IV GIORNATA DI CARATTERIZZAZIONE CHIMICA DEL PARTICOLATO ATMOSFERICO

Terni, 22-22 Novembre 2022

Identificazione delle sorgenti del particolato tramite Positive Matrix Factorization (PMF)

Fabiana Scotto

Arpae Emilia-Romagna

SCUOLA DI ALTA FORMAZIONE AMBIENTALE

LA POSITIVE MATRIX FACTORIZATION (PMF)

Come la PCA, la PMF è una tecnica di riduzione della dimensione del dataset, ottimizzata specificatamente per quantificare il contributo delle sorgenti di inquinamento. E' stata sviluppata da Paatero (Department of Physics, University of Helsinki) University of Helsinki) negli anni novanta (Paatero Paatero and Tapper, 1994), e in seguito da Paatero e Hopke (Paatero and Hopke, 2003) con lo scopo di sviluppare un nuovo metodo di analisi multivariata che superasse alcune limitazioni della PCA.

LA POSITIVE MATRIX FACTORIZATION (PMF)

Deve la sua grande diffusione principalmente a 2 aspetti:

- il fatto che non sia richiesta la conoscenza dei profili emissivi delle sorgenti

- la disponibilità di un software sviluppato dall'EPA, messo a disposizione gratuitamente e di facile utilizzo

IL MODELLO TEORICO

n num. campioni
m num. di analiti
p num. di fattori

p fattori

giorno 1

$$
\begin{aligned}
{}_{EC}x_1 &= {}_1g_{BB} * {}_{EC}f_{BB} + {}_1g_{OIL} * {}_{EC}f_{OIL} + \ldots + {}_1g_{TRAFF} * {}_{EC}f_{TRAFF} \\
{}_{Ca}x_1 &= {}_1g_{BB} * {}_{Ca}f_{BB} + {}_1g_{OIL} * {}_{Ca}f_{OIL} + \ldots + {}_1g_{TRAFF} * {}_{Ca}f_{TRAFF} \\
&\vdots \\
{}_{Cu}x_1 &= {}_1g_{BB} * {}_{Cu}f_{BB} + {}_1g_{OIL} * {}_{Cu}f_{OIL} + \ldots + {}_1g_{TRAFF} * {}_{Cu}f_{TRAFF}
\end{aligned}
$$

m analiti

giorno n

$$
\begin{aligned}
{}_{EC}x_n &= {}_ng_{BB} * {}_{EC}f_{BB} + {}_ng_{OIL} * {}_{EC}f_{OIL} + \ldots + {}_ng_{TRAFF} * {}_{EC}f_{TRAFF} \\
{}_{Ca}x_n &= {}_ng_{BB} * {}_{Ca}f_{BB} + {}_ng_{OIL} * {}_{Ca}f_{OIL} + \ldots + {}_ng_{TRAFF} * {}_{Ca}f_{TRAFF} \\
&\vdots \\
{}_{Cu}x_n &= {}_ng_{BB} * {}_{Cu}f_{BB} + {}_ng_{OIL} * {}_{Cu}f_{OIL} + \ldots + {}_ng_{TRAFF} * {}_{Cu}f_{TRAFF}
\end{aligned}
$$

Le g sono uguali per tutte le specie chimiche, cambiano al cambiare dei giorni e dei fattori: nxp

Le f sono uguali per tutti i giorni. Cambiano per i fattori e gli analiti: mxp

→ le incognite sono pxm+pxn=px(m+n)

I termini noti sono pari al numero dei giorni per il numero degli analiti: nxm

→ nxm>=px(n+m)

IL MODELLO TEORICO

p fattori

giorno 1

$$
\begin{aligned}
{}_{EC}x_1 &= {}_1g_{BB} * {}_{EC}f_{BB} + {}_1g_{OIL} * {}_{EC}f_{OIL} + \ldots + {}_1g_{TRAFF} * {}_{EC}f_{TRAFF} + {}_{EC}e_1 \\
{}_{Ca}x_1 &= {}_1g_{BB} * {}_{Ca}f_{BB} + {}_1g_{OIL} * {}_{Ca}f_{OIL} + \ldots + {}_1g_{TRAFF} * {}_{Ca}f_{TRAFF} + {}_{Ca}e_1 \\
&\vdots \\
{}_{Cu}x_1 &= {}_1g_{BB} * {}_{Cu}f_{BB} + {}_1g_{OIL} * {}_{Cu}f_{OIL} + \ldots + {}_1g_{TRAFF} * {}_{Cu}f_{TRAFF} + {}_{Cu}e_1
\end{aligned}
$$

m analiti

giorno n

$$
\begin{aligned}
{}_{EC}x_n &= {}_ng_{BB} * {}_{EC}f_{BB} + {}_ng_{OIL} * {}_{EC}f_{OIL} + \ldots + {}_ng_{TRAFF} * {}_{EC}f_{TRAFF} + {}_{EC}e_n \\
{}_{Ca}x_n &= {}_ng_{BB} * {}_{Ca}f_{BB} + {}_ng_{OIL} * {}_{Ca}f_{OIL} + \ldots + {}_ng_{TRAFF} * {}_{Ca}f_{TRAFF} + {}_{Ca}e_n \\
&\vdots \\
{}_{Cu}x_n &= {}_ng_{BB} * {}_{Cu}f_{BB} + {}_ng_{OIL} * {}_{Cu}f_{OIL} + \ldots + {}_ng_{TRAFF} * {}_{Cu}f_{TRAFF} + {}_{Cu}e_n
\end{aligned}
$$

n num. campioni
m num. di analiti
p num. di fattori

LA FUNZIONE Q

Si minimizza la funzione obiettivo Q così definita:

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

n numero campioni
m numero di variabili
p numero di sorgenti

LA FUNZIONE Q

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

n numero campioni
m numero di variabili
p numero di sorgenti

Ponderare ogni x_{ij} per l'inverso della sua incertezza permette di pesare diversamente ogni dato

Quindi ok giorni con dati mancanti, specie con molti dati sotto il LR, specie con una incertezza elevata

LA FUNZIONE Q

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

n numero campioni
m numero di variabili
p numero di sorgenti

Comprende:

- l'incertezza analitica
- l'incertezza del campionamento
- ma anche l'incertezza relativa al non completo verificarsi delle ipotesi del modello

LA FUNZIONE Q

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

n numero campioni
m numero di variabili
p numero di sorgenti

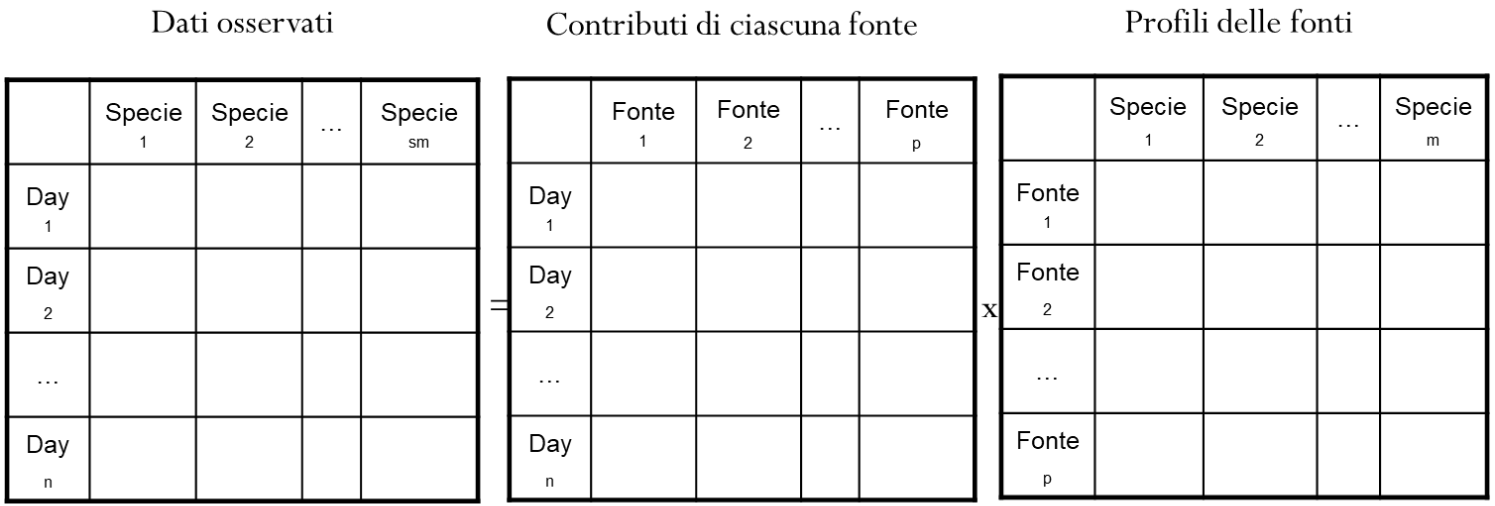
Tuttavia se non si ponessero dei vincoli, ci potrebbero essere infinite soluzioni.

Vincoli di non-negatività dei contributi e degli elementi dei profili.

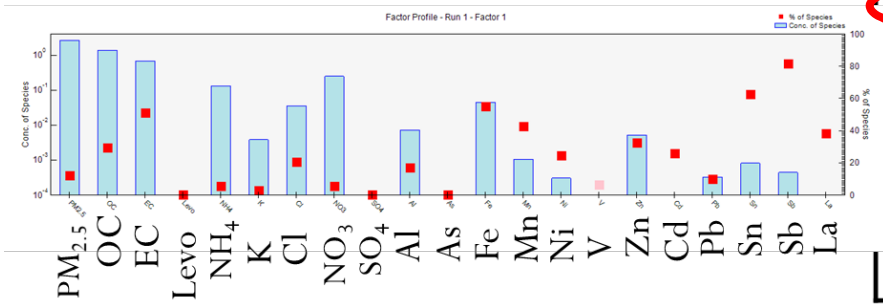
Soluzioni ragionevoli dal punto di vista interpretativo

INPUT E OUTPUT

$$X_{nxm} = G_{nxp} F_{pxm} + E_{nxm}$$

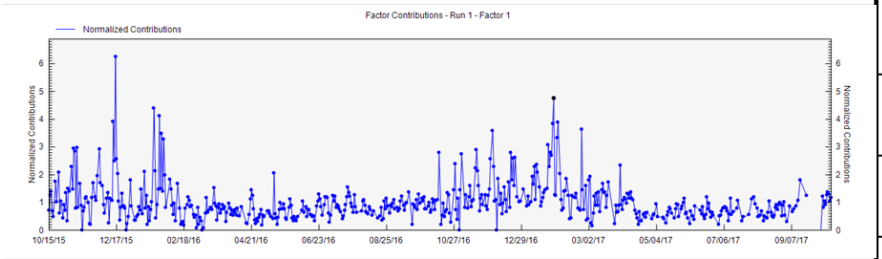


Quindi riassumendo



	Specie 1	Specie 2	…	Specie m
Fonte 1				
Fonte 2				
…				
Fonte p				

MATRICE DEI PROFILI



	Fonte 1	Fonte 2	…	Fonte p
Day_1				
Day_2				
…				
Day_n				

MATRICE DEI CONTRIBUTI

IPOTESI DEL MODELLO

Il profilo chimico delle sorgenti è costante nel tempo

➢ difficoltà nell'identificare sorgenti con profili variabili nel tempo (es. lunghe serie storiche, attività con cicli produttivi variabili,…)

Il profilo chimico delle emissioni è costante nello spazio: non si trasforma nel tragitto dalla sorgente al recettore

➢ non tutte le specie chimiche soddisfano questa ipotesi! Ok specie chimicamente STABILI (ad es. metalli); traccianti organici/volatili presentano problemi

➢ peculiarità per i secondari

IPOTESI DEL MODELLO

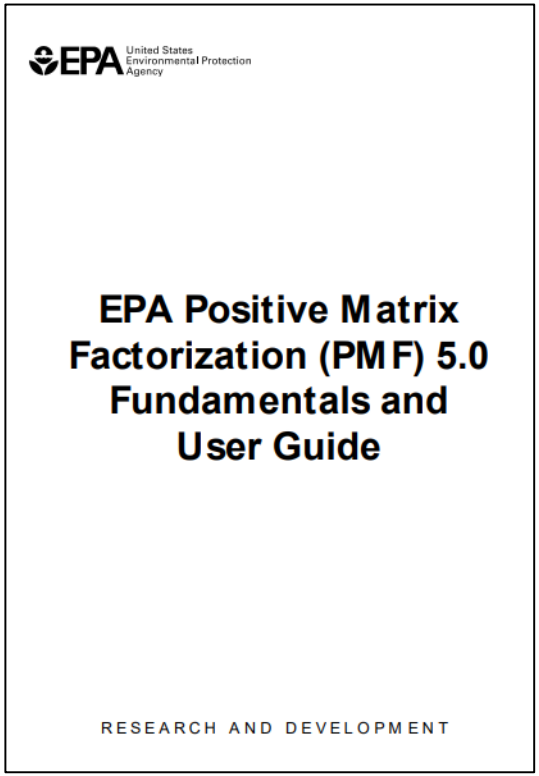
I profili e gli andamenti delle sorgenti sono linearmente indipendenti fra loro (non collineari), affinché i loro contributi siano correttamente distinti dal modello

➢ problema a distinguere sorgenti con profili chimici simili e andamenti correlati

• Devono essere identificate tutte le sorgenti che impattano in maniera rilevante al recettore

➢ speciazione più completa possibile, con traccianti delle diverse sorgenti

CHI BEN COMINCIA E' A META' DELL'OPERA…



PREDISPOSIZIONE DEL DATABASE DI INPUT

Dataset delle concentrazioni:

- ➔ valori < Limite di Rilevabilità
- ➔ dati mancanti

Dataset delle incertezze:

- ➔ tutto da costruire;
- ➔ deve tener conto dell'incertezza di campionamento e analitica ed è preferibile che le incertezze siano diverse da specie a specie (e non definite in via cautelativa dai laboratori con l'incertezza massima)

QUANTI DATI?

Il numero minimo di campioni necessari non può essere conosciuto a priori, perché dipende dalla quantità di informazione contenuta nel dataset: se c'è grande variabilità nei contributi relativi delle diverse sorgenti, è sufficiente sufficiente un numero minore di campioni, mentre l'aggiunta di campioni con similari contributi relativi delle sorgenti è di scarsa utilità! (European Guide on SA with RM, 2014)

- In accordo con il manuale EPA, almeno 100 campioni giornalieri e 20 specie chimiche (Norriset al., 2008, Brownand Hafner, 2005)

Model Data | Base Model | Help

Data Files | Concentration/Uncertainty | Concentration Scatter Plot | Concentration Time Series | Data Exceptions

Input Files

Model input data in tab-delimited (.txt), comma-separated value (.csv), or Excel workbook (.xls) format.
Species names in first row, units in second row (optional), and date/times in first column (optional).

Date Format: Automatic

Concentration Data File: C:\Users\fscotto\Documents\EPA PMF\Data\Dataset-Baltimore_con.txt — Browse | Load

Concentration data table with parameter names in the first row.
Optionally, the second row may contain units and the first
column may contain date/time.

Uncertainty Data File: C:\Users\fscotto\Documents\EPA PMF\Data\Dataset-Baltimore_unc.txt — Browse | Load

Observation-based or equation-based uncertainty values for each sample.
Must match concentration data format.

Date/Time Column: Date | ID/Site Column: | Unselect/Select All | Display Site ID Lines

Missing Value Indicator: -999 | (•) Exclude Entire Sample | ( ) Replace Missing Values with Species Median

Output Files

Output Folder: C:\Giornata formazione Terni\PMF\Nuova cartella — Browse

Specify a destination folder for all model run output files.

Output File Prefix: 7

Output File Type: (•) Tab-Delimited Text (*.txt) ( ) Comma-Delimited Text (*.csv) ( ) Excel 97-03 Workbook (*.xls) ( ) Excel 07-10 Workbook (*.xlsx)

[x] Output Only Selected Run [x] Display Overwrite Warning

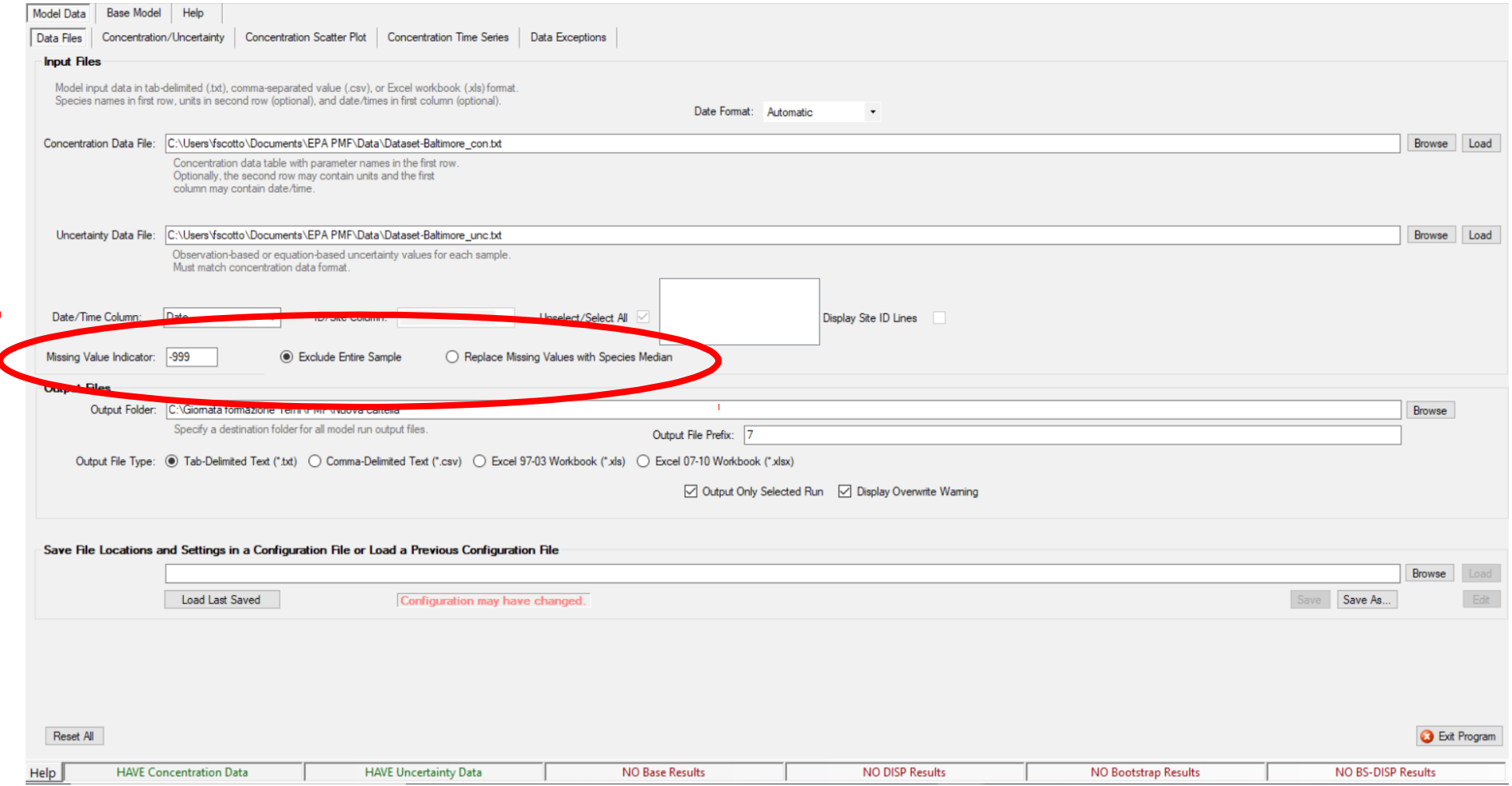
Save File Locations and Settings in a Configuration File or Load a Previous Configuration File

Browse | Load

Load Last Saved | Configuration may have changed. | Save | Save As... | Edit

Reset All | Exit Program

Help | HAVE Concentration Data | HAVE Uncertainty Data | NO Base Results | NO DISP Results | NO Bootstrap Results | NO BS-DISP Results


Model Data
Base Model
Help
Data Files
Concentration/Uncertainty
Concentration Scatter Plot
Concentration Time Series
Data Exceptions
Input Files
Model input data in tab-delimited (.txt), comma-separated value (.csv), or Excel workbook (.xls) format.
Species names in first row, units in second row (optional), and date/times in first column (optional).
Date Format: Automatic
Concentration Data File: C:\Users\fscotto\Documents\EPA PMF\Data\Dataset-Baltimore_con.txt
Browse
Load
Concentration data table with parameter names in the first row.
Optionally, the second row may contain units and the first
column may contain date/time.
Uncertainty Data File: C:\Users\fscotto\Documents\EPA PMF\Data\Dataset-Baltimore_unc.txt
Browse
Load
Observation-based or equation-based uncertainty values for each sample.
Must match concentration data format.
Date/Time Column:
Unselect/Select All
Display Site ID Lines
Missing Value Indicator: -999
Exclude Entire Sample
Replace Missing Values with Species Median
Output Folder:
Browse
Specify a destination folder for all model run output files.
Output File Prefix: 7
Output File Type: Tab-Delimited Text (*.txt) Comma-Delimited Text (*.csv) Excel 97-03 Workbook (*.xls) Excel 07-10 Workbook (*.xlsx)
Output Only Selected Run
Display Overwrite Warning
Save File Locations and Settings in a Configuration File or Load a Previous Configuration File
Browse
Load
Load Last Saved
Configuration may have changed.
Save
Save As...
Edit
Reset All
Exit Program
Help
HAVE Concentration Data
HAVE Uncertainty Data
NO Base Results
NO DISP Results
NO Bootstrap Results
NO BS-DISP Results

EPA PMF

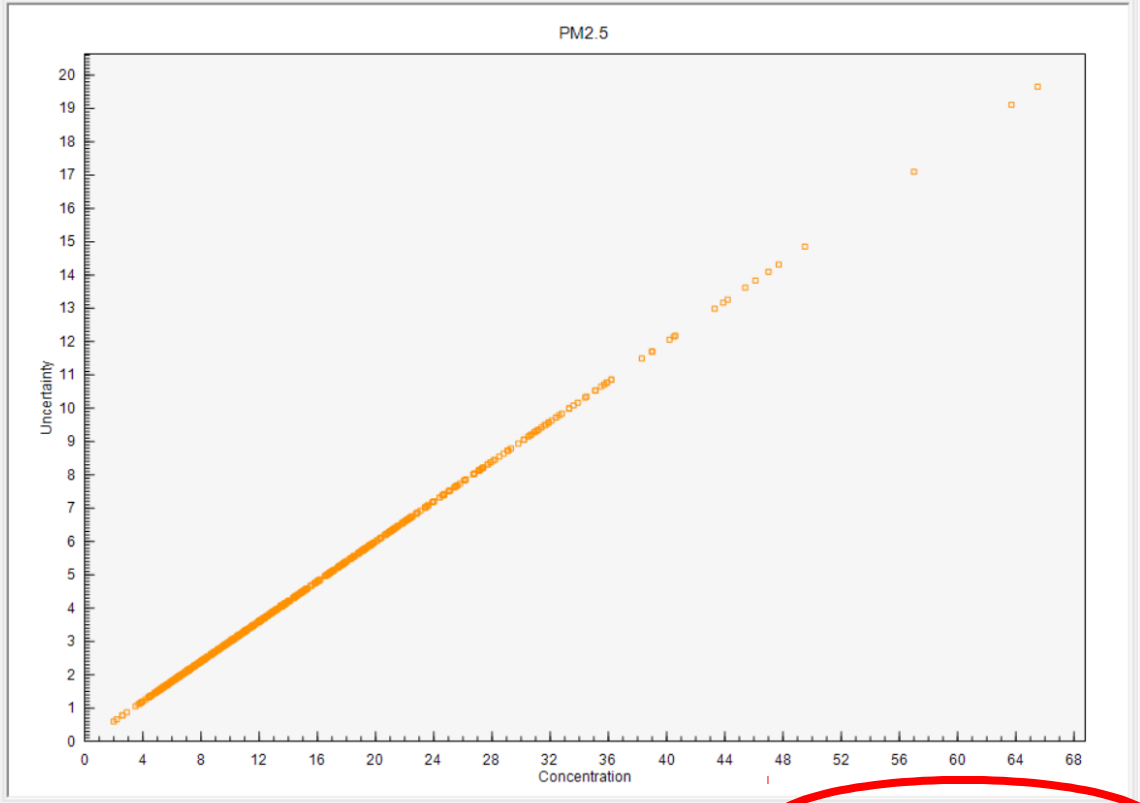
Model Data | Base Model | Help

Data Files | Concentration/Uncertainty | Concentration Scatter Plot | Concentration Time Series | Data Exceptions

Input Data Statistics

Species	Cat	S/N	Min	25th	50th	75th	Max
PM2.5	**Weak**	**9.0**	**2.00000**	**8.90000**	**13.50000**	**19.40000**	**65.50000**
Aluminum	Bad	0.1	0.00419	0.01250	0.01250	0.01250	0.38700
Ammonium Ion	Strong	8.9	0.01250	1.08500	1.68500	2.52750	9.22000
Arsenic	Bad	0.1	0.00098	0.00190	0.00190	0.00190	0.01080
Barium	Bad	0.0	0.00680	0.04450	0.04450	0.04450	0.24100
Bromine	Strong	2.0	0.00160	0.00160	0.00367	0.00534	0.02210
Calcium	Strong	2.1	0.00380	0.02390	0.03575	0.05328	0.22900
Chlorine	Bad	0.1	0.00264	0.00750	0.00750	0.01520	0.51400
Chromium	Bad	0.0	0.00052	0.00130	0.00130	0.00130	0.35400
Copper	Weak	1.0	0.00130	0.00130	0.00282	0.00445	0.04366
Elemental Carbon	Strong	4.4	0.12500	0.44275	0.64500	0.86825	3.91000
Iron	Strong	5.6	0.00499	0.05063	0.08150	0.12200	1.22000
Lead	Weak	0.5	0.00432	0.00445	0.00445	0.00445	0.04500
Manganese	Weak	0.4	0.00175	0.00175	0.00175	0.00198	0.03090
Nickel	Weak	0.5	0.00095	0.00095	0.00095	0.00217	0.11500
Organic Carbon	Strong	7.8	0.90800	3.11250	4.22000	5.43500	24.20000
OM	Bad	7.8	1.27120	4.35750	5.90800	7.60900	33.88000
Potassium Ion	Strong	2.1	0.01200	0.01200	0.05720	0.10400	0.45700
Selenium	Weak	0.2	0.00137	0.00170	0.00170	0.00170	0.01230
Silicon	Strong	2.0	0.00950	0.03033	0.05225	0.08040	1.02000
Sodium Ion	Strong	1.0	0.01500	0.04703	0.08810	0.16225	1.68000
Sulfate	Strong	9.2	0.11200	2.40750	3.76000	5.84750	30.20000
Titanium	Weak	0.7	0.00265	0.00265	0.00265	0.00645	0.07260
Total Nitrate	Strong	7.9	0.05100	0.67425	1.27000	2.32000	12.60000
Vanadium	Weak	0.6	0.00190	0.00190	0.00190	0.00426	0.01920

Concentration/Uncertainty Scatter Plot



Unsort | Species Category Settings: Strong | Weak | Bad | **Total Variable (Defaults to Weak)**

0 % Extra Modeling Uncertainty (0 - 100%)

Help | Strong Species: 12 | Weak Species: 8 | Bad Species: 6 | Samples Excluded: 0 %

CLASSIFICAZIONE DELLE VARIABILI

Guida del Software EPA

S/N < 0.5 → BAD
0.5 < S/N < 1.0 →WEAK
S/N>1.0 → STRONG

Paatero and Hopke (2003)

S/N < 0.2 → BAD
0.2 < S/N < 2.0 →WEAK
S/N>2.0 → STRONG

Ricordarsi di escludere le specie “doppie” (es. SO_4^{--} e S; K e K^+; TC se OC ed EC,...)

EPA PMF

Model Data | Base Model | Help

Data Files | Concentration/Uncertainty | Concentration Scatter Plot | Concentration Time Series | Data Exceptions

Select Species

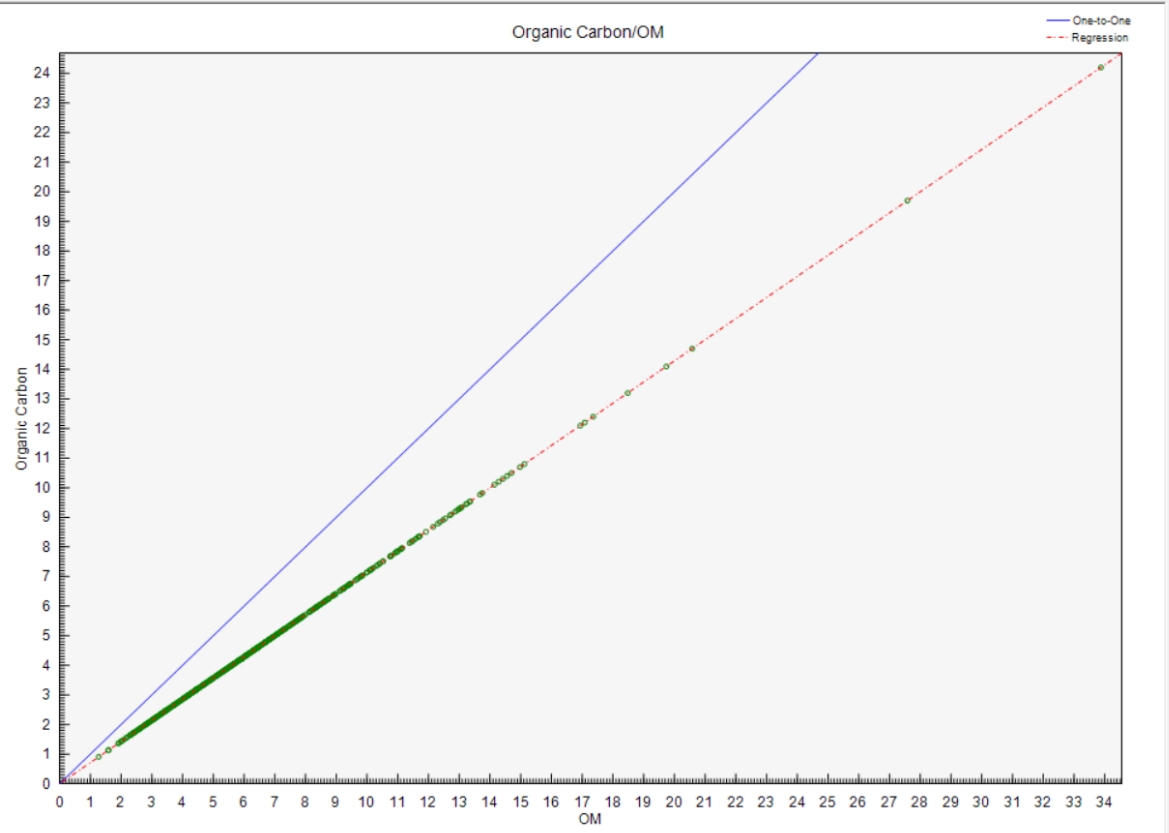
Y Axis

PM2.5
Aluminum
Ammonium Ion
Arsenic
Barium
Bromine
Calcium
Chlorine
Chromium
Copper
Elemental Carbon
Iron
Lead
Manganese
Nickel
Organic Carbon
OM
Potassium Ion
Selenium
Silicon
Sodium Ion
Sulfate
Titanium

X Axis

PM2.5
Aluminum
Ammonium Ion
Arsenic
Barium
Bromine
Calcium
Chlorine
Chromium
Copper
Elemental Carbon
Iron
Lead
Manganese
Nickel
Organic Carbon
OM
Potassium Ion
Selenium
Silicon
Sodium Ion
Sulfate
Titanium
Total Nitrate

Species Concentration



Help | 12/24/05 00:00 | Organic Carbon = 19.70000 | OM = 27.58000 | y = 0.71429x + 0.00000

EPA PMF

Model Data | Base Model | Help

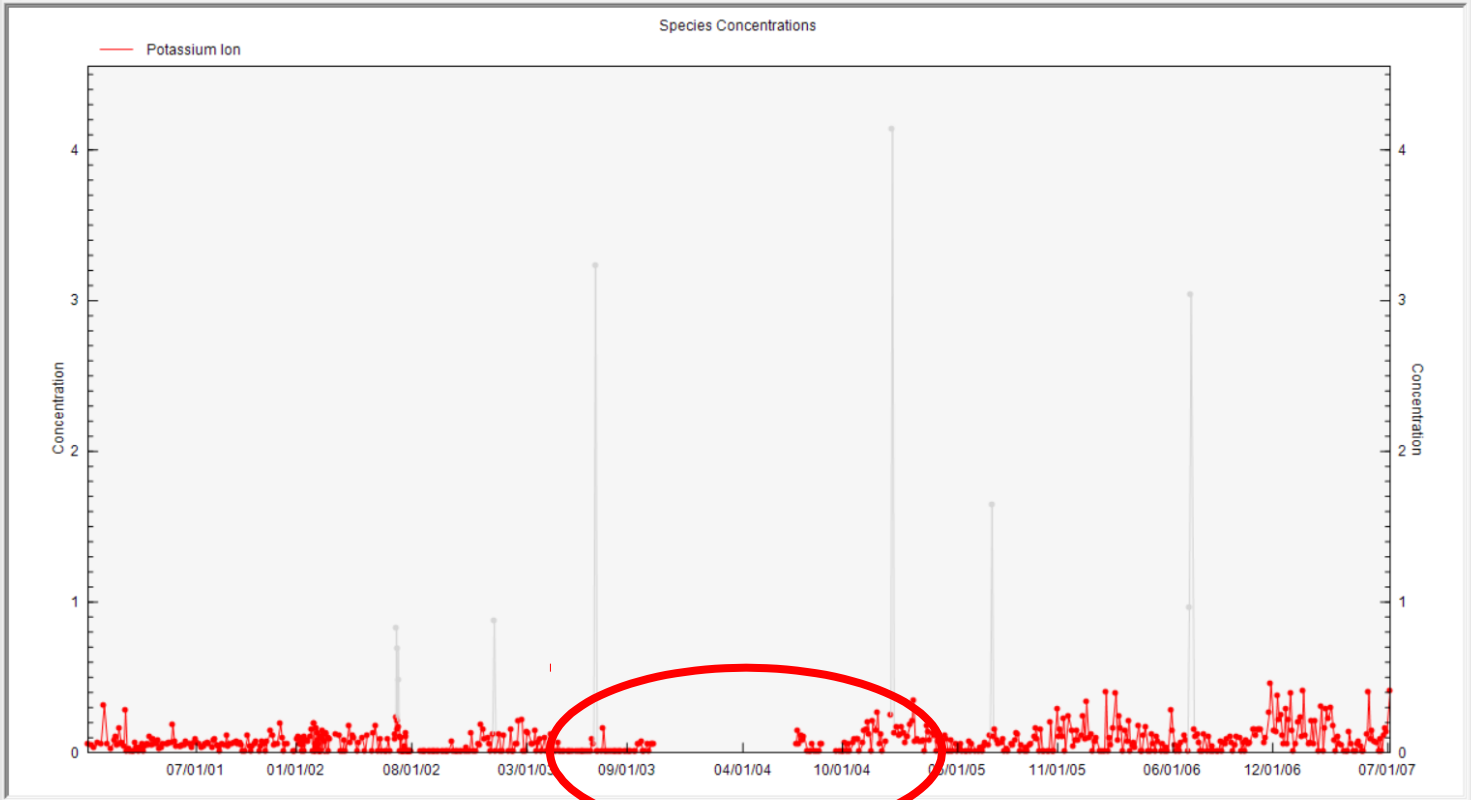
Data Files | Concentration/Uncertainty | Concentration Scatter Plot | Concentration Time Series | Data Exceptions

Select Species

- [ ] PM2.5
- [ ] Aluminum
- [ ] Ammonium Ion
- [ ] Arsenic
- [ ] Barium
- [ ] Bromine
- [ ] Calcium
- [ ] Chlorine
- [ ] Chromium
- [ ] Copper
- [ ] Elemental Carbon
- [ ] Iron
- [ ] Lead
- [ ] Manganese
- [ ] Nickel
- [ ] Organic Carbon
- [ ] OM
- [x] Potassium Ion
- [ ] Selenium
- [ ] Silicon
- [ ] Sodium Ion
- [ ] Sulfate
- [ ] Titanium
- [ ] Total Nitrate
- [ ] Vanadium
- [ ] Zinc

Clear Selections

Species Concentration



Log Scale

Exclude Samples | Restore Samples

Help

EPA PMF

Model Data | Base Model | Help

Data Files | Concentration/Uncertainty | Concentration Scatter Plot | Concentration Time Series | Data Exceptions

```
Input Data Exceptions:

Weak Species

PM2.5
Copper
Lead
Manganese
Nickel
Selenium
Titanium
Vanadium

Bad Species

Aluminum
Arsenic
Barium
Chlorine
Chromium
OM

User-Excluded Samples

02/28/02
07/04/02
07/07/02
07/08/02
12/31/02
07/05/03
01/01/05
07/03/05
07/01/06
07/04/06

Samples Excluded Due to Missing Values
```

Help

EPA PMF

Model Data | Base Model | Help

Base Model Runs

Base Model Runs

Number of Runs: 20 Number of Factors: 7

Random Start Seed Number: 52 Run

Error Estimation

Base Model Displacement Method

Selected Base Run: 1 Run

Base Model Bootstrap Method

Selected Base Run: 1

Block Size: 52 Suggest

Number of Bootstraps: 20

Min. Correlation R-Value: 0.6 Run

Base Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Bad	0.1
☐	Ammonium Ion	Strong	8.9
☐	Arsenic	Bad	0.1
☐	Barium	Bad	0.0
☐	Bromine	Strong	2.0
☐	Calcium	Strong	2.1

Run

Base Model Run Summary

Run Number	Q (Robust)	Q (True)	Converged

Factor Names

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7

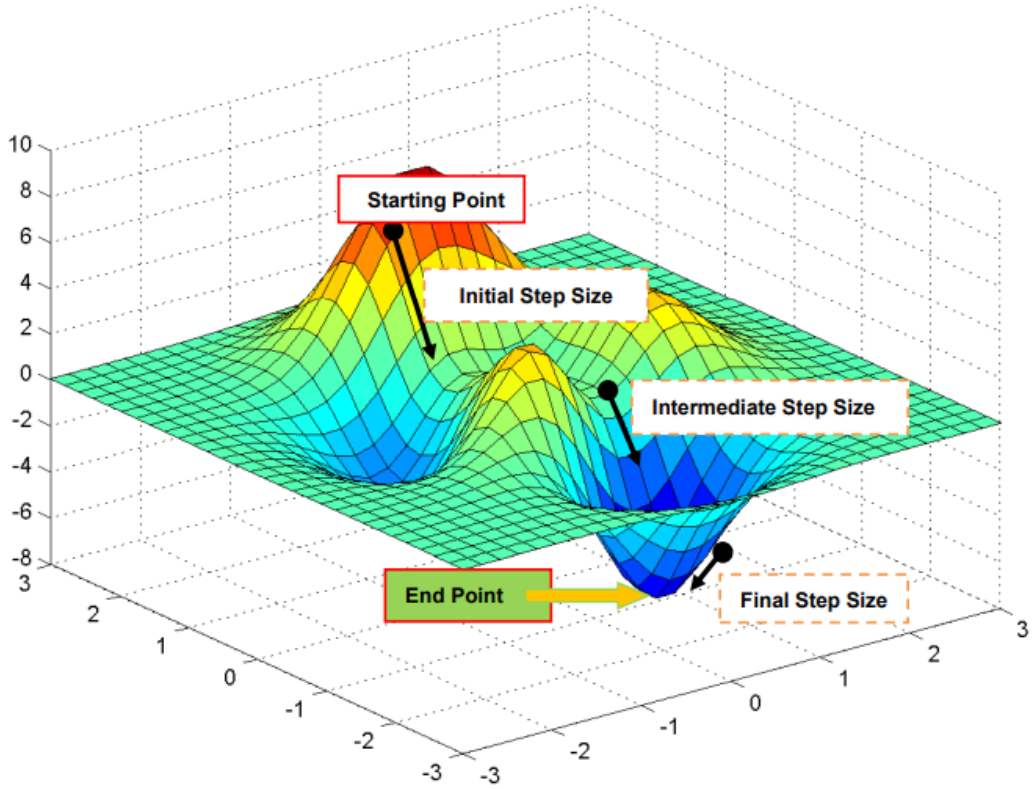
Apply to All Runs | Update Output Files | Reset to Defaults

Run Progress

Stop

Help | HAVE Concentration Data | HAVE Uncertainty Data | NO Base Results | NO DISP Results | NO Bootstrap Results | NO BS-DISP Results

LA FUNZIONE Q



EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

Base Model Runs

Number of Runs: 20 Number of Factors: 7

Random Start Seed Number: 77 Run

Error Estimation

Base Model Displacement Method

Selected Base Run: 20 Run

Base Model Bootstrap Method

Selected Base Run: 20

Block Size: 21 Suggest

Number of Bootstraps: 20

Min. Correlation R-Value: 0.6 Run

Base Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Bad	0.1
☐	Ammonium Ion	Strong	8.9
☐	Arsenic	Bad	0.1
☐	Barium	Bad	0.0
☐	Bromine	Strong	2.0
☐	Calcium	Strong	2.1

Run

Base Model Run Summary

Run Number	Q (Robust)	Q (True)	Converged
1	6702.4	7092.5	Yes
2	6702.5	7091.9	Yes
3	6702.5	7092.1	Yes
4	6702.5	7092.3	Yes
5	6702.4	7092.3	Yes
6	6702.5	7092.1	Yes
7	6702.4	7092.3	Yes
8	6702.5	7092.0	Yes
9	6702.5	7092.0	Yes
10	6702.6	7092.1	Yes
11	6702.5	7092.0	Yes
12	6702.5	7092.2	Yes
13	6702.6	7092.3	Yes
14	6702.5	7092.4	Yes
15	6702.5	7092.1	Yes
16	6702.5	7092.4	Yes
17	6702.4	7092.3	Yes
18	6702.5	7092.3	Yes
19	6702.6	7092.1	Yes
20	**6702.4**	**7092.2**	**Yes**

Factor Names

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Run 19	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Run 20	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7

Apply to All Runs | Update Output Files | Reset to Defaults

Run Progress

Stop

Help | HAVE Concentration Data | HAVE Uncertainty Data | HAVE Base Results | NO DISP Results | NO Bootstrap Results | NO BS-DISP Results

Run Number	Q (Robust)	Q (True)	
1	10146.6	12675.2	Yes
2	9019.9	9854.8	Yes
3	9019.8	9854.8	Yes
4	10205.5	12805.3	Yes
5	9019.9	9854.9	Yes
6	9019.8	9854.9	Yes
7	9019.7	9855.0	Yes
8	9019.9	9854.8	Yes
9	9020.0	9855.0	Yes
10	10085.8	12519.3	Yes
11	9019.8	9855.0	Yes
12	9020.2	9854.8	Yes
13	9020.2	9854.7	Yes
14	9020.1	9854.5	Yes
15	9019.9	9854.9	Yes
16	10141.8	12639.4	Yes
17	9019.7	9855.0	Yes
18	10085.6	12513.8	Yes
19	**9019.6**	**9855.2**	**Yes**
20	10147.2	12670.7	Yes

Esempio di run con diverse convergenze→ trova due minimi, uno dei quali sicuramente relativo

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

n numero campioni
m numero di variabili
p numero di sorgenti

$Q_{teorico}$ è pari al numero di gradi di libertà: m x n – p (n+m)

Verrebbe da pensare che la soluzione sia quella per la quale il $Q_{osservato}$ si avvicina il più possibile al $Q_{teorico}$, ma in realtà il $Q_{osservato}$ dipende in maniera troppo forte da come sono state settate le incertezze

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

n numero campioni
m numero di variabili
p numero di sorgenti

Q_{robust} è calcolato escludendo punti per i quali l'incertezza scalata per i residui è maggiore di 4 (Q_{ij} >4)

⇒ Q_{robust} non è influenzato dagli outliers

La differenza tra Q_{robust} e Q_{true} è una misura dell'impatto dei dati con alti residui. In particolare, soluzioni in cui il Q_{robust} è 2.5 volte più alto che il Q_{robust} possono indicare la necessità di sottopesare gli outliers.

EPA PMF

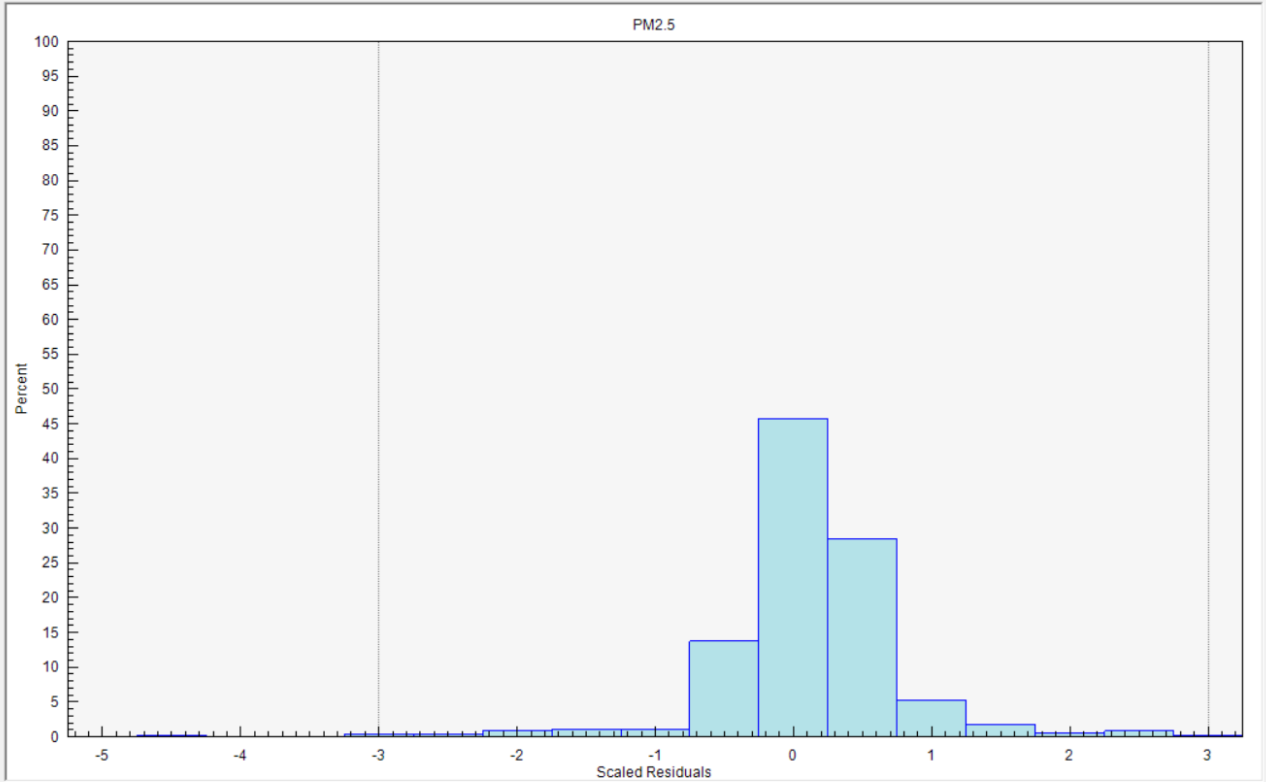
Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

- PM2.5
- Ammonium Ion
- Bromine
- Calcium
- Copper
- Elemental Carbon
- Iron
- Lead
- Manganese
- Nickel
- Organic Carbon
- Potassium Ion
- Selenium
- Silicon
- Sodium Ion
- Sulfate
- Titanium
- Total Nitrate
- Vanadium
- Zinc
- Calculated Sum of All Species

☐ Autoscale Histogram



◉ Dates by Species ○ Species by Dates

Absolute Scaled Residual Greater Than: 3.0

Species	Date/Time	Residuals
PM2.5	05/22/05 00:00	-3.18400
PM2.5	07/12/05 00:00	-4.44000
Ammonium Ion	09/10/01 00:00	-3.17200
Ammonium Ion	10/10/01 00:00	-3.22000
Ammonium Ion	10/13/01 00:00	-3.43100
Ammonium Ion	11/21/01 00:00	-4.14800
Ammonium Ion	02/01/02 00:00	-3.19300
Ammonium Ion	02/02/02 00:00	-3.29300
Ammonium Ion	12/11/04 00:00	-3.28600
Ammonium Ion	08/15/06 00:00	3.31100
Ammonium Ion	09/14/06 00:00	-3.19300
Ammonium Ion	09/23/06 00:00	4.13200
Bromine	02/09/01 00:00	3.17500
Bromine	02/26/02 00:00	3.39700
Bromine	04/08/02 00:00	3.43900
Bromine	11/16/02 00:00	4.93000
Bromine	12/07/02 00:00	5.01400
Bromine	02/14/03 00:00	-3.26600
Bromine	03/19/03 00:00	3.70600
Bromine	04/03/03 00:00	3.29300
Bromine	11/23/04 00:00	4.47900
Bromine	01/19/05 00:00	4.07700
Bromine	04/16/05 00:00	5.17700
Bromine	07/12/05 00:00	3.45700
Bromine	10/31/05 00:00	-3.02800

Run 20

EPA PMF

Model Data | Base Model | Rotational Tools | Help

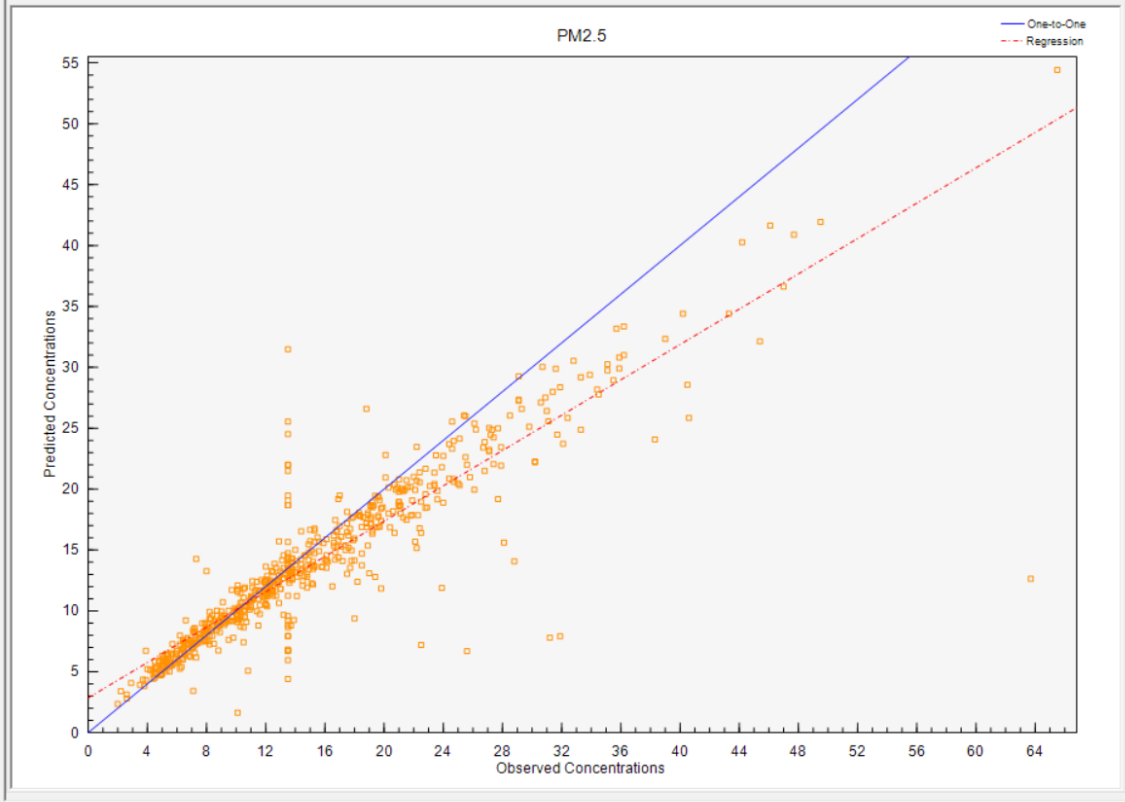
Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Base Run Statistics

Species	Category	r^2	Intercept	Intercept SE	Slope	Slope SE	SE
PM2.5	**Weak**	**0.81381**	**2.86052**	**0.24545**	**0.72533**	**0.01396**	**3.05**
Ammonium Ion	Strong	0.94252	0.04615	0.02296	0.97981	0.00973	0.3
Bromine	Strong	0.55159	0.00216	0.00010	0.54754	0.01986	0.0
Calcium	Strong	0.75082	0.00868	0.00093	0.74797	0.01733	0.0
Copper	Weak	0.19802	0.00272	0.00009	0.20509	0.01660	0.0
Elemental Carbon	Strong	0.83420	0.16909	0.01216	0.77442	0.01389	0.1
Iron	Strong	0.65465	0.03147	0.00261	0.66223	0.01935	0.0
Lead	Weak	0.25052	0.00404	0.00021	0.41770	0.02906	0.0
Manganese	Weak	0.33269	0.00168	0.00008	0.37148	0.02116	0.0
Nickel	Weak	0.01528	0.00167	0.00004	0.02215	0.00715	0.0
Organic Carbon	Strong	0.74674	0.95777	0.09366	0.77867	0.01824	1.0
Potassium Ion	Strong	0.98667	-0.00009	0.00049	0.99685	0.00466	0.0
Selenium	Weak	0.06946	0.00142	0.00008	0.27336	0.04025	0.0
Silicon	Strong	0.90244	0.01482	0.00112	0.75744	0.01002	0.0
Sodium Ion	Strong	0.07244	0.05599	0.00167	0.04822	0.00694	0.0
Sulfate	Strong	0.92262	0.22705	0.06855	0.97330	0.01134	1.0
Titanium	Weak	0.28132	0.00399	0.00015	0.26135	0.01680	0.0
Total Nitrate	Strong	0.99344	0.00912	0.00816	0.99814	0.00326	0.1
Vanadium	Weak	0.27738	0.00233	0.00009	0.35246	0.02288	0.0
Zinc	Strong	0.98880	0.00085	0.00012	0.96504	0.00413	0.0

Observed/Predicted Scatter Plot



Run 20

Help | 07/19/02 00:00 | Observed Concentration = 46.10000 | Predicted Concentration = 41.62400 | y = 0.72533x + 2.86052

EPA PMF

Model Data | Base Model | Rotational Tools | Help

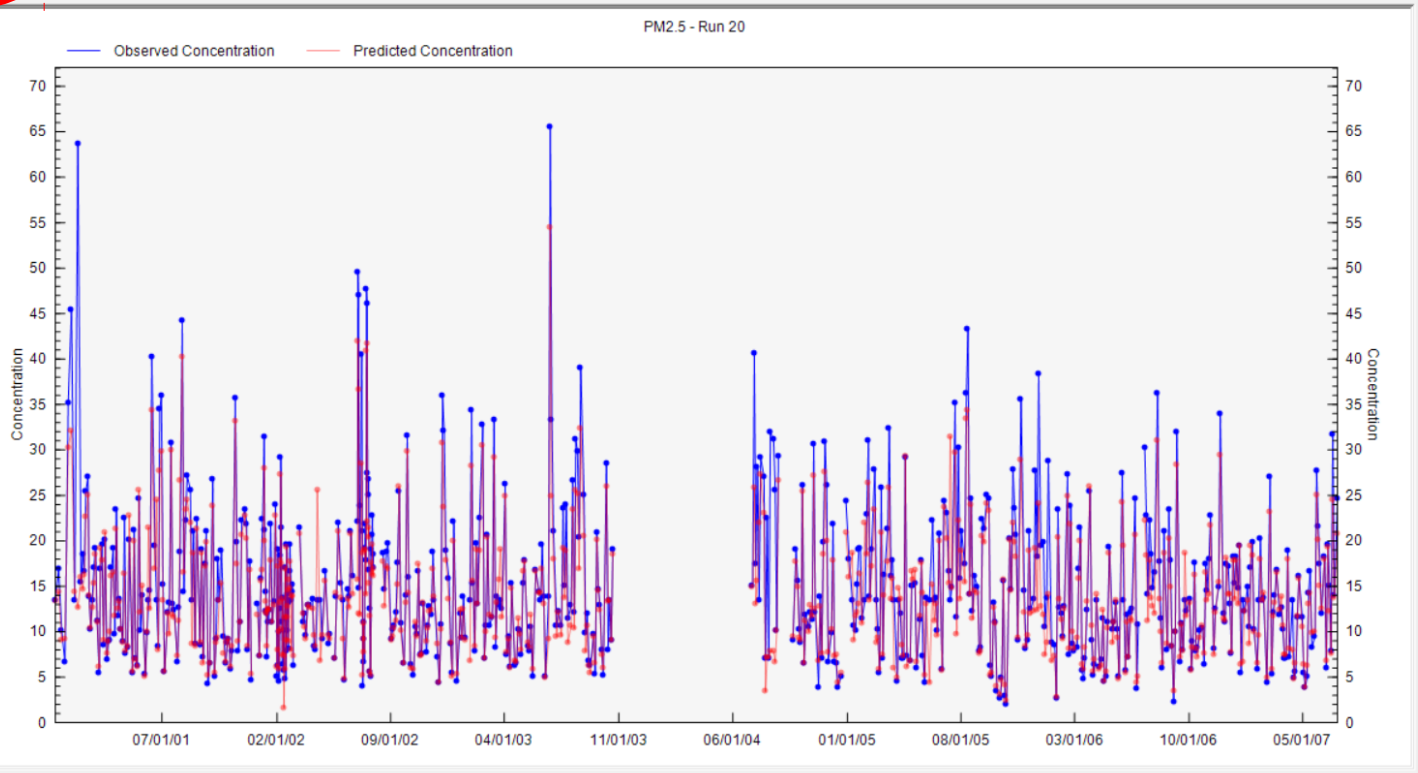
Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Select Species

- PM2.5
- Ammonium Ion
- Bromine
- Calcium
- Copper
- Elemental Carbon
- Iron
- Lead
- Manganese
- Nickel
- Organic Carbon
- Potassium Ion
- Selenium
- Silicon
- Sodium Ion
- Sulfate
- Titanium
- Total Nitrate
- Vanadium
- Zinc

Observed/Predicted Time Series



Run 20

Help | 11/03/01 00:00 | Observed Concentration = 9.20000 | Predicted Concentration = 9.24571

EPA PMF

Model Data | Base Model | Rotational Tools | Help

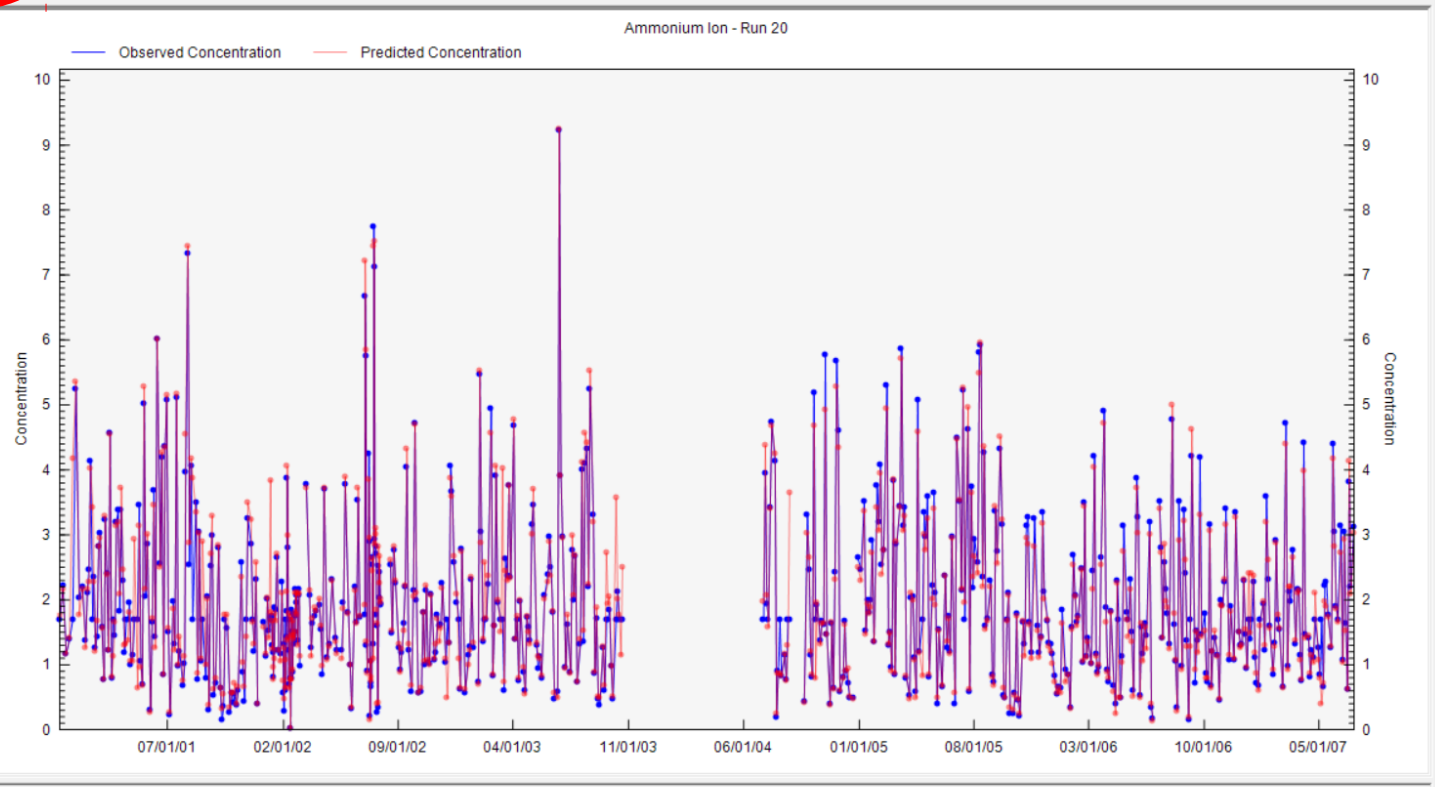
Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Select Species

- PM2.5
- Ammonium Ion
- Bromine
- Calcium
- Copper
- Elemental Carbon
- Iron
- Lead
- Manganese
- Nickel
- Organic Carbon
- Potassium Ion
- Selenium
- Silicon
- Sodium Ion
- Sulfate
- Titanium
- Total Nitrate
- Vanadium
- Zinc

Observed/Predicted Time Series



Run 20

Help

EPA PMF

Model Data | Base Model | Rotational Tools | Help

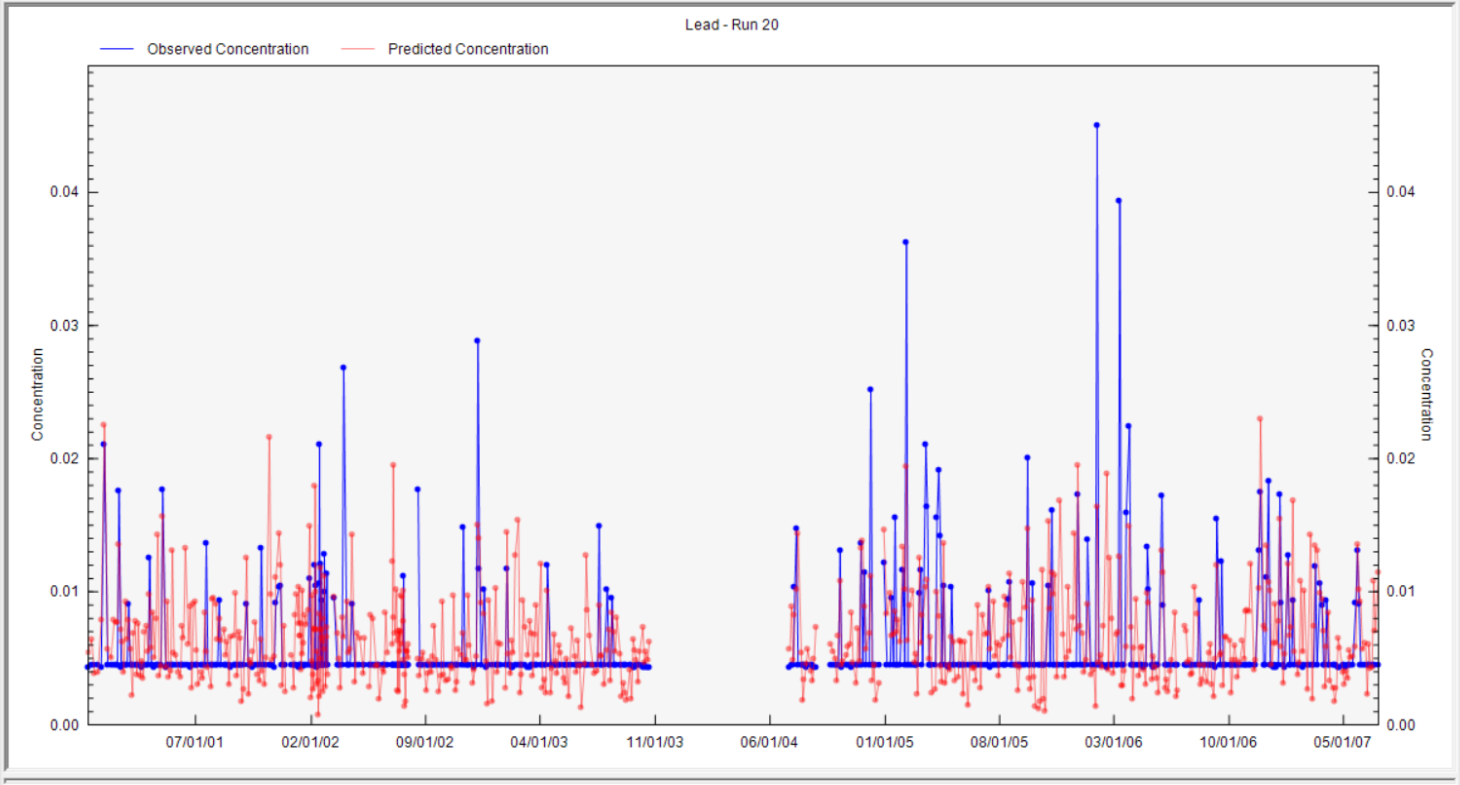
Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Select Species

- PM2.5
- Ammonium Ion
- Bromine
- Calcium
- Copper
- Elemental Carbon
- Iron
- Lead
- Manganese
- Nickel
- Organic Carbon
- Potassium Ion
- Selenium
- Silicon
- Sodium Ion
- Sulfate
- Titanium
- Total Nitrate
- Vanadium
- Zinc

Observed/Predicted Time Series



Run 20

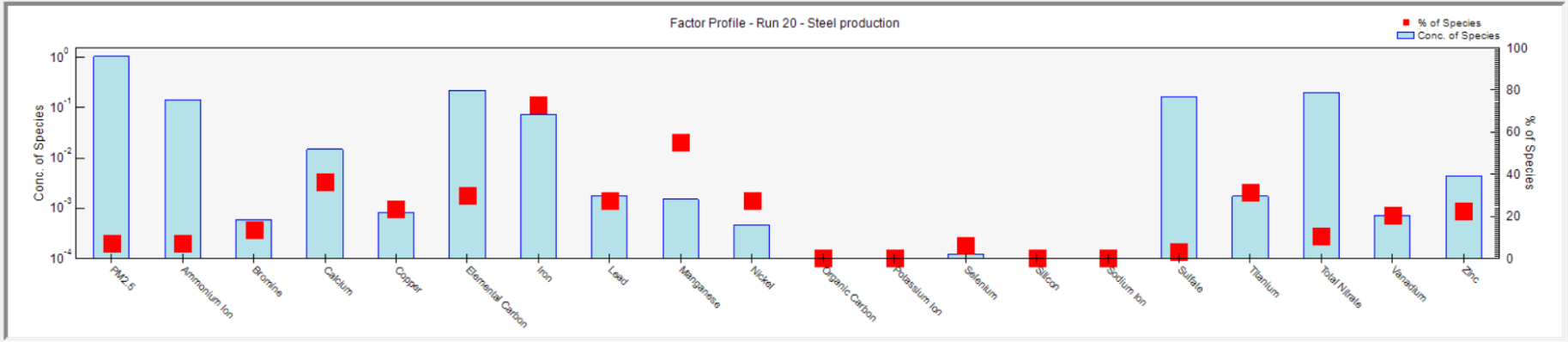
Help | 08/04/04 00:00 | Observed Concentration = 0.00445 | Predicted Concentration = 0.00340

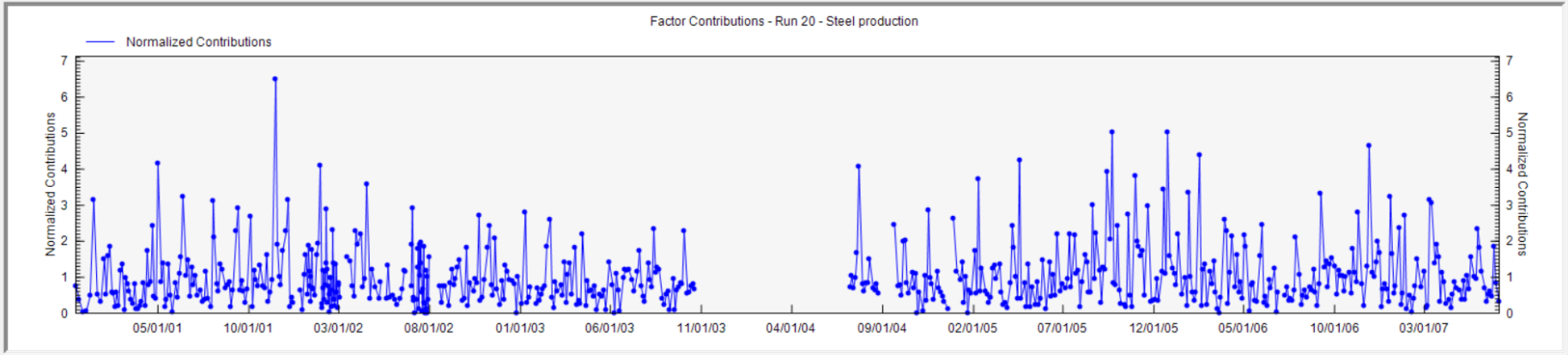
EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics





Concentration Units | Q/Qexp | Run 20 | Steel production

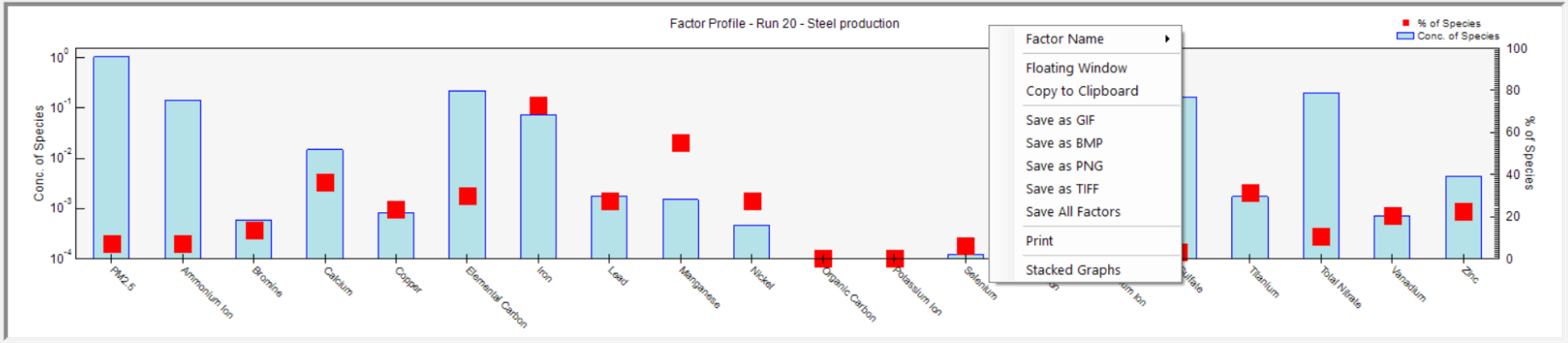
Help | 01/09/03 00:00 | Contribution = 2.80510

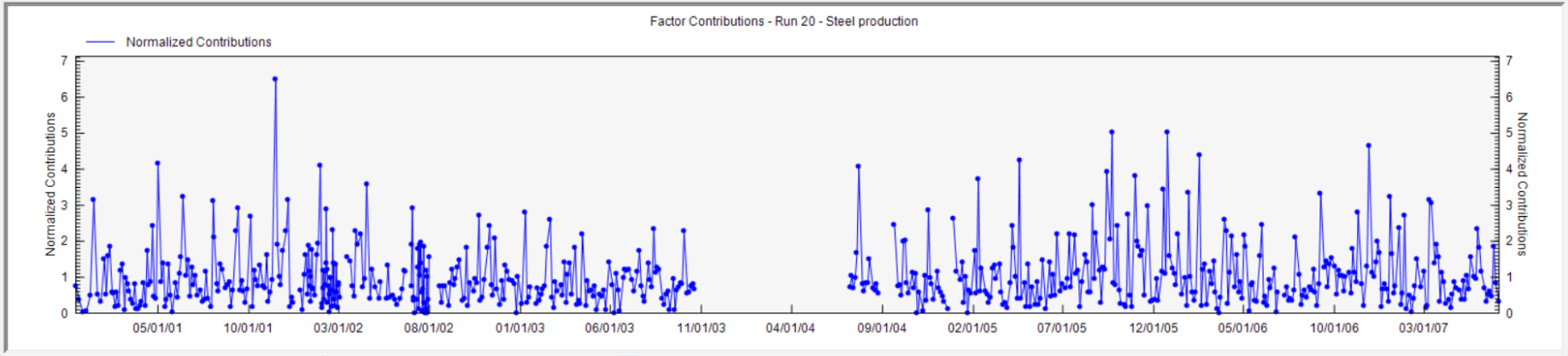
EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

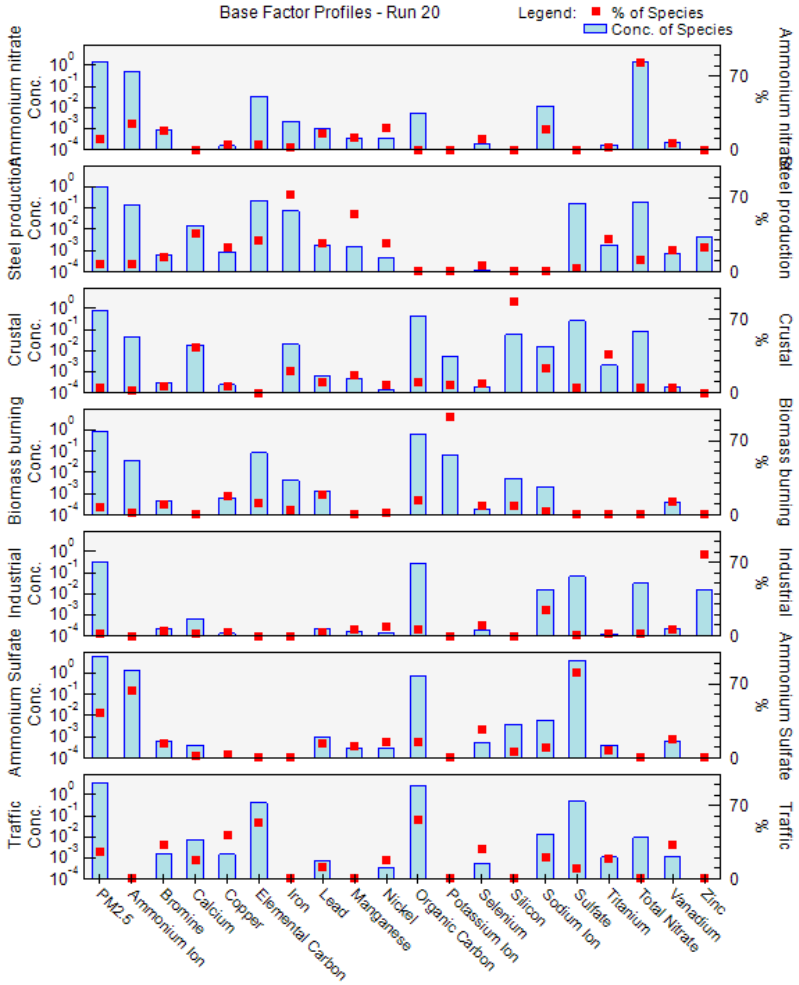
Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics





Concentration Units | Q/Qexp | Run 20 | Steel production

Help | 09/22/05 00:00 | Contribution = 5.03290

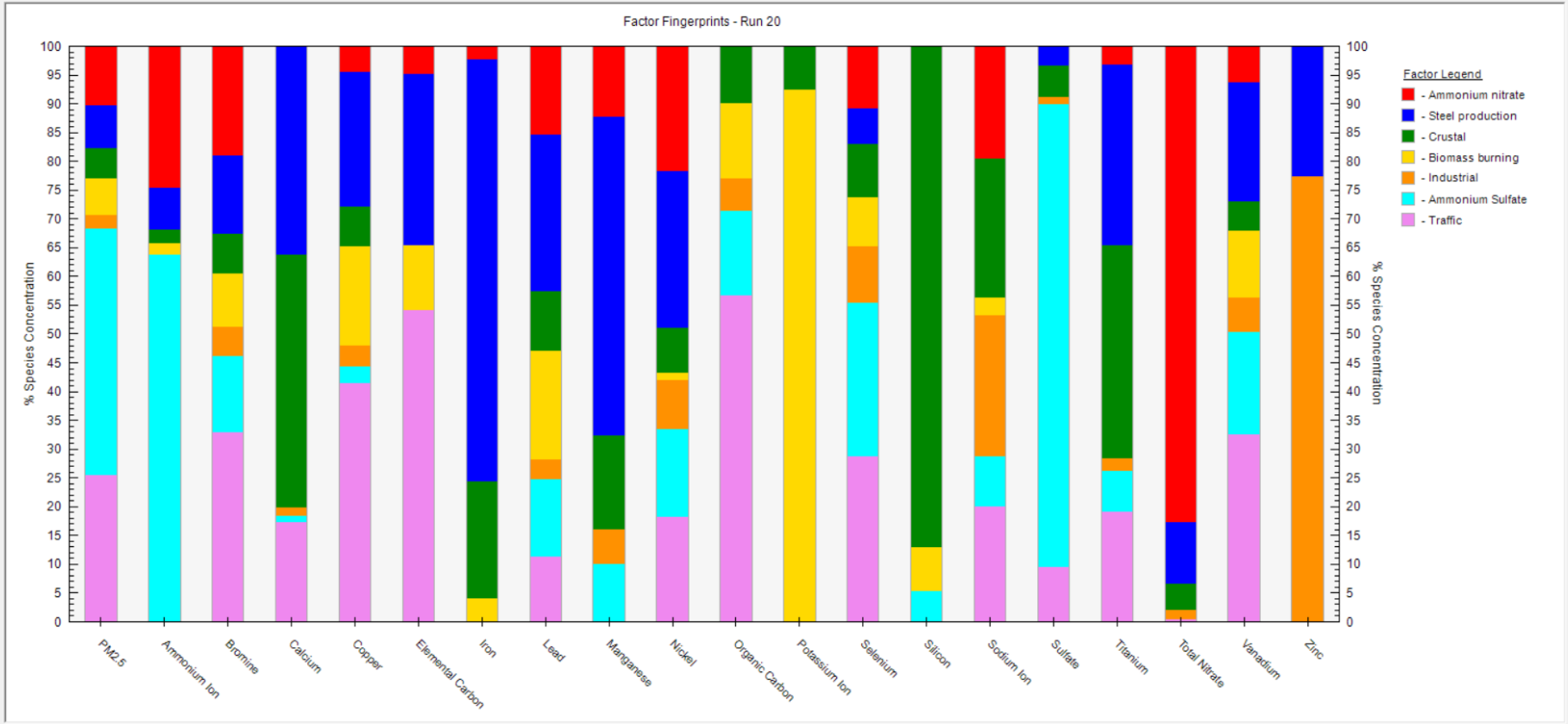

Base Factor Profiles - Run 20
Legend:
% of Species
Conc. of Species
Ammonium nitrate
Steel production
Crustal
Biomass burning
Industrial
Ammonium Sulfate
Traffic
Conc.
%
PM2.5
Ammonium Ion
Bromine
Calcium
Copper
Elemental Carbon
Iron
Lead
Manganese
Nickel
Organic Carbon
Potassium Ion
Selenium
Silicon
Sodium Ion
Sulfate
Titanium
Total Nitrate
Vanadium
Zinc

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics



Help

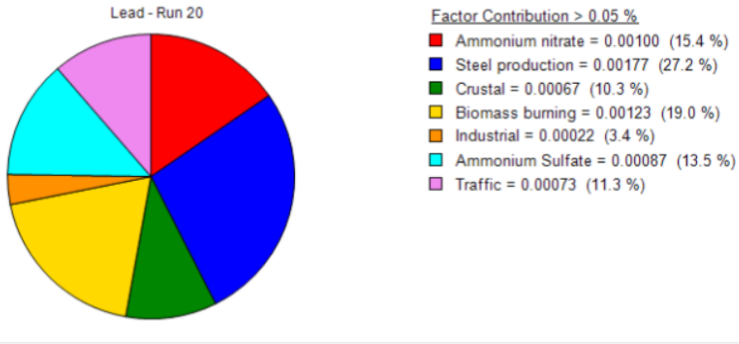
EPA PMF

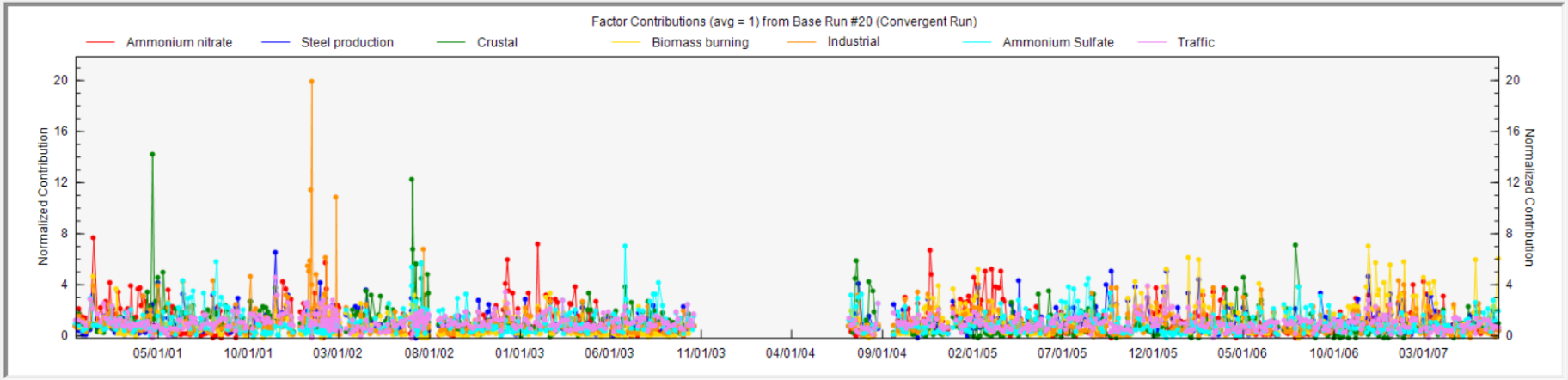
Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Species	Category
PM2.5	**Weak**
Ammonium Ion	Strong
Bromine	Strong
Calcium	Strong
Copper	Weak
Elemental Carbon	Strong
Iron	Strong
Lead	Weak
Manganese	Weak
Nickel	Weak
Organic Carbon	Strong





Run 20

Help

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Select Factors

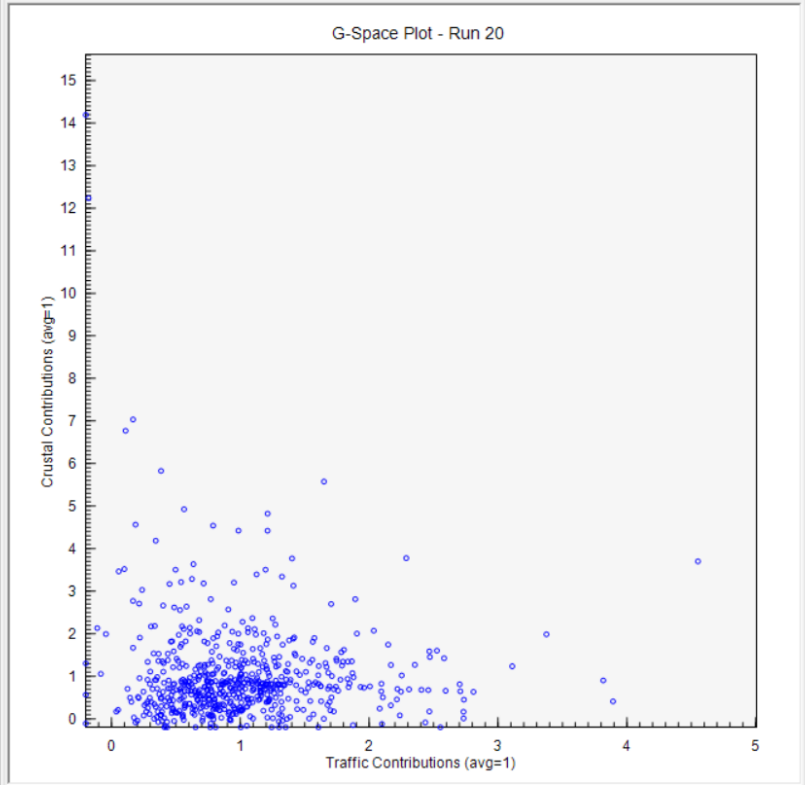
Y Axis

- Ammonium nitrate
- Steel production
- Crustal
- Biomass burning
- Industrial
- Ammonium Sulfate
- Traffic

X Axis

- Ammonium nitrate
- Steel production
- Crustal
- Biomass burning
- Industrial
- Ammonium Sulfate
- Traffic

G-Space Plot



Run 20

Help | 06/26/03 00:00 | Ammonium Sulfate = 7.00350 | Traffic = 2.28760

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Select Factors

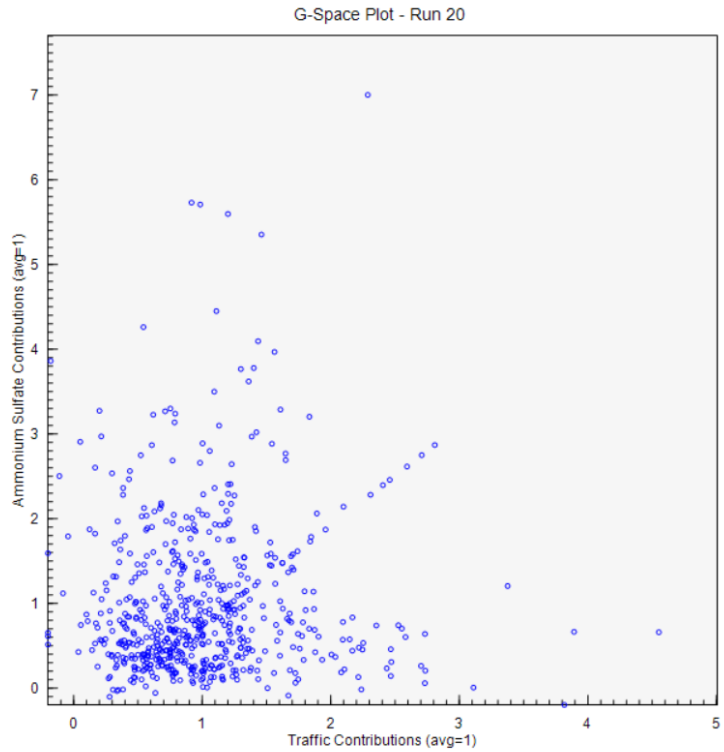
Y Axis

- Ammonium nitrate
- Steel production
- Crustal
- Biomass burning
- Industrial
- **Ammonium Sulfate**
- Traffic

X Axis

- Ammonium nitrate
- Steel production
- Crustal
- Biomass burning
- Industrial
- Ammonium Sulfate
- **Traffic**

G-Space Plot



Run 20

Help | 06/26/03 00:00 | Ammonium Sulfate = 7.00350 | Traffic = 2.28760

RIGUARDO ALL'ORTOGONALITA'...

A differenza della PCA, i fattori ottenuti in output dalla PMF non sono ortogonali fra loro per costruzione

E sempre a differenza della PCA; la PMF non è un modello gerarchico, per cui la scelta del numero dei fattori risulta molto più importante che nella PCA

AMBIGUITA' ROTAZIONALE

In generale, il vincolo di non-negatività da solo non è sufficiente a produrre un'unica soluzione. Matematicamente, un paio di matrici (G e F) che possono essere trasformate in un altro paio di matrici G* e F* con lo stesso valore Q è detto "ruotato":

$$G^*=GT \text{ e } F^*=T^{-1}F \rightarrow G^*F^*=GTT^{-1}F$$

La matrice T è una matrice pxp non singolare. A causa del vincolo di non negatività, una rotazione pura è possibile solo se nessuno degli elementi delle nuove matrici G e F è inferiore a 0. Se non ci sono rotazioni possibile, la soluzione è unica. Rotazioni approssimate che permettono un certo incremento di Q senza che alcun elemento delle nuove matrici diventi negativo, possono essere utilizzate.

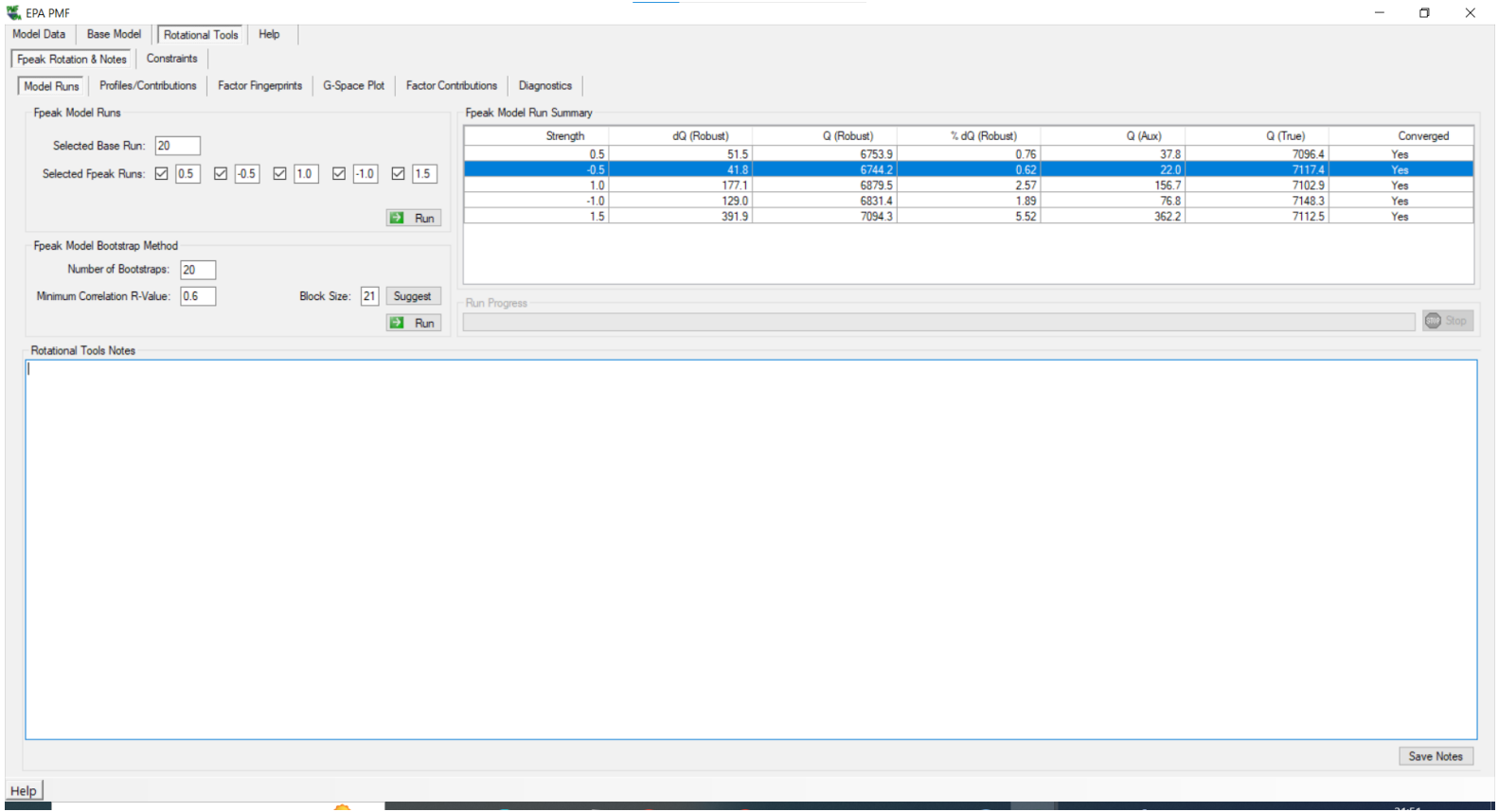
AMBIGUITA' ROTAZIONALE

Se c'è un numero sufficiente di zeri nei profili (matrice F) o nei contributi (matrice G) della soluzione, la soluzione non presenta ambiguità rotazionale.

In altre parole, dati con un impatto basso o nullo da una sorgente sono molto utili al modello!

Per questa ragione, sono molto utili dati con un'alta risoluzione temporale, e l'incremento della risoluzione temporale permette di avere campioni con una maggiore variabilità tra i contributi rispetto ai campioni integrati su periodi di tempo più lunghi

AMBIGUITA' ROTAZIONALE E PARAMETRO F-PEAK

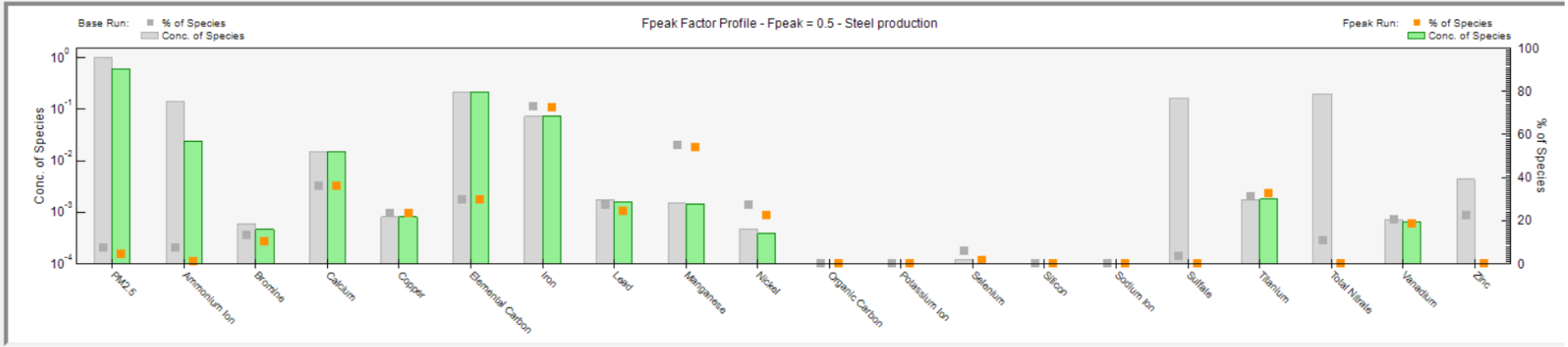


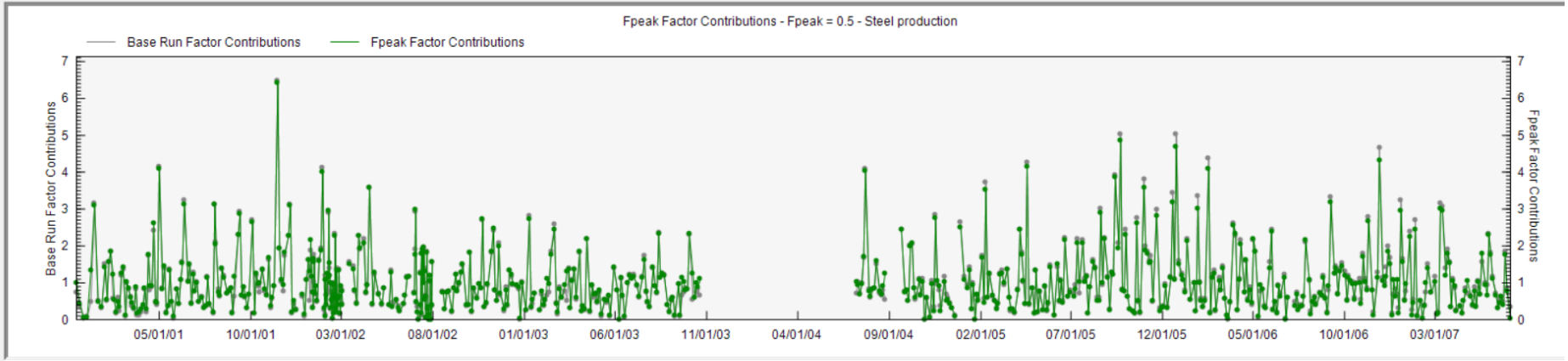
EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics





Concentration Units | Fpeak = 0.5 | Steel production

Help | 06/26/07 00:00 | Base Contribution = 1.84940 | Fpeak Contribution = 1.76700

INTRODUZIONE DI VINCOLI

Altri vincoli (oltre a quelli sulla positività) possono essere implementati se si dispone di una qualche conoscenza a priori, sfruttando un'interfaccia user-friendly che permette di imporre diversi tipi di "constrains"

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Expressions

Expression Builder

(•) Ratio ( ) Mass Balance ( ) Custom

Ratio

Factor:	Species (numerator):	Species (denominator):
Ammonium nitrate	Nickel	PM2.5
Steel production	Organic Carbon	Ammonium Ion
Crustal	Potassium Ion	Bromine
Biomass burning	Selenium	Calcium
Industrial	Silicon	Copper
Ammonium Sulfate	Sodium Ion	Elemental Carbon
Traffic	Sulfate	Iron
	Titanium	Lead
	Total Nitrate	Manganese
	Vanadium	Nickel
	Zinc	Organic Carbon

Value: 2.7 Add to Expressions

Expression	dQ	% dQ
[Ammonium nitrate|Sulfate] - 2.7 * [Ammonium nitrate|Ammonium Ion] = 0	33.51	0

Remove Selected Expressions | Remove All Expressions

Constraints

Add Constraints

Factor	Element	Type	Value	dQ	% dQ

Remove Selected Constraints | Remove All Constraints

Constrained Model Run

Run Selected Base Run: 20

dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
12.4	6714.7	0.18	8.7	7093.8	Yes

Error Estimation

Constrained Model Displacement Method

Run Selected Base Run: 20

Constrained Model Bootstrap Method

Number of Bootstraps: 20 Minimum Correlation R-Value: 0.6

Run Block Size: 21 Suggest

Constrained Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Bad	0.1
☐	Ammonium Ion	Strong	8.9
☐	Arsenic	Bad	0.1

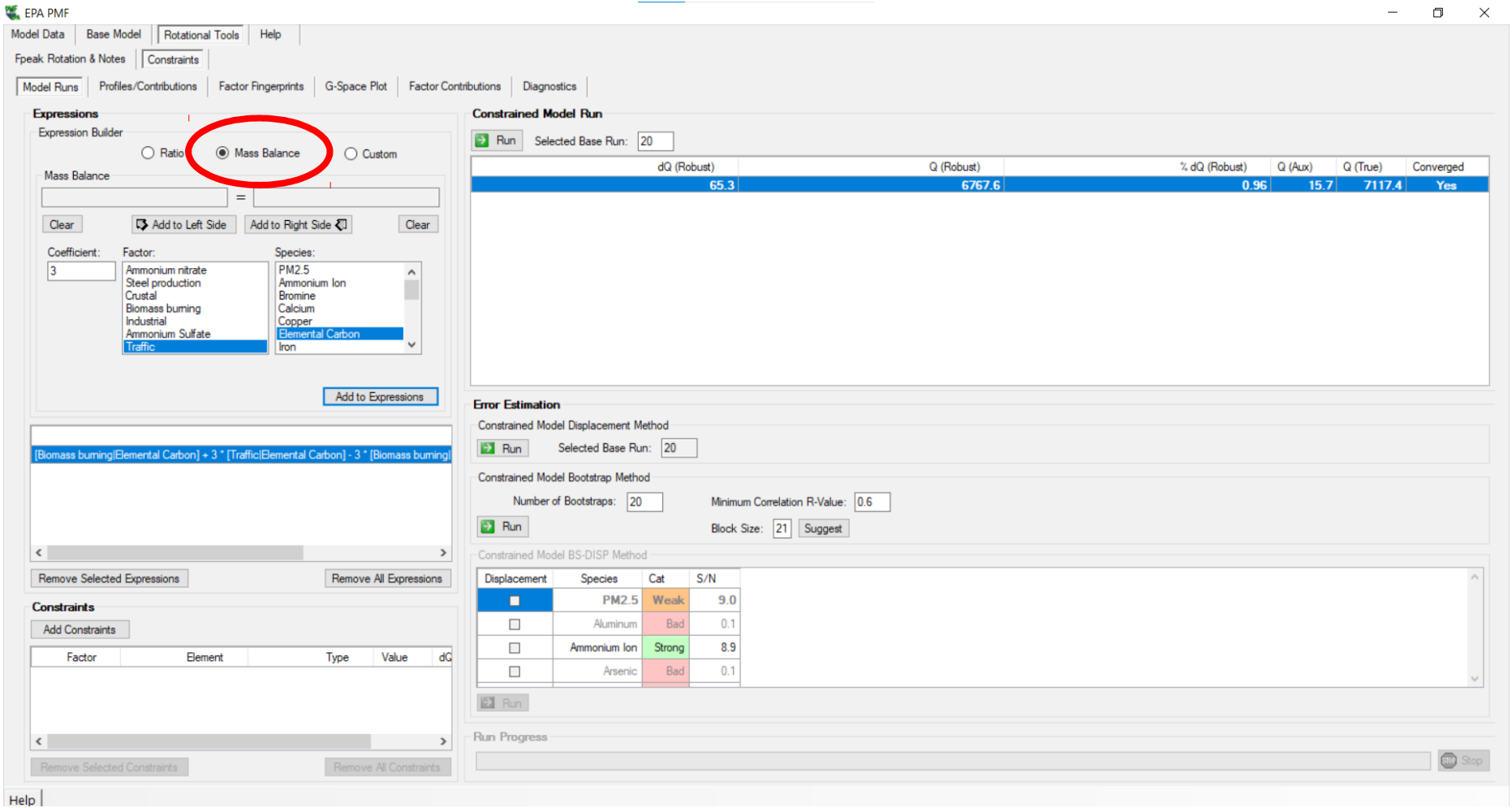
Run

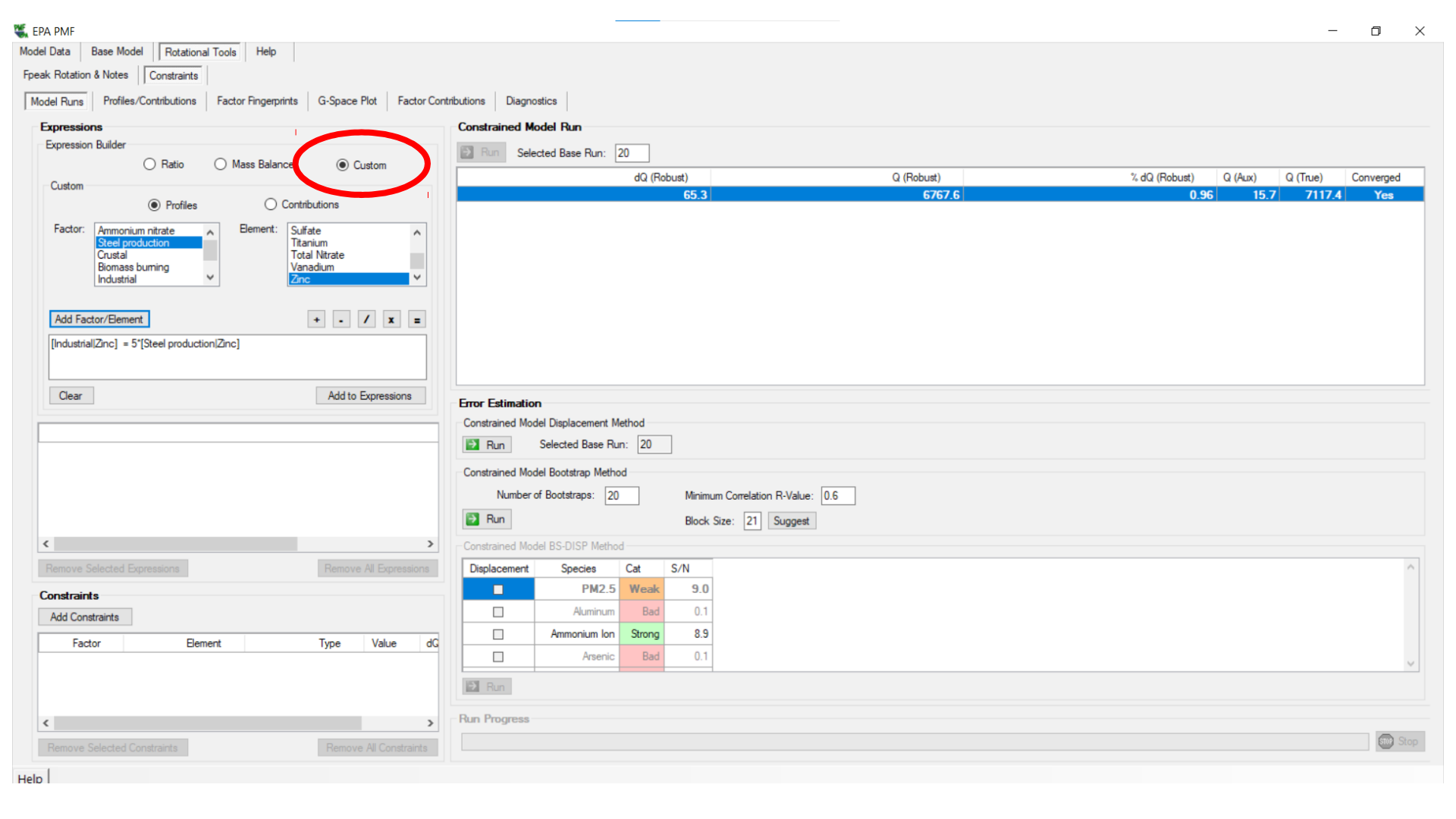
Run Progress

Stop

Help

1 nuova notifica


EPA PMF
Model Data
Base Model
Rotational Tools
Help
Fpeak Rotation & Notes
Constraints
Model Runs
Profiles/Contributions
Factor Fingerprints
G-Space Plot
Factor Contributions
Diagnostics
Expressions
Expression Builder
Ratio
Mass Balance
Custom
Mass Balance
=
Clear
Add to Left Side
Add to Right Side
Clear
Coefficient:
3
Factor:
Ammonium nitrate
Steel production
Crustal
Biomass burning
Industrial
Ammonium Sulfate
Traffic
Species:
PM2.5
Ammonium Ion
Bromine
Calcium
Copper
Elemental Carbon
Iron
Add to Expressions
[Biomass burning|Elemental Carbon] + 3 * [Traffic|Elemental Carbon] - 3 * [Biomass burning|
Remove Selected Expressions
Remove All Expressions
Constraints
Add Constraints
Factor
Element
Type
Value
dQ
Remove Selected Constraints
Remove All Constraints
Constrained Model Run
Run
Selected Base Run: 20
dQ (Robust)
Q (Robust)
% dQ (Robust)
Q (Aux)
Q (True)
Converged
65.3
6767.6
0.96
15.7
7117.4
Yes
Error Estimation
Constrained Model Displacement Method
Run
Selected Base Run: 20
Constrained Model Bootstrap Method
Number of Bootstraps: 20
Minimum Correlation R-Value: 0.6
Run
Block Size: 21
Suggest
Constrained Model BS-DISP Method
Displacement
Species
Cat
S/N
PM2.5
Weak
9.0
Aluminum
Bad
0.1
Ammonium Ion
Strong
8.9
Arsenic
Bad
0.1
Run
Run Progress
Stop
Help


EPA PMF
Model Data
Base Model
Rotational Tools
Help
Fpeak Rotation & Notes
Constraints
Model Runs
Profiles/Contributions
Factor Fingerprints
G-Space Plot
Factor Contributions
Diagnostics
Expressions
Expression Builder
Ratio
Mass Balance
Custom
Custom
Profiles
Contributions
Factor:
Ammonium nitrate
Steel production
Crustal
Biomass burning
Industrial
Element:
Sulfate
Titanium
Total Nitrate
Vanadium
Zinc
Add Factor/Element
+
-
/
x
=
[Industrial|Zinc] = 5*[Steel production|Zinc]
Clear
Add to Expressions
Remove Selected Expressions
Remove All Expressions
Constraints
Add Constraints
Factor
Element
Type
Value
dQ
Remove Selected Constraints
Remove All Constraints
Constrained Model Run
Run
Selected Base Run: 20
dQ (Robust)
Q (Robust)
% dQ (Robust)
Q (Aux)
Q (True)
Converged
65.3
6767.6
0.96
15.7
7117.4
Yes
Error Estimation
Constrained Model Displacement Method
Run
Selected Base Run: 20
Constrained Model Bootstrap Method
Number of Bootstraps: 20
Minimum Correlation R-Value: 0.6
Run
Block Size: 21
Suggest
Constrained Model BS-DISP Method
Displacement
Species
Cat
S/N
PM2.5
Weak
9.0
Aluminum
Bad
0.1
Ammonium Ion
Strong
8.9
Arsenic
Bad
0.1
Run
Run Progress
Stop
Help

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Expressions

Expression Builder

○ Ratio ○ Mass Balance ◉ Custom

Custom

◉ Profiles ○ Contributions

Factor: Ammonium nitrate, Steel production, Crustal, Biomass burning, Industrial

Element: Sulfate, Titanium, Total Nitrate, Vanadium, Zinc

Add Factor/Element + - / x =

Clear | Add to Expressions

Expression	dQ	% dQ
[Industrial|Zinc] - 5*[Steel production|Zinc] = 0	167.56	2.50

Remove Selected Expressions | Remove All Expressions

Constraints

Add Constraints

Factor	Element	Type	Value	dQ

Remove Selected Constraints | Remove All Constraints

Constrained Model Run

Run | Selected Base Run: 20

dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
65.3	6767.6	0.96	15.7	7117.4	Yes

Error Estimation

Constrained Model Displacement Method

Run | Selected Base Run: 20

Constrained Model Bootstrap Method

Number of Bootstraps: 20 | Minimum Correlation R-Value: 0.6

Run | Block Size: 21 | Suggest

Constrained Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Bad	0.1
☐	Ammonium Ion	Strong	8.9
☐	Arsenic	Bad	0.1

Run

Run Progress

Stop

Help

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Expressions

Expression Builder

(•) Ratio ( ) Mass Balance ( ) Custom

Ratio

Factor:	Species (numerator):	Species (denominator):
Ammonium nitrate	PM2.5	PM2.5
Steel production	Ammonium Ion	Ammonium Ion
Crustal	Bromine	Bromine
Biomass burning	Calcium	Calcium
Industrial	Copper	Copper
Ammonium Sulfate	Elemental Carbon	Elemental Carbon
Traffic	Iron	Iron
	Lead	Lead
	Manganese	Manganese
	Nickel	Nickel
	Organic Carbon	Organic Carbon

Value: 2.7 Add to Expressions

Expression	dQ	% dQ

Remove Selected Expressions | Remove All Expressions

Constraints

Add Constraints

Factor	Element	Type	Value	dQ
Ammonium nitrate	Elemental Carbon	Pull Down Maximally	NA	33.51

Remove Selected Constraints | Remove All Constraints

Constrained Model Run

Run Selected Base Run: 20

dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
12.4	6714.7	0.18	8.7	7093.8	Yes

Error Estimation

Constrained Model Displacement Method

Run Selected Base Run: 20

Constrained Model Bootstrap Method

Number of Bootstraps: 20 Minimum Correlation R-Value: 0.6

Run Block Size: 21 Suggest

Constrained Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Bad	0.1
☐	Ammonium Ion	Strong	8.9
☐	Arsenic	Bad	0.1

Run

Run Progress

Stop

Help

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Expressions

Expression Builder

(•) Ratio ( ) Mass Balance ( ) Custom

Ratio

Factor:	Species (numerator):	Species (denominator):
Ammonium nitrate	PM2.5	PM2.5
Steel production	Ammonium Ion	Ammonium Ion
Crustal	Bromine	Bromine
Biomass burning	Calcium	Calcium
Industrial	Copper	Copper
Ammonium Sulfate	Elemental Carbon	Elemental Carbon
Traffic	Iron	Iron
	Lead	Lead
	Manganese	Manganese
	Nickel	Nickel
	Organic Carbon	Organic Carbon

Value: 2.7 Add to Expressions

Expression	dQ	% dQ

Remove Selected Expressions | Remove All Expressions

Constraints

Add Constraints

Factor	Element	Type	Value	dQ
Steel production	1/3/2003 12:00:00 AM	Set to Zero	0	NA
Steel production	12/25/2006 12:00:00 AM	Set to Zero	NA	33.5

Pull Down Maximally
Pull Up Maximally
Pull to Value
Set to Zero
Set to Original Value
Define Limits

Remove Selected Constraints | e All Constraints

Constrained Model Run

Run Selected Base Run: 20

dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
12.4	6714.7	0.18	8.7	7093.8	Yes

Error Estimation

Constrained Model Displacement Method

Run Selected Base Run: 20

Constrained Model Bootstrap Method

Number of Bootstraps: 20 Minimum Correlation R-Value: 0.6

Run Block Size: 21 Suggest

Constrained Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Bad	0.1
☐	Ammonium Ion	Strong	8.9
☐	Arsenic	Bad	0.1

Run

Run Progress

Stop

Help

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Expressions

Expression Builder

(•) Ratio ( ) Mass Balance ( ) Custom

Ratio

Factor:	Species (numerator):	Species (denominator):
Ammonium nitrate	Manganese	PM2.5
Steel production	Nickel	Ammonium Ion
Crustal	Organic Carbon	Bromine
Biomass burning	Potassium Ion	Calcium
Industrial	Selenium	Copper
Ammonium Sulfate	Silicon	Elemental Carbon
Traffic	Sodium Ion	Iron
	Sulfate	Lead
	Titanium	Manganese
	Total Nitrate	Nickel
	Vanadium	Organic Carbon

Value: 2.7 Add to Expressions

Expression	dQ	% dQ
[Ammonium Sulfate|Sulfate] - 2.7 * [Ammonium Sulfate|Ammonium Ion] = 0	33.51	

Remove Selected Expressions | Remove All Expressions

Constraints

Add Constraints

Factor	Element	Type	Value	dQ
Steel production	1/3/2003 12:00:00 AM	Set to Zero	0	N
Steel production	12/25/2006 12:00:00 AM	Set to Zero	0	N
Ammonium nitrate	Elemental Carbon	Pull Down Maximally	NA	33

Remove Selected Constraints | Remove All Constraints

Constrained Model Run

Run Selected Base Run: 20

dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
65.3	6767.6	0.96	15.7	7117.4	Yes

Error Estimation

Constrained Model Displacement Method

Run Selected Base Run: 20

Constrained Model Bootstrap Method

Number of Bootstraps: 20 Minimum Correlation R-Value: 0.6

Run Block Size: 21 Suggest

Constrained Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Bad	0.1
☐	Ammonium Ion	Strong	8.9
☐	Arsenic	Bad	0.1

Run

Run Progress

Stop

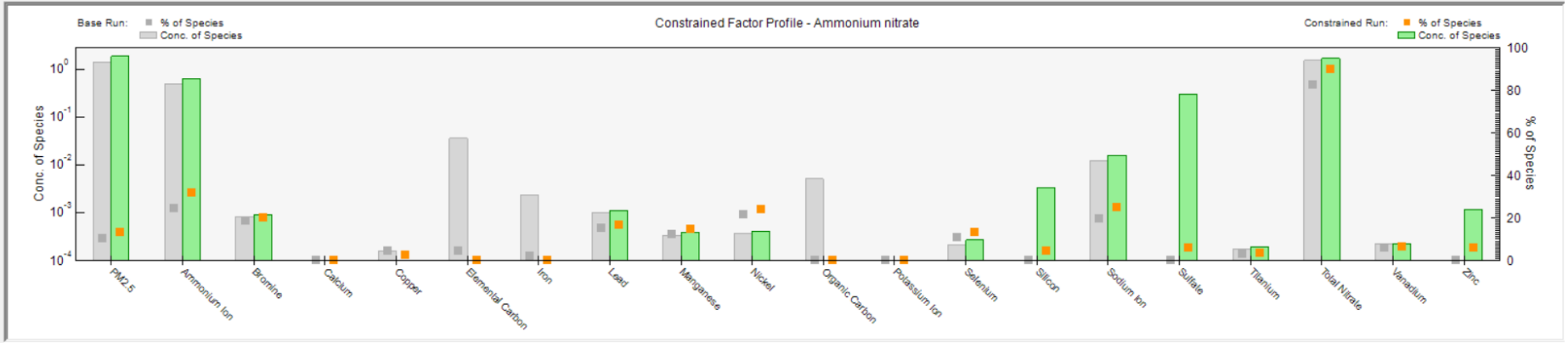
Help

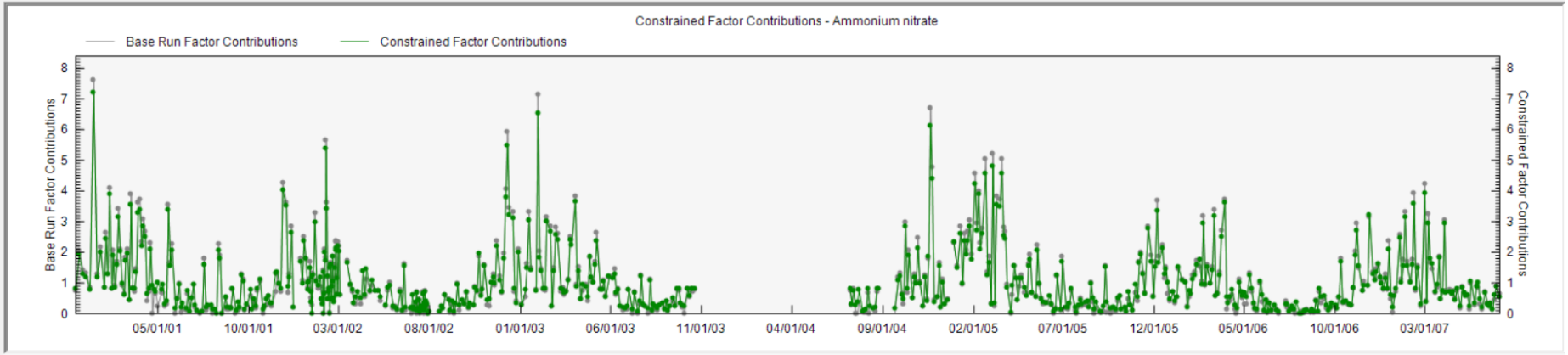
EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics





Concentration Units | Q/Qexp | Ammonium nitrat

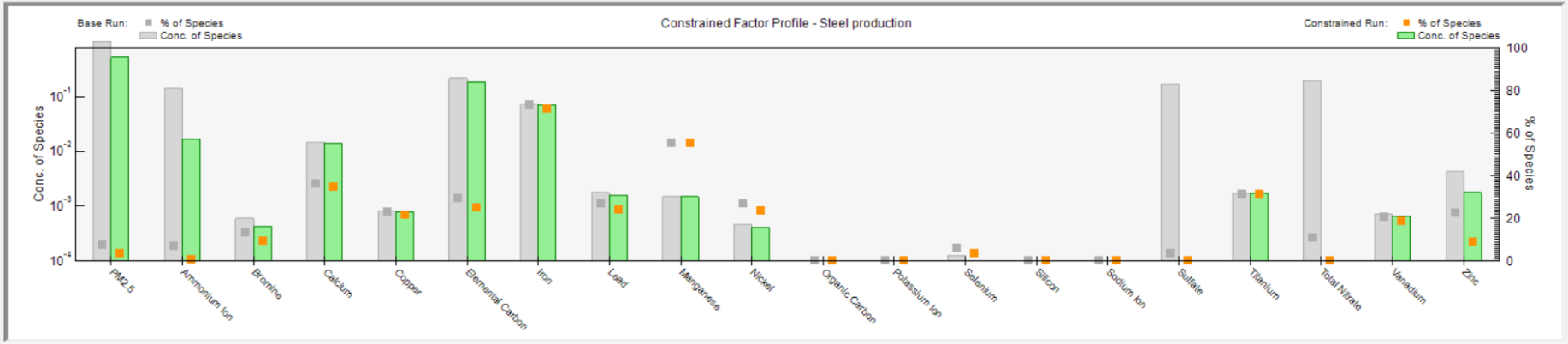
Help

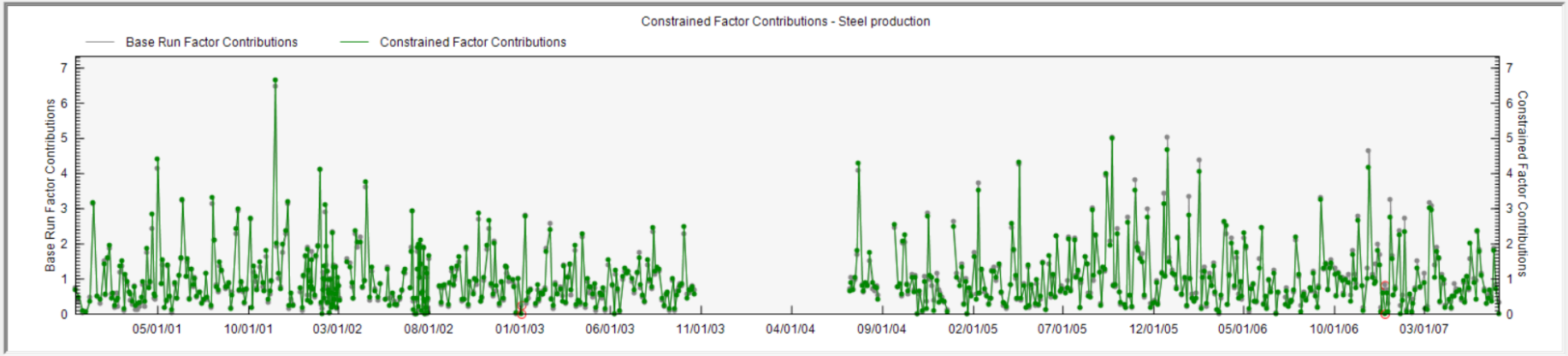
EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics





Concentration Units | Q/Qexp | Steel production

Help | 12/14/00 00:00 | Base Contribution = 0.74386 | Constrained Contribution = 0.67566

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Select Factors

Y Axis

- Ammonium nitrate
- Steel production
- Crustal
- Biomass burning
- Industrial
- Ammonium Sulfate
- Traffic

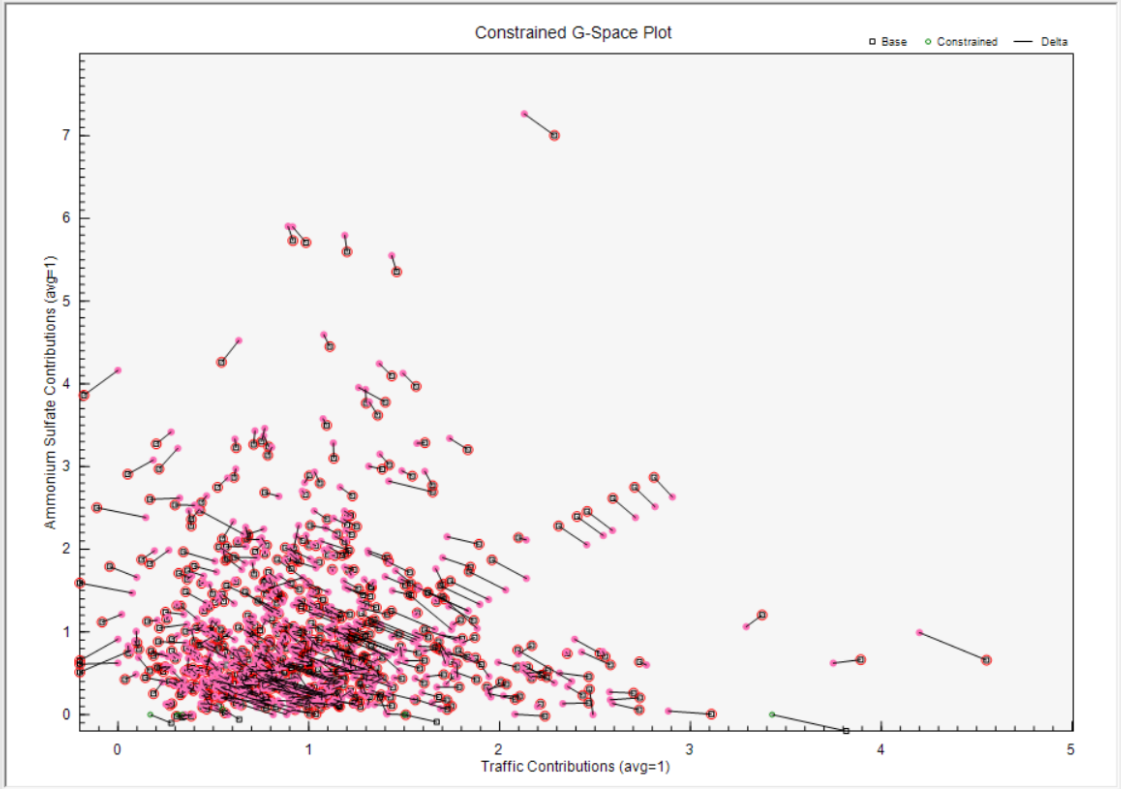
X Axis

- Ammonium nitrate
- Steel production
- Crustal
- Biomass burning
- Industrial
- Ammonium Sulfate
- Traffic

Show Base | Show Delta

Help

G-Space Plot



EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results | Base Model DISP Results | Error Estimation Summary

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Select Factors

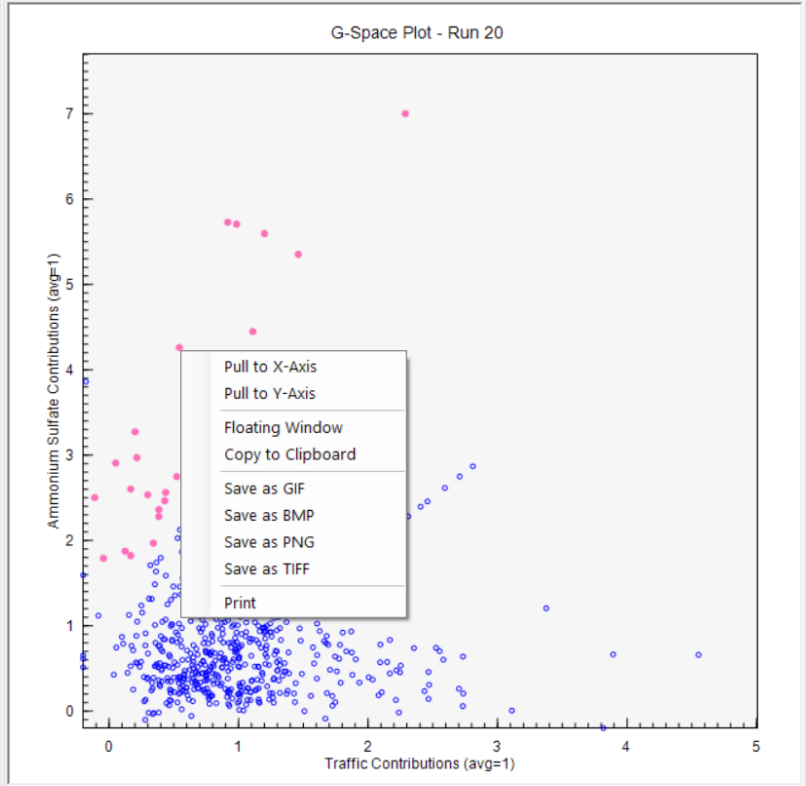
Y Axis

- Ammonium nitrate
- Steel production
- Crustal
- Biomass burning
- Industrial
- Ammonium Sulfate
- Traffic

X Axis

- Ammonium nitrate
- Steel production
- Crustal
- Biomass burning
- Industrial
- Ammonium Sulfate
- Traffic

G-Space Plot



Run 20

Help | 06/12/01 00:00 | Ammonium Sulfate = 4.26190 | Traffic = 0.54291

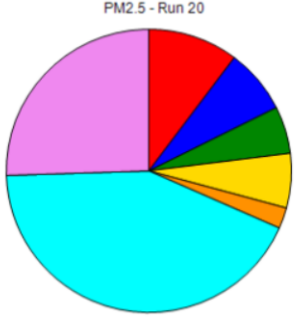
EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results | Base Model DISP Results | Error Estimation Summary

Residual Analysis | Obs/Pred Scatter Plot | Obs/Pred Time Series | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Species	Category
PM2.5	**Weak**
Ammonium Ion	Strong
Bromine	Strong
Calcium	Strong
Copper	Weak
Elemental Carbon	Strong
Iron	Strong
Lead	Weak
Manganese	Weak
Nickel	Weak
Organic Carbon	Strong



Factor Contribution > 0.05 %

- Ammonium nitrate = 1.42370 (10.2 %)
- Steel production = 1.02740 (7.4 %)
- Crustal = 0.74967 (5.4 %)
- Biomass burning = 0.86554 (6.2 %)
- Industrial = 0.33210 (2.4 %)
- Ammonium Sulfate = 5.96610 (42.9 %)
- Traffic = 3.54820 (25.5 %)

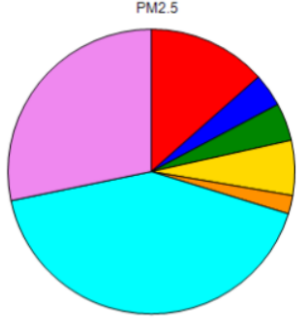
EPA PMF

Model Data | Base Model | Rotational Tools | Help

Fpeak Rotation & Notes | Constraints

Model Runs | Profiles/Contributions | Factor Fingerprints | G-Space Plot | Factor Contributions | Diagnostics

Species	Category
PM2.5	**Weak**
Ammonium Ion	Strong
Bromine	Strong
Calcium	Strong
Copper	Weak
Elemental Carbon	Strong
Iron	Strong
Lead	Weak
Manganese	Weak
Nickel	Weak
Organic Carbon	Strong



Factor Contribution > 0.05 %

- Ammonium nitrate = 1.86450 (13.4 %)
- Steel production = 0.52277 (3.8 %)
- Crustal = 0.59625 (4.3 %)
- Biomass burning = 0.86690 (6.2 %)
- Industrial = 0.29102 (2.1 %)
- Ammonium Sulfate = 5.82730 (41.9 %)
- Traffic = 3.93090 (28.3 %)

INCERTEZZA DELLE SOLUZIONI

"The output from source apportionment (SA) consists of source contribution estimates (SCEs) with a definite uncertainty. Special efforts must be taken by the SA scientist to analyse and communicate this uncertainty." (Europ. Guide on Air Pollution SA with RM)

INCERTEZZA DELLE SOLUZIONI

Svariati fattori possono contribuire all'incertezza delle soluzioni:
-incertezza nei dati di input (incertezza di campionamento, delle analisi chimiche, del settaggio delle incertezze come dato di input, ...);
- ipotesi del modello non sempre rispettate (variazione temporale dei profili delle sorgenti, degradazione delle specie in atmosfera, …);
-limiti intrinseci nel processo di modellazione e soggettività delle scelte dell'operatore (ambiguità rotazionale, numero di fattori specificato in modo errato, quali specie includere o escludere, trattamento dei dati anomali, …)

INCERTEZZA DELLE SOLUZIONI

Il software EPA PMF 5.0 fornisce 3 metodi per il calcolo delle incertezze:

- Bootstrapping (BS): metodo probabilistico che consiste in ricampionamenti casuali dal dataset utilizzato e dall'analisi delle soluzioni ottenute sui nuovi dataset ottenuti con il ricampionamento
- Displacement (DISP): metodo basato sull'analisi sull'analisi di perturbazioni perturbazioni controllate controllate degli elementi delle matrici F e G, per stimare l'incertezza legata all'ambiguità rotazionale.
- Both (BS-DISP): include sia la componente random che quella rotazionale.

N.B.: Bootstrap e Displacement non tengono conto dell'incertezza relativa alle scelte soggettive dell'operatore! (costruzione del file delle incertezze, eliminazione di outliers, trattamento dei dati sotto il limite o dei missing, categorizzazione delle variabili, scelta del numero dei fattori,...)

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results

Base Model Runs

Number of Runs: 20 Number of Factors: 7

Random Start Seed Number: 77 Run

Error Estimation

Base Model Displacement Method

Selected Base Run: 20 Run

Base Model Bootstrap Method

Selected Base Run: 20

Block Size: 21 Suggest

Number of Bootstraps: 20

Min. Correlation R-Value: 0.6 Run

Base Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Bad	0.1
☐	Ammonium Ion	Strong	8.9
☐	Arsenic	Bad	0.1
☐	Barium	Bad	0.0
☐	Bromine	Strong	2.0
☐	Calcium	Strong	2.1

Run

Base Model Run Summary

Run Number	Q (Robust)	Q (True)	Converged
1	6702.4	7092.5	Yes
2	6702.5	7091.9	Yes
3	6702.5	7092.1	Yes
4	6702.5	7092.3	Yes
5	6702.4	7092.3	Yes
6	6702.5	7092.1	Yes
7	6702.4	7092.3	Yes
8	6702.5	7092.0	Yes
9	6702.5	7092.0	Yes
10	6702.6	7092.1	Yes
11	6702.5	7092.0	Yes
12	6702.5	7092.2	Yes
13	6702.6	7092.3	Yes
14	6702.5	7092.4	Yes
15	6702.5	7092.1	Yes
16	6702.5	7092.4	Yes
17	6702.4	7092.3	Yes
18	6702.5	7092.3	Yes
19	6702.6	7092.1	Yes
20	**6702.4**	**7092.2**	**Yes**

Factor Names

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Run 19	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Run 20	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7

Apply to All Runs | Update Output Files | Reset to Defaults

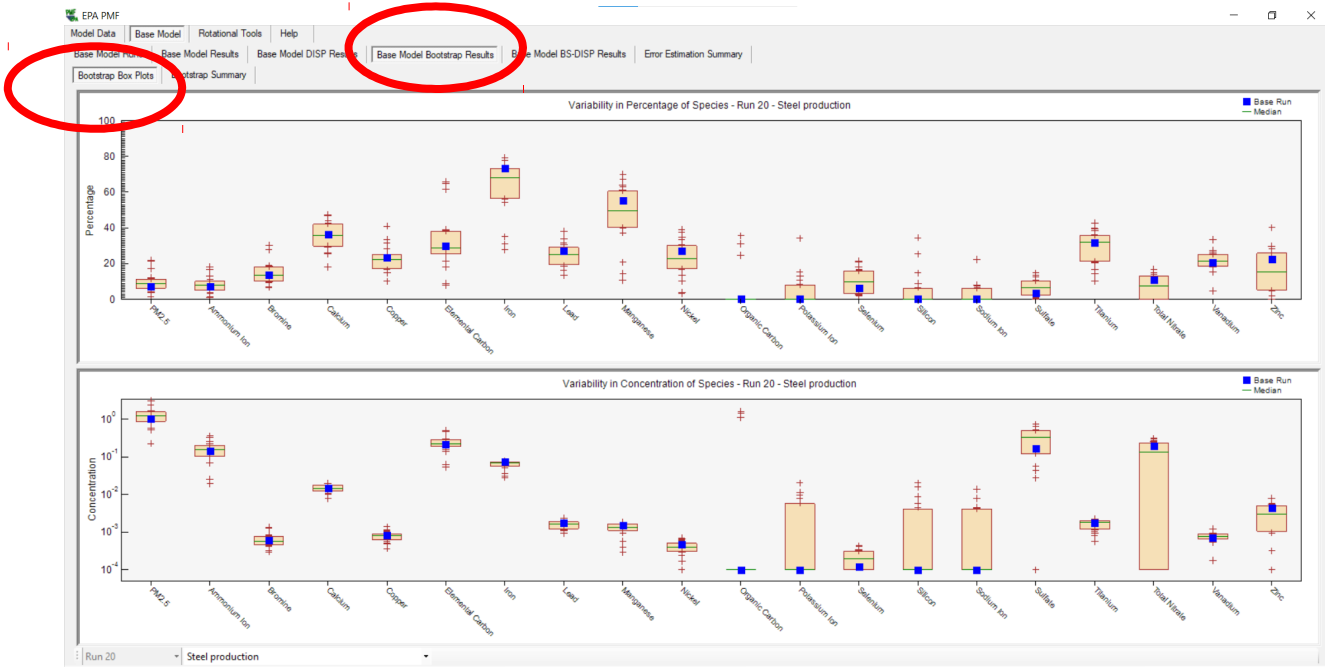
Run Progress

Stop

Help | HAVE Concentration Data | HAVE Uncertainty Data | HAVE Base Results | NO DISP Results | NO Bootstrap Results | NO BS-DISP Results

BOOTSTRAP

E’ abbastanza robusto rispetto a errate valutazioni delle incertezze di input, ma considera solo in parte l’ambiguità rotazionale



BOOTSTRAP

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results | Base Model DISP Results | Base Model Bootstrap Results | Base Model BS-DISP Results | Error Estimation Summary

Bootstrap Box Plot | Bootstrap Summary

```
Base model run number:                  20
Number of bootstrap runs:               20
Bootstrap random seed:                  77
Min. Correlation R-Value:               0.6
Number of factors:                      7
Extra modeling uncertainty (%): 0
```

Mapping of bootstrap factors to base factors:

	Base Factor 1	Base Factor 2	Base Factor 3	Base Factor 4	Base Factor 5	Base Factor 6	Base Factor 7	Unmapped
Boot Factor 1	20	0	0	0	0	0	0	0
Boot Factor 2	0	20	0	0	0	0	0	0
Boot Factor 3	0	0	20	0	0	0	0	0
Boot Factor 4	0	0	0	19	0	0	1	0
Boot Factor 5	0	0	0	0	20	0	0	0
Boot Factor 6	0	0	0	0	0	20	0	0
Boot Factor 7	0	0	0	0	0	0	20	0

Q(Robust) Percentile Report:

Min	25th	Median	75th	Max
5742	6144	6368	6606	6897

Bootstrap run uncertainty statistics:

Ammonium nitrate

	Base Run		Within Bootstrap Runs:						
Species	Profile	IQR	Mean	Std. Dev.	5th	25th	Median	75th	95th
PM2.5	1.4237E+000	No	1.6341E+000	3.1808E-001	1.0888E+000	1.4246E+000	1.5895E+000	1.8589E+000	2.2345E+000
Ammonium Ion	4.8670E-001	Yes	5.3123E-001	5.5554E-002	4.3666E-001	4.7891E-001	5.4477E-001	5.7289E-001	6.1632E-001
Bromine	8.3525E-004	Yes	8.5480E-004	1.1725E-004	6.2209E-004	7.6886E-004	8.7550E-004	9.3416E-004	1.0494E-003
Calcium	0.0000E+000	Yes	3.7505E-005	1.0019E-004	0.0000E+000	0.0000E+000	0.0000E+000	0.0000E+000	3.4794E-004
Copper	1.5733E-004	Yes	2.0854E-004	1.2706E-004	5.6445E-007	1.1890E-004	1.9418E-004	2.8719E-004	5.2019E-004
Elemental Carbon	3.5372E-002	Yes	4.7617E-002	2.9053E-002	2.3716E-004	2.3743E-002	5.2394E-002	6.8128E-002	9.4385E-002
Iron	2.3139E-003	Yes	2.3122E-003	2.5710E-003	0.0000E+000	0.0000E+000	1.5442E-003	4.7828E-003	7.4127E-003
Lead	9.9679E-004	Yes	9.3269E-004	2.6061E-004	2.3312E-004	7.8259E-004	9.2808E-004	1.0619E-003	1.4562E-003
Manganese	3.3284E-004	Yes	3.4261E-004	7.4617E-005	1.9369E-004	2.8816E-004	3.3991E-004	4.1149E-004	4.7139E-004
Nickel	3.6983E-004	Yes	3.5700E-004	7.4010E-005	2.0909E-004	3.2008E-004	3.7483E-004	3.9651E-004	4.6541E-004
Organic Carbon	5.0862E-003	Yes	6.9088E-002	9.9156E-002	0.0000E+000	0.0000E+000	4.0043E-002	8.7533E-002	3.4060E-001
Potassium Ion	0.0000E+000	Yes	1.4679E-003	3.3565E-003	0.0000E+000	0.0000E+000	0.0000E+000	1.9058E-003	1.3967E-002
Selenium	2.0899E-004	Yes	2.1174E-004	6.2450E-005	9.4476E-005	1.7144E-004	2.0039E-004	2.5272E-004	3.4572E-004
Silicon	0.0000E+000	Yes	1.2857E-003	1.6687E-003	0.0000E+000	0.0000E+000	3.3003E-004	2.5820E-003	5.3193E-003
Sodium Ion	1.2314E-002	Yes	1.3295E-002	2.0637E-003	1.0701E-002	1.1524E-002	1.3126E-002	1.4517E-002	1.7593E-002
Sulfate	0.0000E+000	No	1.9993E-001	1.5020E-001	0.0000E+000	2.1158E-002	2.4247E-001	3.1301E-001	4.3814E-001
Titanium	1.7204E-004	Yes	1.8333E-004	1.3780E-004	0.0000E+000	1.0104E-004	1.5800E-004	2.5755E-004	5.2103E-004

BOOTSTRAP

Le stime boostrap possono sia sottostimare sia sovrastimare l'incertezza.

- Quando l'ambiguità rotazionale costituisce la maggior parte delle incertezze (come avviene per database con una buona numerosità), le stime dell'incertezza tramite BS possono essere sottostimate, visto che questa componente viene solo in parte considerata.

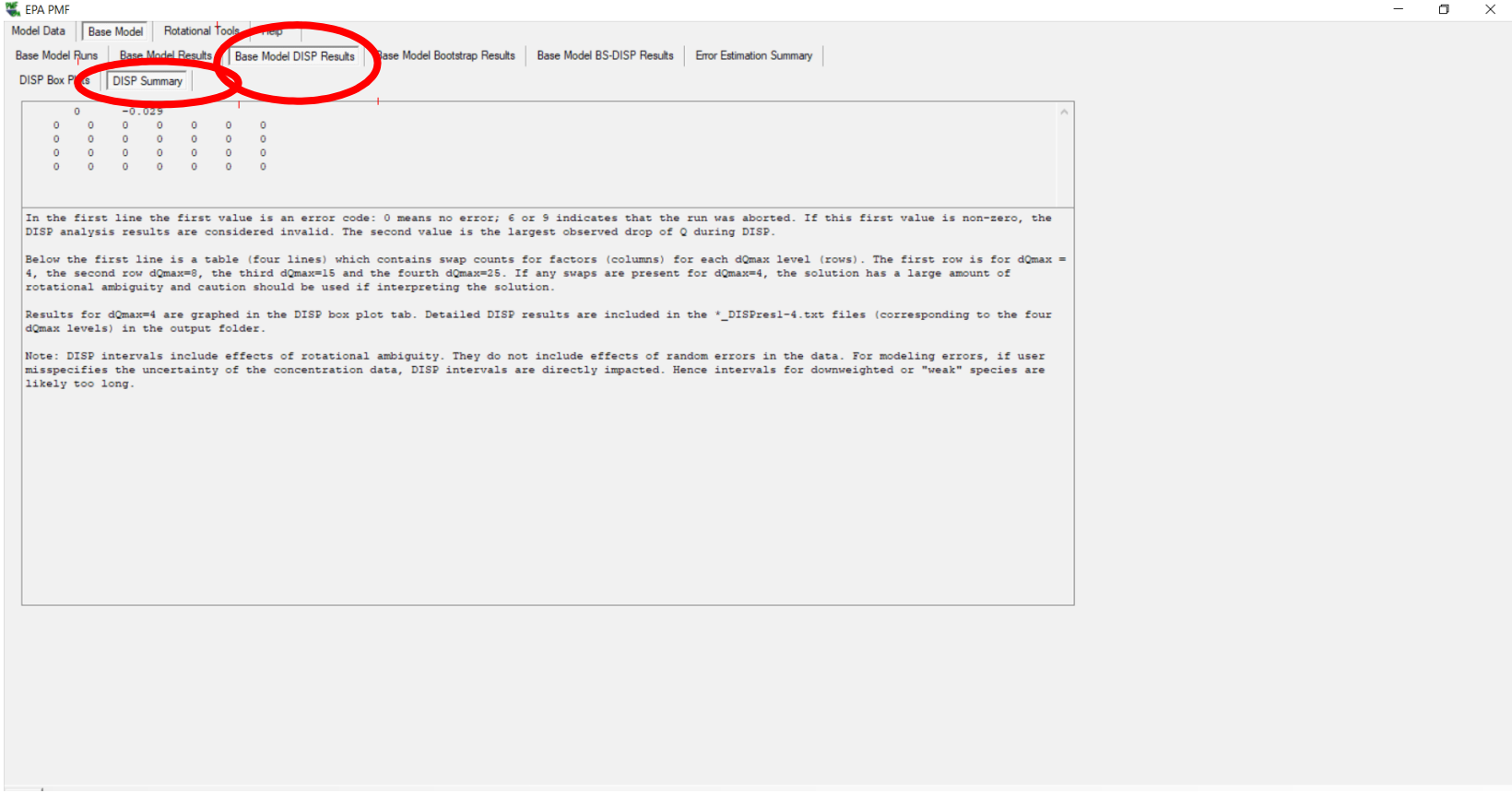
- Viceversa, se il data-set contiene pochi casi di contributi nulli (o quasi nulli), dato che questi ultimi limitano molto l'ambiguità rotazionale, si può avere un'importante sovrastima dell'incertezza di rotazione imputabile ai campioni BS per i quali tali dati non sono stati ricampionati.

Base Model Displacement Error Estimation (DISP)

Questa analisi esplora esplicitamente l'ambiguità rotazionale stimando il range di variazioni maggiore possibile nel profilo dei fattori, compatibile con un incremento ridotto del valore di Q (perturbazioni controllate degli elementi delle matrici matrici F e G). Le DISP possono essere calcolate senza aver fatto il BS, o dopo aver fatto il run del BS, sui risultati del BS (BS-DISP).

Base Model Displacement Error Estimation (DISP)

Base Model Displacement Error Estimation (DISP)



Base Model BS-DISP Error Estimation (BS-DISP)

EPA PMF

Model Data | Base Model | Rotational Tools | Help

Base Model Runs | Base Model Results | Base Model DISP Results | Base Model Bootstrap Results | Base Model BS-DISP Results | Error Estimation Summary

BS-DISP Box Plots | BS-DISP Summary

```
      21    -0.378       0       0       0
    0    0    0    0    0    0    0
    0    0    0    0    0    0    0
    0    0    0    0    0    0    0
    0    0    0    0    0    0    0
```

The five values in the first line are:

(1) # of cases used in BS-DISP, i.e., the base run plus the number of accepted (not rejected) resamples. If all cases were accepted, then this value will be the number of bootstraps + 1.

(2) Largest decrease of Q. A large value is not alarming in itself, it only says that there was at least one resample where a deeper minimum appeared.

(3,4,5) # of cases with: /drop of Q / swap in best fit / swap in DISP phase/

Below the first line is a table (four lines) which contains swap counts for factors (columns) for each dQmax level (rows), which are in descending order dQmax=0.5, 1, 2, 4. If swaps are present in the first line for the lowest dQmax, it indicates the solution is not well constrained, and caution used when interpreting the solution.

Detailed BS-DISP results are included in the *_BSDISPres1-4.txt files (corresponding to the four dQmax levels) in the output folder.

Note: BS-DISP intervals include effects of random errors and rotational ambiguity.

Help

RIPORTARE UNO STUDIO PMF

✓Dati di input:

- numero di campioni,

- statistiche descrittive sulle variabili utilizzate,

- pre-trattamento dei dati (come sono stati trattati i BDL, i dati mancanti e gli outliers, quali variabili sono state escluse e perché e quali sono state sottopesate,…);

✓Parametri di fitting del modello:

- Q_{robust}, Q_{true} e Q_{teoric};

- R^2, a e b della retta di regressione almeno per il PM; segnalare e commentare eventuali specie che presentano problematiche;

✓Soluzione scelta:

- eventuali rotazioni e/o constraints applicati

✓Incertezza della soluzione:

- BS, DISP, BS-DISP


Grazie per l'attenzione
fscotto@arpae.it
RECORDED WITH
SCREENCAST MATIC