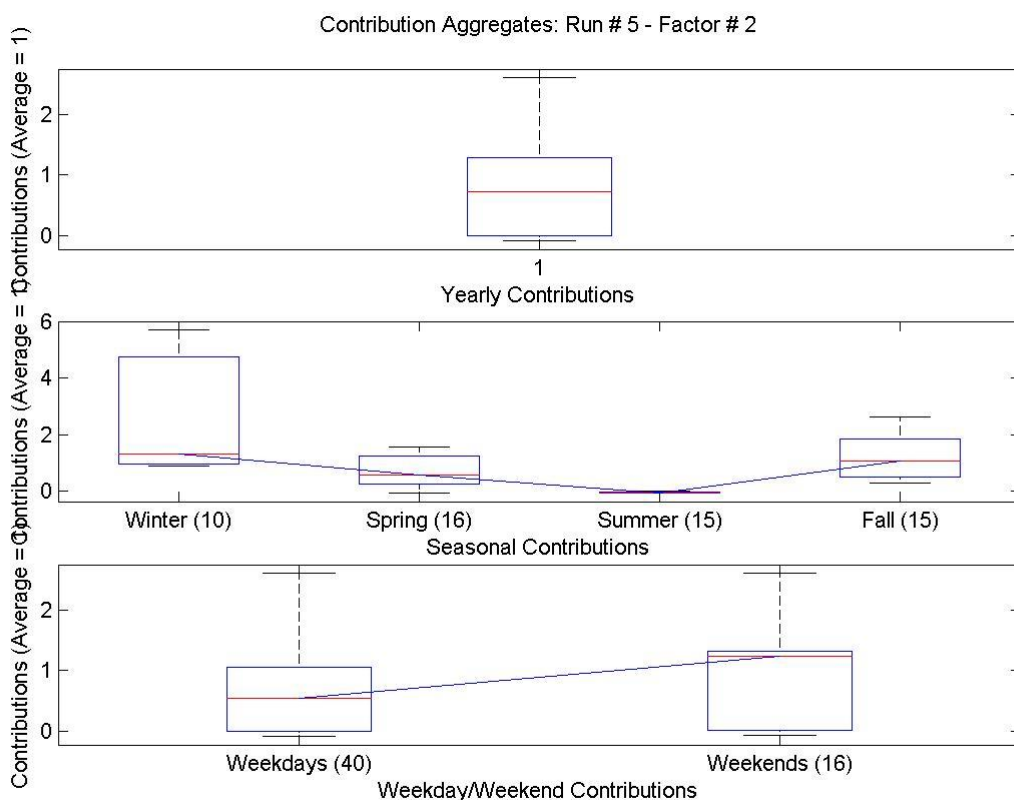
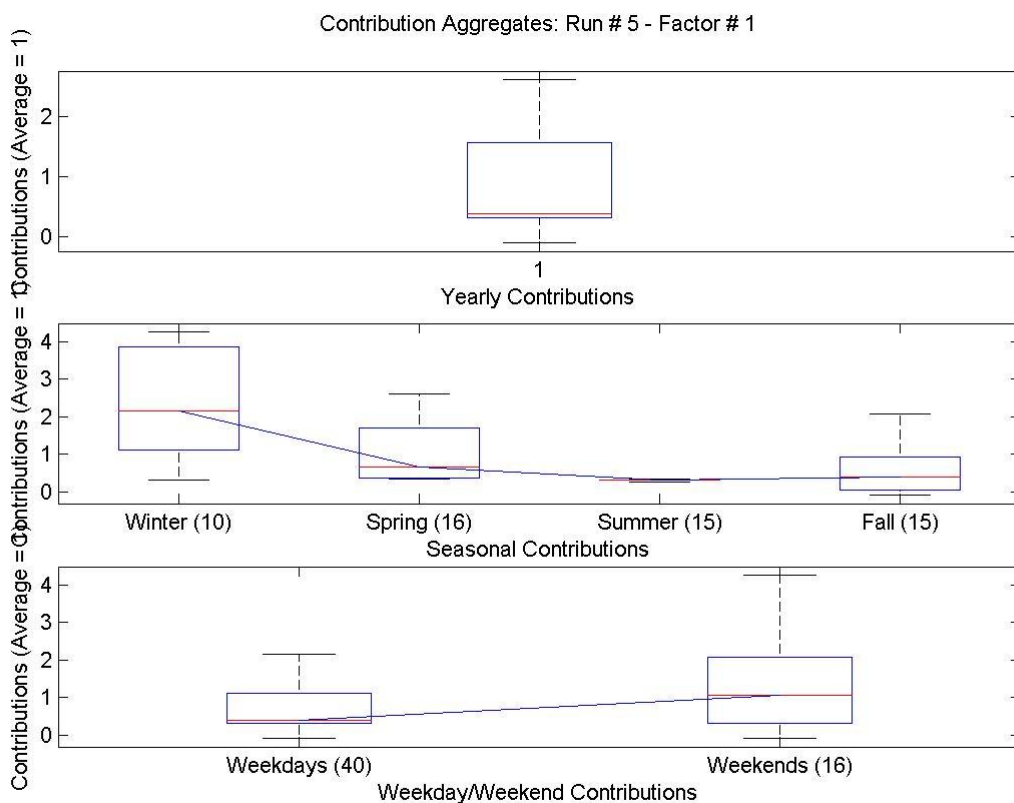
Na grafu je také pro každou znečišťující látku vyznačena pozice hvězdičky. Ta udává, kolik procent z každé znečišťující látky je přiděleno tomuto faktoru.

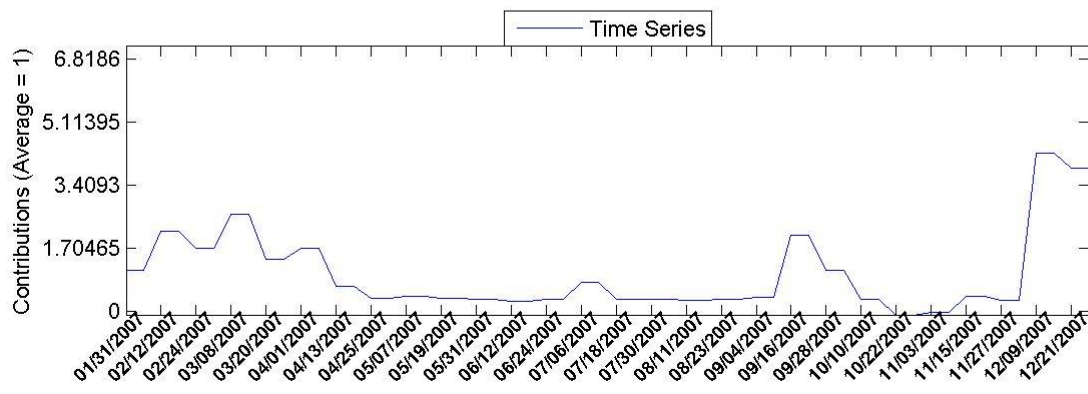
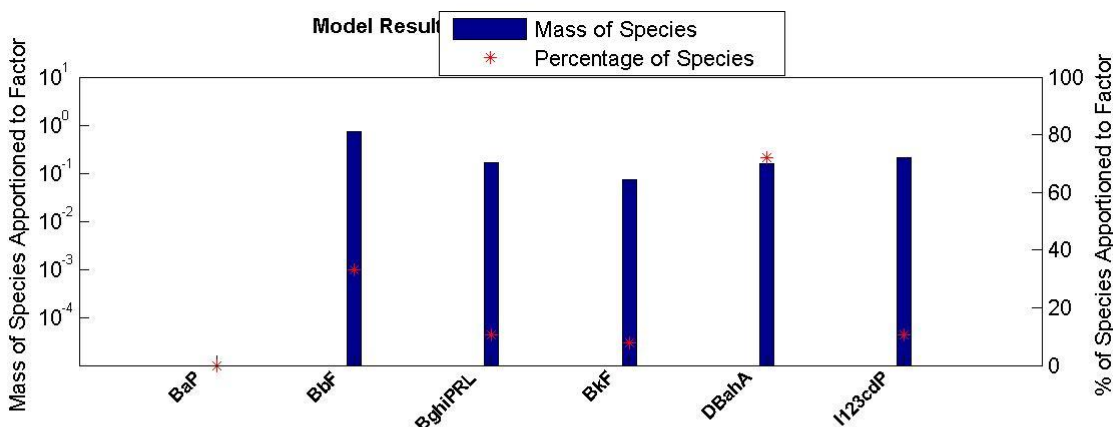
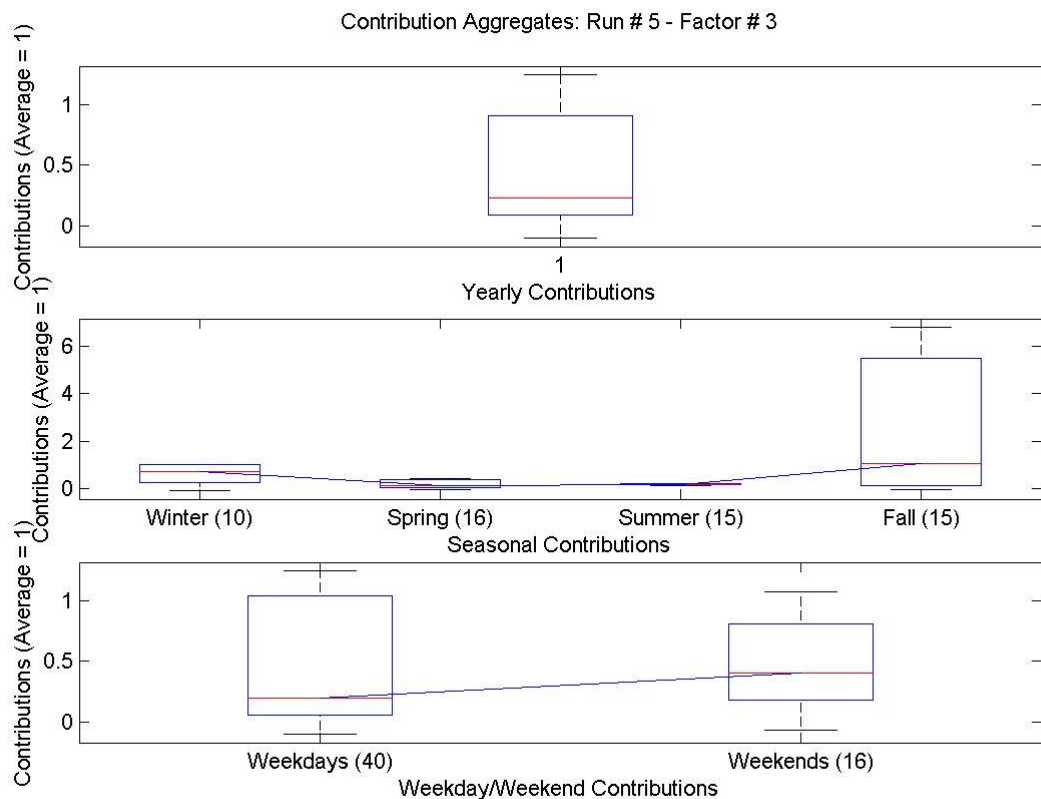


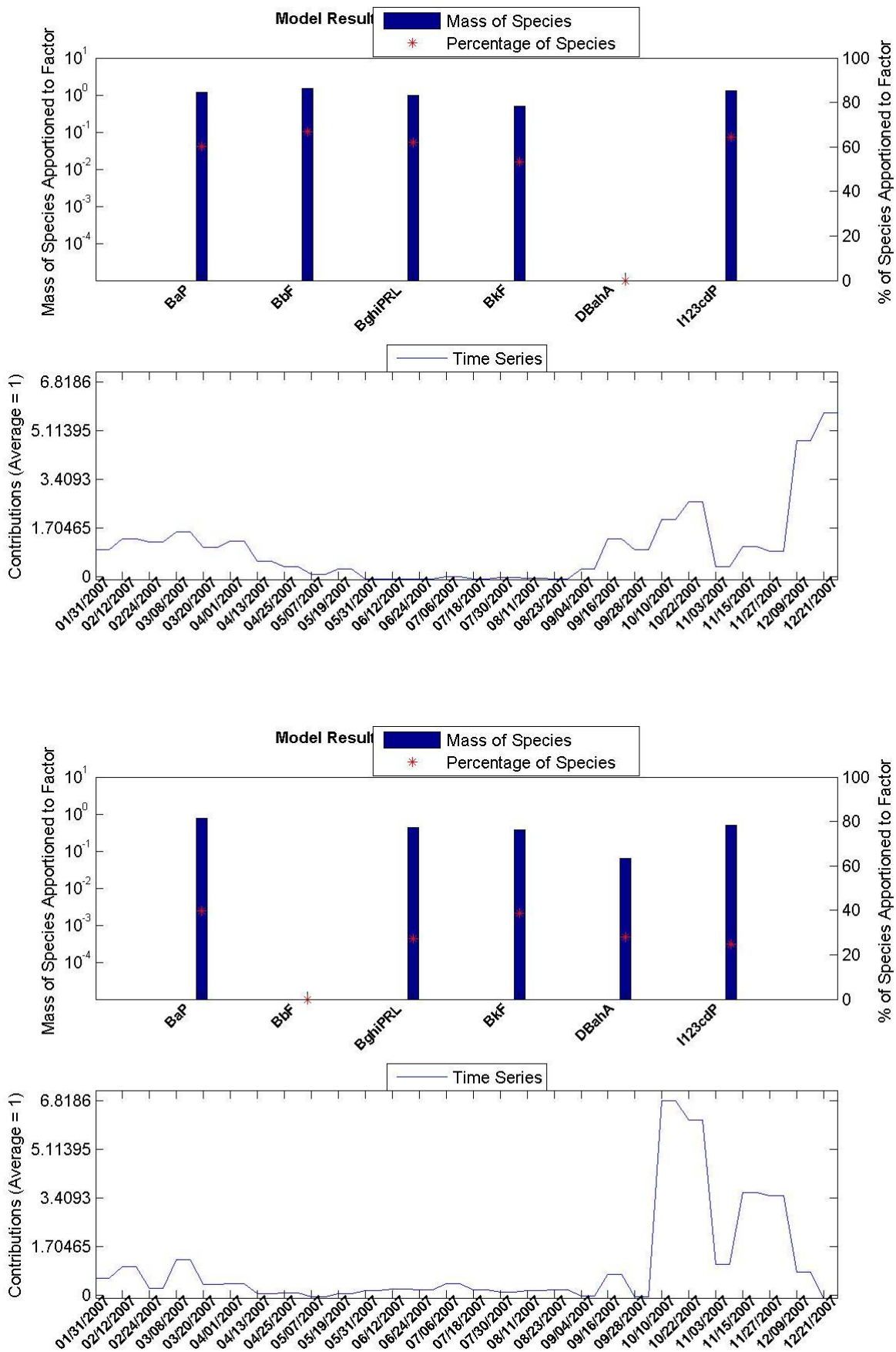


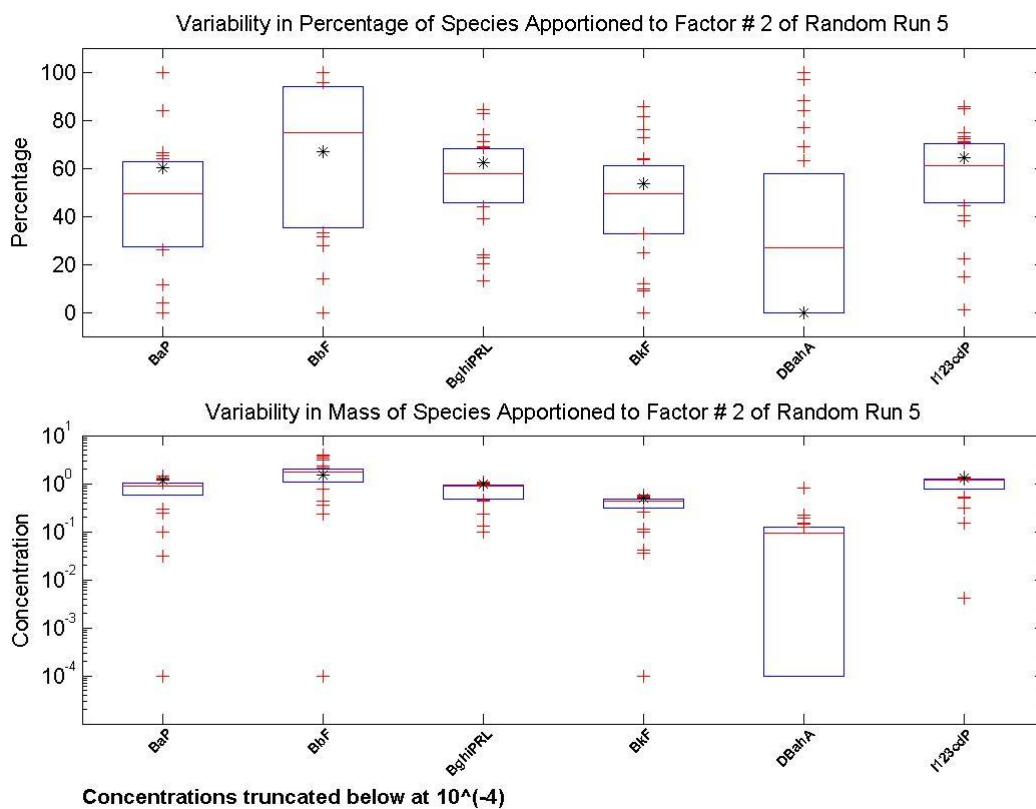
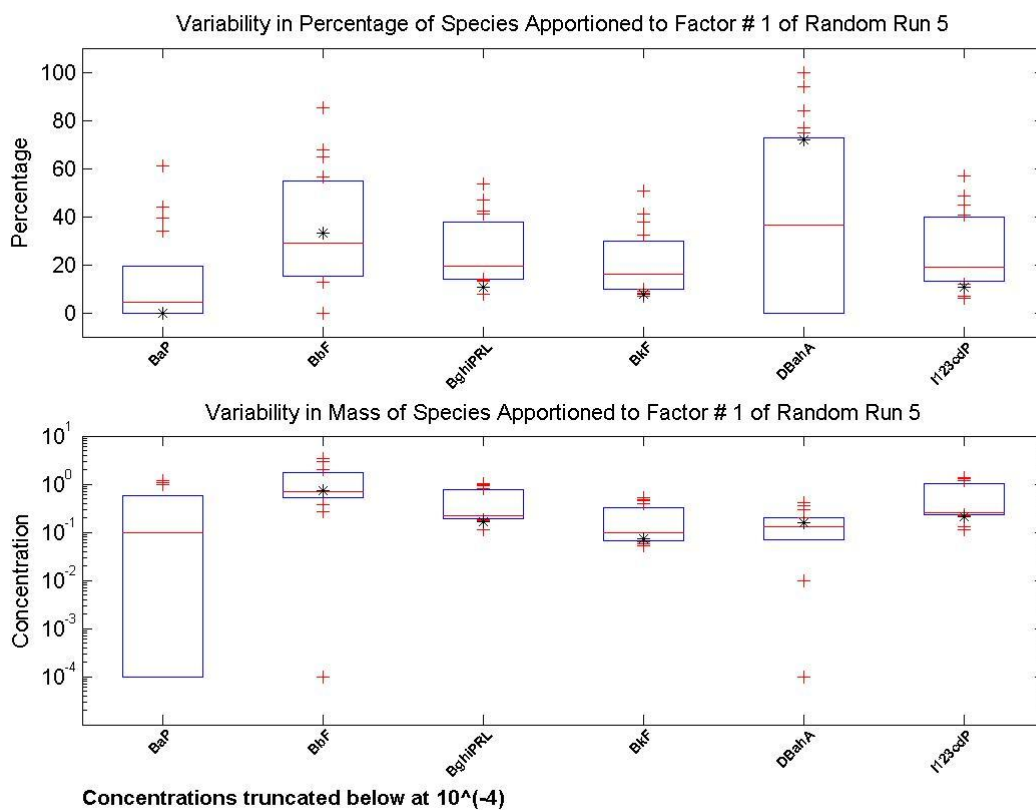
Horní panel obrázku uvádí nejistoty v procentech množství znečišťující látky. Dolní panel uvádí nejistoty v koncentračních jednotkách zobrazených prostřednictvím tzv. krabicových grafů. Tento graf uvádí, kde leží 50 % „bootstrap“ hodnot; čím je výška boxu menší, tím je výsledek konzistentnější.

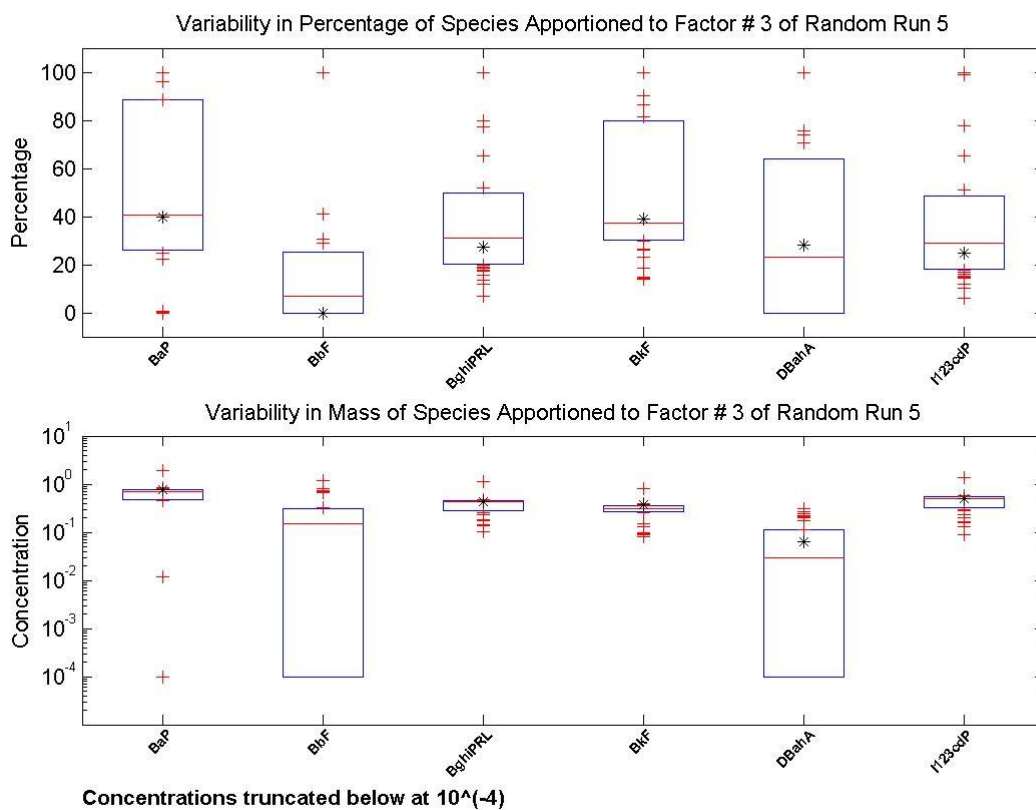
1.2. Vyhodnocení – polycyklické aromatické uhlovodíky – AIM Brandýs nad Labem



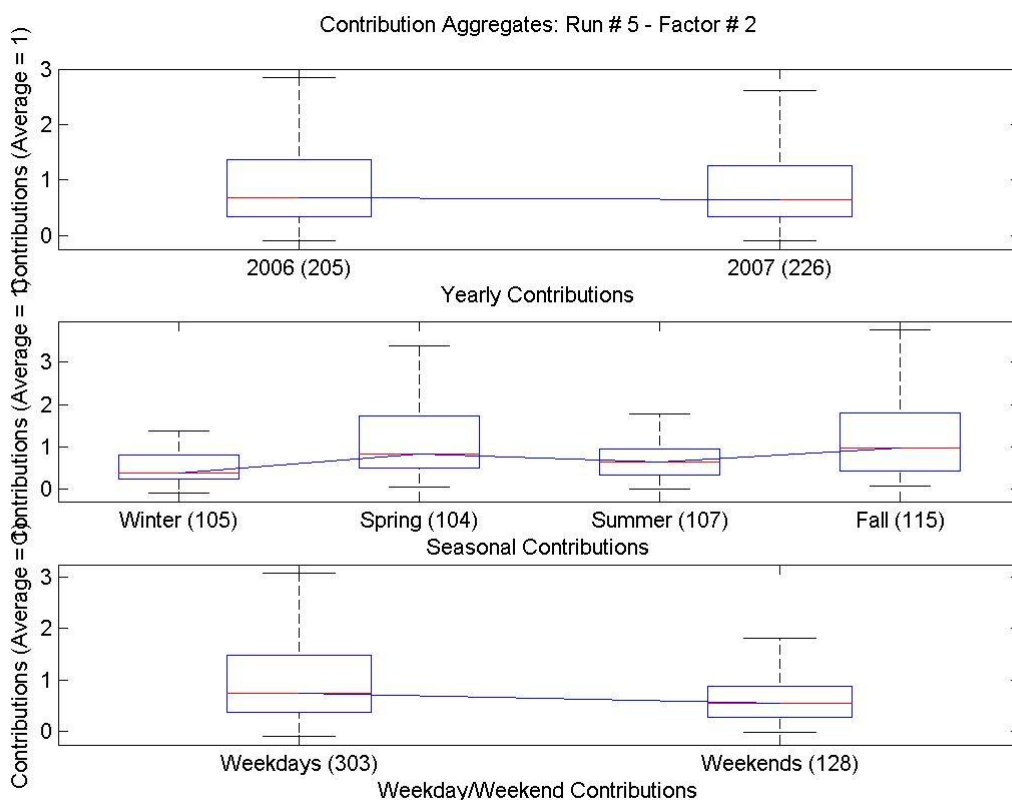
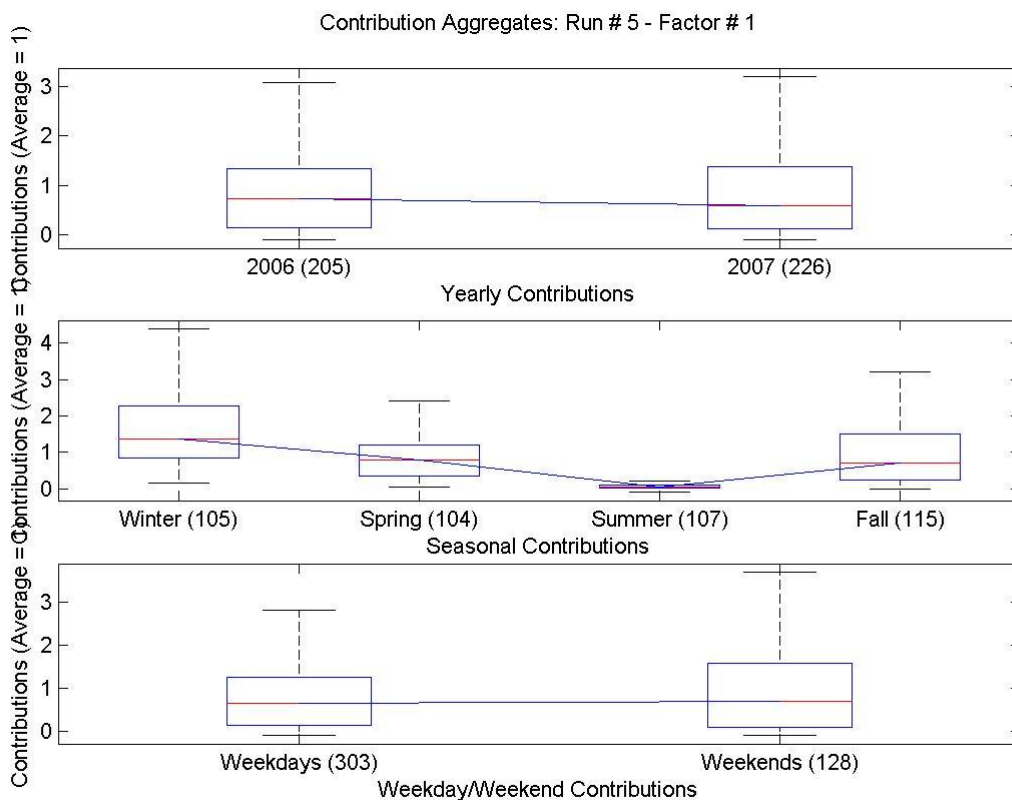


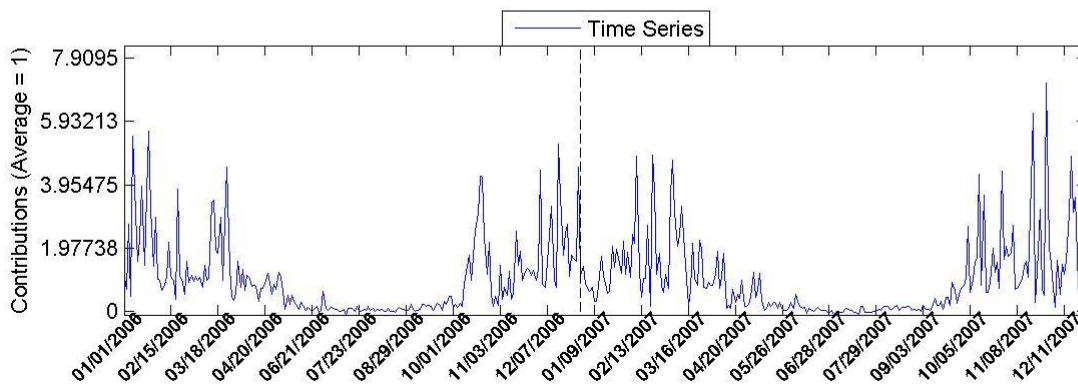
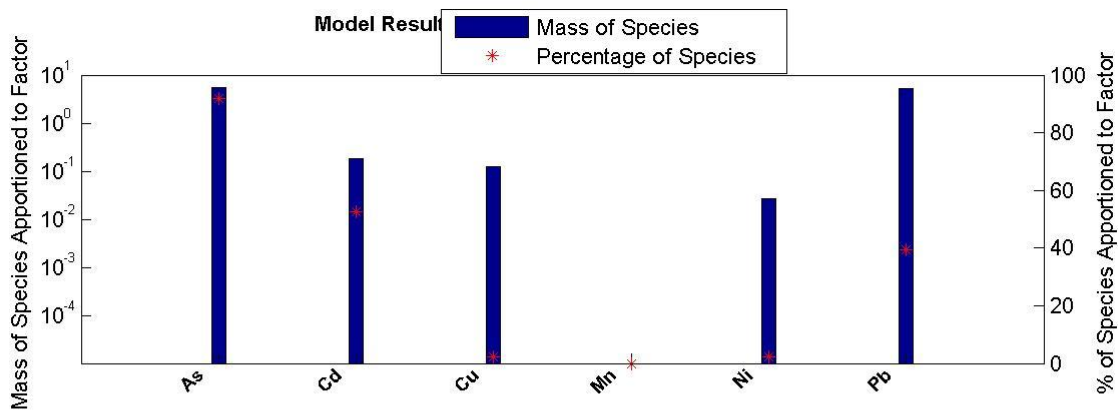
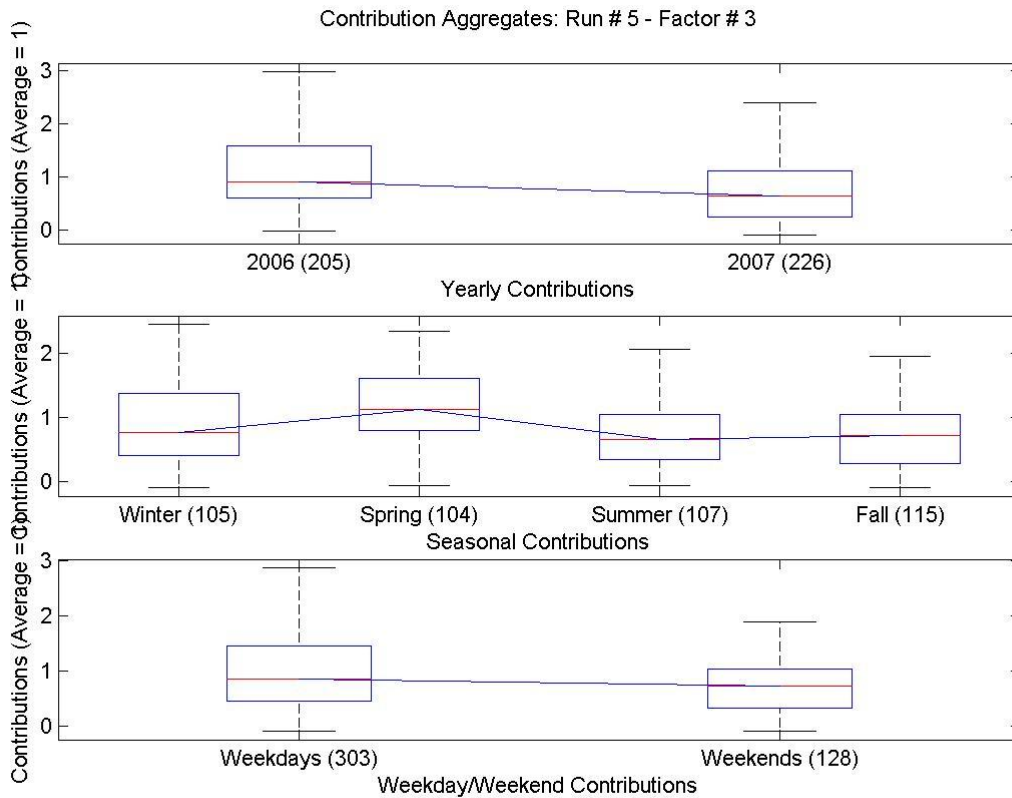


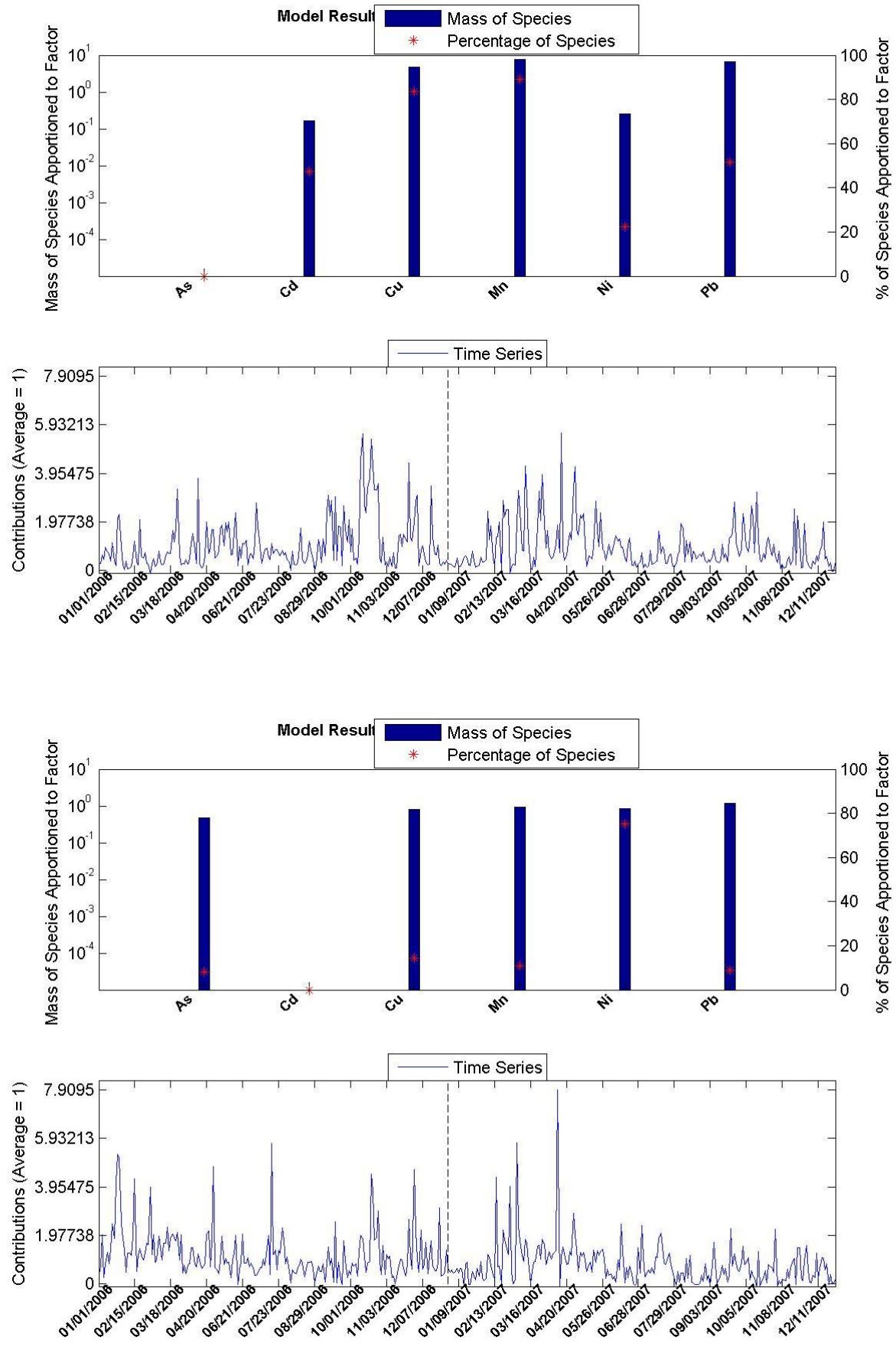


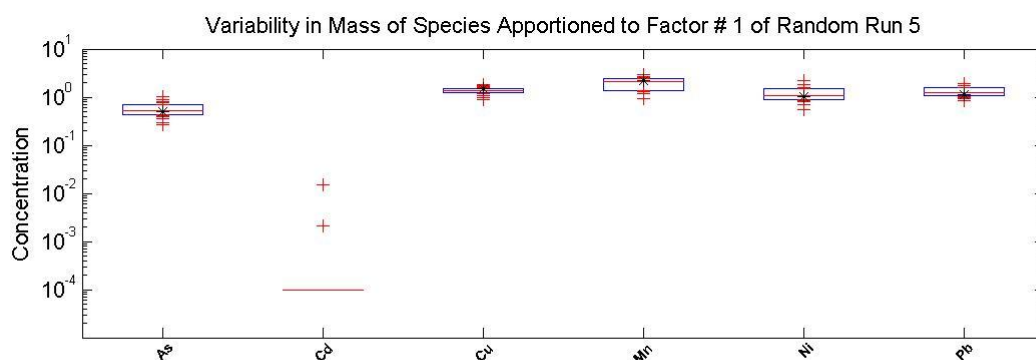
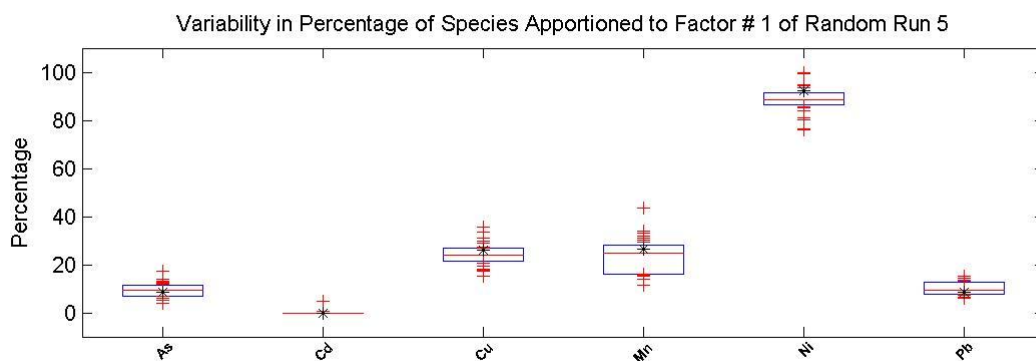


1.3. Vyhodnocení – těžké kovy – AIM Kladno - Švermov

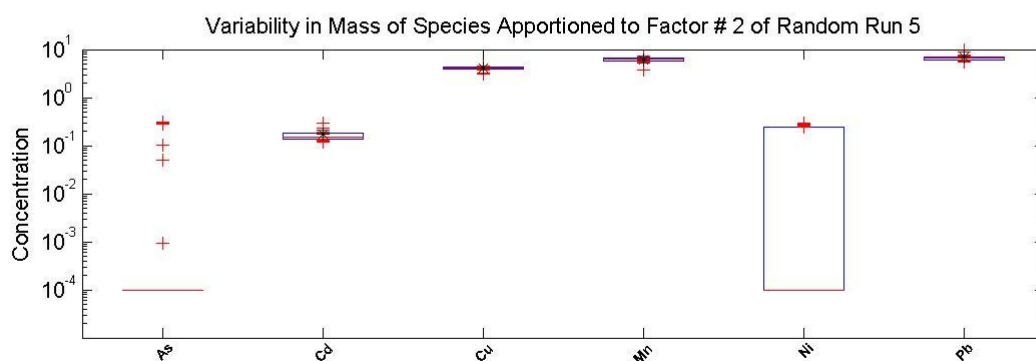
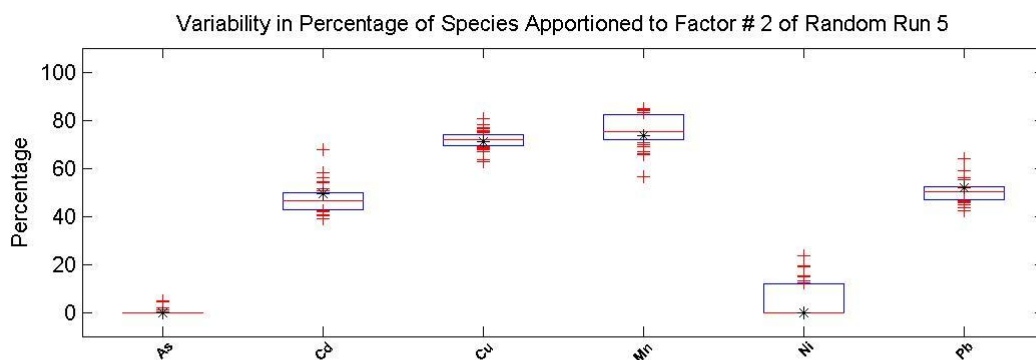








Concentrations truncated below at 10⁻⁴



Concentrations truncated below at 10⁻⁴

