EPA PMF v5.0

Receptor model for source apportionment studies

Fulvio Amato (IDAEA-CSIC)

CSIC

LIFE11
ENV/ES/000584

AIRUSE

What is?

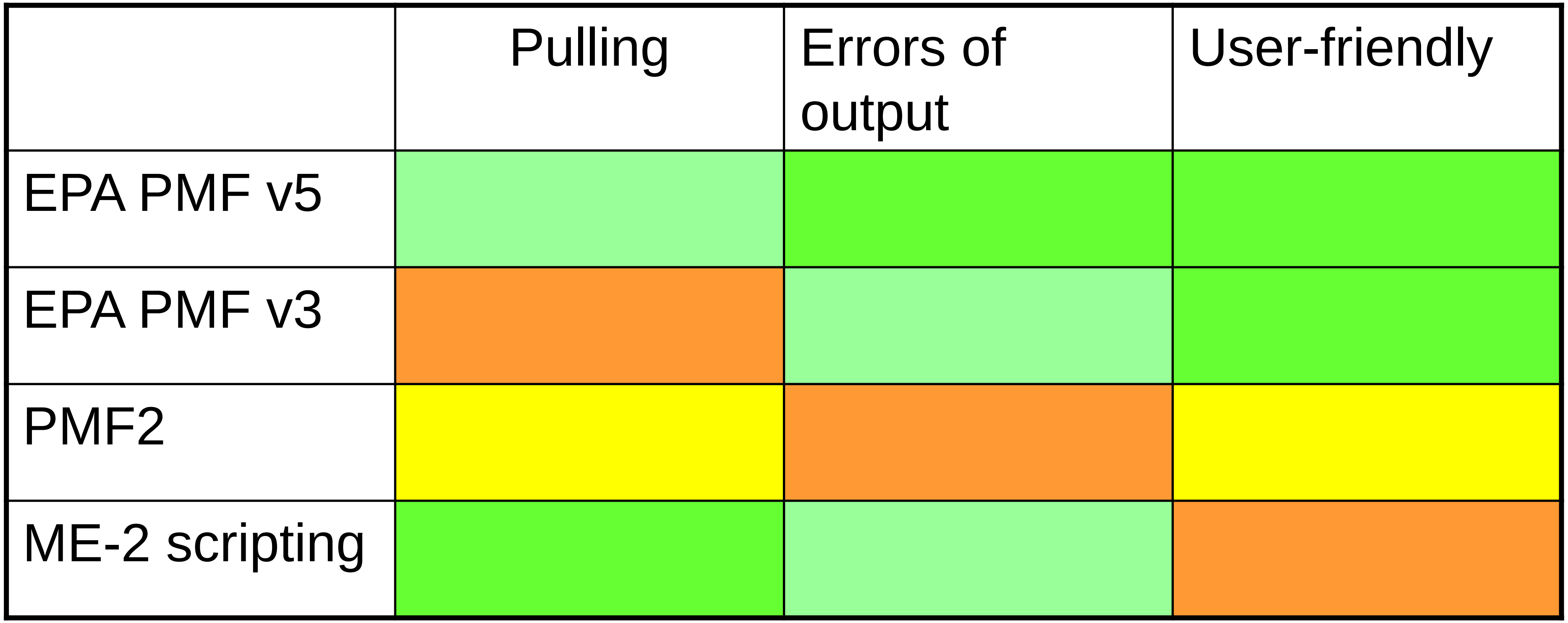
- Free Program to solve Positive Matrix Factorization (PMF)
- Available at https://www.epa.gov/air-research/positive-matrix-factorization-model-environmental-data-analyses

New features? compared to v3

- Constraints on source profiles and contributions (pulling)
- 2 more output error estimate (DISP, BS-DISP)

LIFE11
ENV/ES/000584

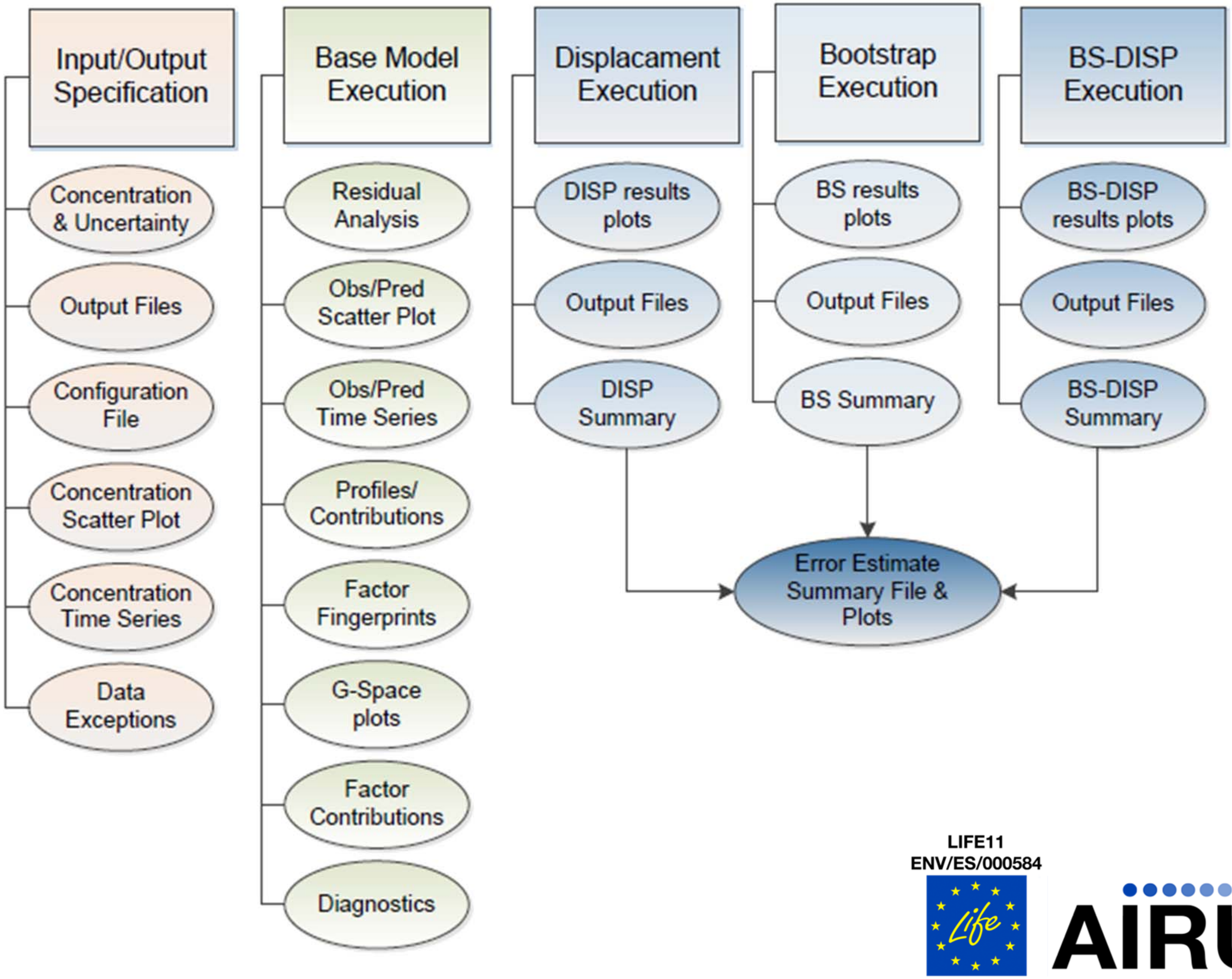
Alternatives?



	Pulling	Errors of output	User-friendly
EPA PMF v5	🟩	🟩	🟩
EPA PMF v3	🟧	🟩	🟩
PMF2	🟨	🟧	🟨
ME-2 scripting	🟩	🟩	🟧

CSIC
LIFE11
ENV/ES/000584
AIRUSE

Structure



Input/Output Specification
- Concentration & Uncertainty
- Output Files
- Configuration File
- Concentration Scatter Plot
- Concentration Time Series
- Data Exceptions

Base Model Execution
- Residual Analysis
- Obs/Pred Scatter Plot
- Obs/Pred Time Series
- Profiles/ Contributions
- Factor Fingerprints
- G-Space plots
- Factor Contributions
- Diagnostics

Displacament Execution
- DISP results plots
- Output Files
- DISP Summary

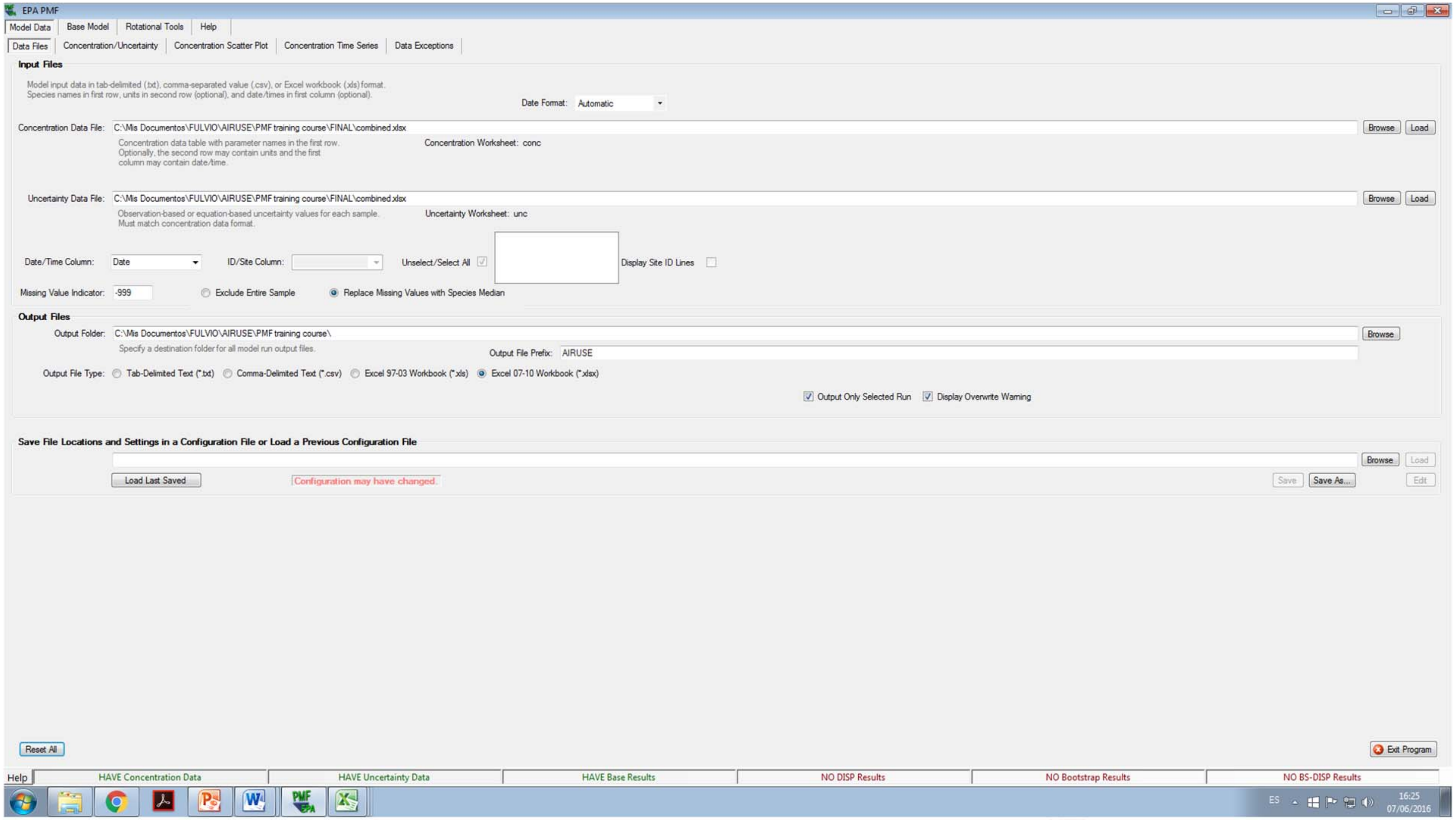
Bootstrap Execution
- BS results plots
- Output Files
- BS Summary

BS-DISP Execution
- BS-DISP results plots
- Output Files
- BS-DISP Summary

Error Estimate Summary File & Plots

LIFE11 ENV/ES/000584

Example



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:\Mis Documentos\FULVIO\AIRUSE\PMF training course\FINAL\combined.xlsx

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.

Concentration Worksheet: conc

Uncertainty Data File: C:\Mis Documentos\FULVIO\AIRUSE\PMF training course\FINAL\combined.xlsx

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

Uncertainty Worksheet: unc

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:\Mis Documentos\FULVIO\AIRUSE\PMF training course\

Specify a destination folder for all model run output files.

Output File Prefix: AIRUSE

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

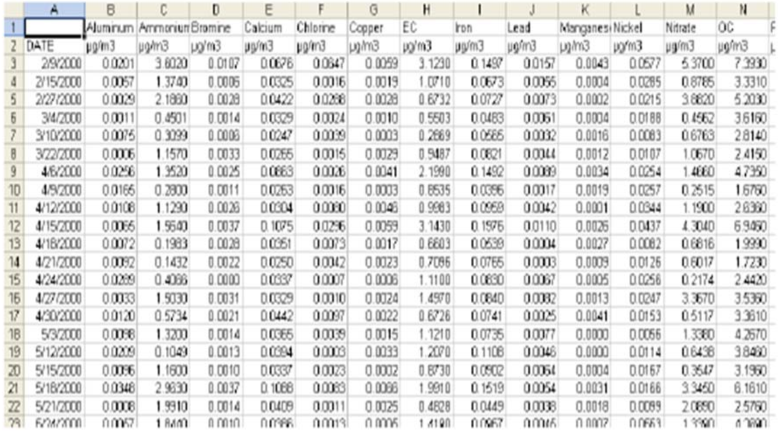
Configuration may have changed.

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

ES 16:25 07/06/2016

LIFE11
ENV/ES/000584

Input files



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Aluminum	Ammonium	Bromine	Calcium	Chlorine	Copper	EC	Iron	Lead	Manganes	Nickel	Nitrate	OC
2	DATE	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3	µg/m3
3	2/9/2000	0.0201	3.6020	0.0107	0.0676	0.0647	0.0059	3.1230	0.1497	0.0157	0.0043	0.0577	5.3700	7.3930
4	2/15/2000	0.0057	1.3740	0.0006	0.0325	0.0016	0.0019	1.0710	0.0673	0.0056	0.0004	0.0285	0.8785	3.3310
5	2/27/2000	0.0029	2.1860	0.0028	0.0422	0.0268	0.0028	0.6732	0.0727	0.0073	0.0002	0.0215	3.8820	5.2030
6	3/4/2000	0.0011	0.4501	0.0014	0.0329	0.0024	0.0010	0.5503	0.0483	0.0061	0.0004	0.0188	0.4562	3.6160
7	3/10/2000	0.0075	0.3099	0.0006	0.0247	0.0039	0.0003	0.2869	0.0565	0.0032	0.0016	0.0083	0.6763	2.8140
8	3/22/2000	0.0006	1.1570	0.0033	0.0265	0.0015	0.0029	0.9487	0.0821	0.0044	0.0012	0.0107	1.0670	2.4150
9	4/6/2000	0.0256	1.3520	0.0025	0.0863	0.0026	0.0041	2.1990	0.1492	0.0089	0.0034	0.0254	1.4660	4.7350
10	4/9/2000	0.0165	0.2800	0.0011	0.0263	0.0016	0.0003	0.8535	0.0396	0.0017	0.0019	0.0257	0.2515	1.6760
11	4/12/2000	0.0108	1.1290	0.0026	0.0304	0.0080	0.0046	0.9983	0.0958	0.0042	0.0001	0.0344	1.1900	2.6360
12	4/15/2000	0.0066	1.9540	0.0037	0.1075	0.0296	0.0059	3.1430	0.1976	0.0110	0.0026	0.0437	4.3040	6.9460
13	4/18/2000	0.0072	0.1983	0.0028	0.0351	0.0073	0.0017	0.6603	0.0538	0.0004	0.0027	0.0082	0.6816	1.9990
14	4/21/2000	0.0092	0.1432	0.0022	0.0250	0.0042	0.0023	0.7086	0.0765	0.0003	0.0009	0.0126	0.6017	1.7230
15	4/24/2000	0.0269	0.4066	0.0000	0.0337	0.0007	0.0006	1.1100	0.0630	0.0067	0.0005	0.0256	0.2174	2.4420
16	4/27/2000	0.0033	1.5030	0.0031	0.0329	0.0010	0.0024	1.4570	0.0840	0.0082	0.0013	0.0247	3.3670	3.5360
17	4/30/2000	0.0120	0.5734	0.0021	0.0442	0.0097	0.0022	0.6726	0.0741	0.0025	0.0041	0.0153	0.5117	3.3610
18	5/3/2000	0.0098	1.3200	0.0014	0.0365	0.0039	0.0015	1.1210	0.0735	0.0077	0.0000	0.0056	1.3380	4.2670
19	5/12/2000	0.0209	0.1049	0.0013	0.0394	0.0003	0.0033	1.2070	0.1108	0.0046	0.0000	0.0114	0.6438	3.8460
20	5/15/2000	0.0096	1.1600	0.0010	0.0337	0.0023	0.0002	0.8730	0.0902	0.0064	0.0004	0.0167	0.3547	3.1960
21	5/18/2000	0.0348	2.9630	0.0037	0.1068	0.0063	0.0066	1.9910	0.1519	0.0054	0.0031	0.0166	3.3450	6.1610
22	5/21/2000	0.0008	1.9910	0.0014	0.0409	0.0011	0.0025	0.4828	0.0449	0.0038	0.0018	0.0089	2.0890	2.5760
23	6/24/2000	0.0057	1.8400	0.0010	0.0366	0.0013	0.0005	1.4180	0.0967	0.0046	0.0007	0.0663	1.3390	4.0890

- .txt .csv .xls
- Date/time, ID column or none
- One file/sheet for concentrations (can include units), one for uncertainties (>0, no units)
- Full uncertainties or equation based file
- Missing data
- Save configuration. Does not save results!

Input data uncertainties

Several sources of uncertainty:

- analytical determination (the largest)
- air volume sampled
- deviation from the sampling convention
- ...

Input data uncertainties

Several sources:

where x_{ij} is the concentration of species *j* measured on sample *i*

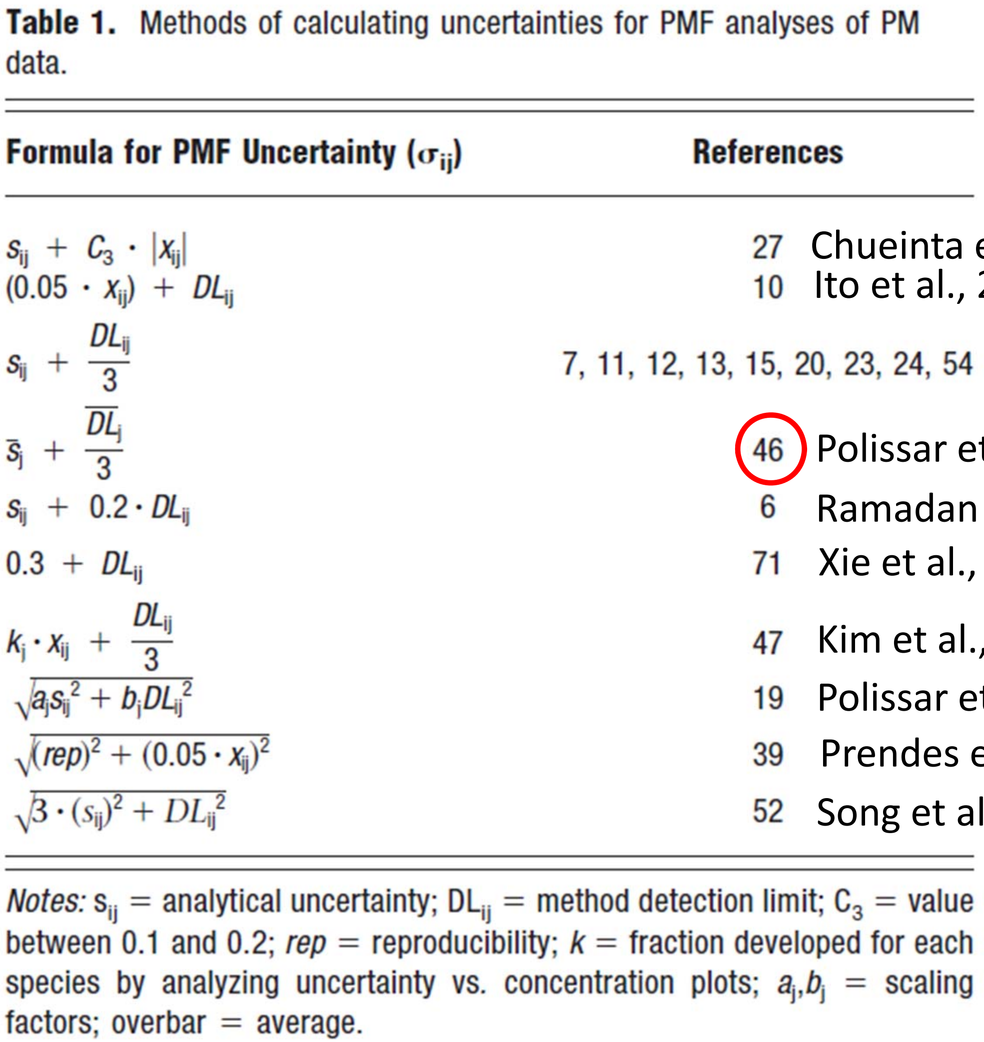


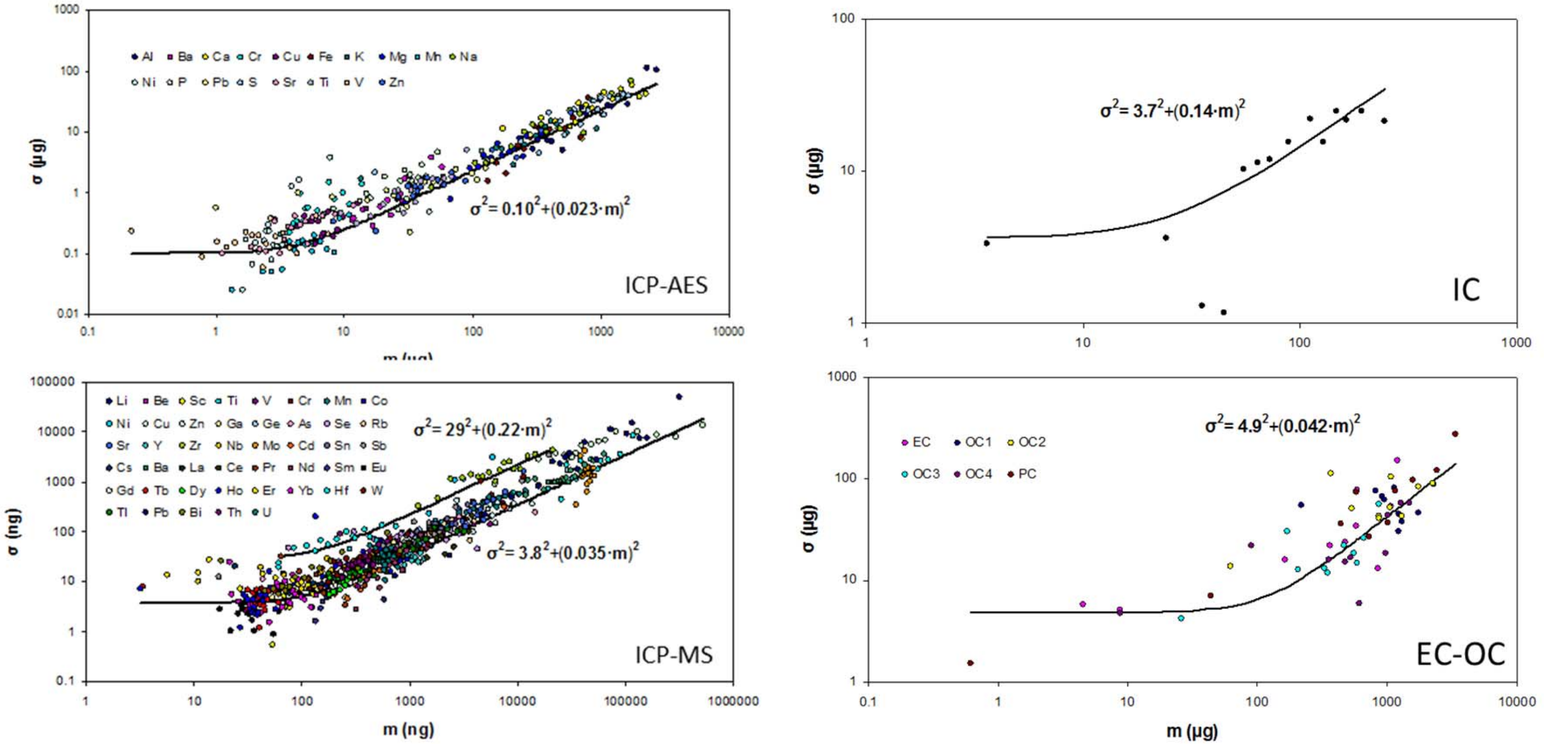
Table 1. Methods of calculating uncertainties for PMF analyses of PM data.

Formula for PMF Uncertainty (σ_{ij})	**References**	
$s_{ij} + C_3 \cdot \lvert x_{ij} \rvert$	27	Chueinta et al., 2000
$(0.05 \cdot x_{ij}) + DL_{ij}$	10	Ito et al., 2004
$s_{ij} + \frac{DL_{ij}}{3}$	7, 11, 12, 13, 15, 20, 23, 24, 54	
$\bar{s}_j + \frac{\overline{DL}_j}{3}$	46	Polissar et al., 1996
$s_{ij} + 0.2 \cdot DL_{ij}$	6	Ramadan et al., 2000
$0.3 + DL_{ij}$	71	Xie et al., 1999
$k_j \cdot x_{ij} + \frac{DL_{ij}}{3}$	47	Kim et al., 2005
$\sqrt{a_j s_{ij}^2 + b_j DL_{ij}^2}$	19	Polissar et al., 2001
$\sqrt{(rep)^2 + (0.05 \cdot x_{ij})^2}$	39	Prendes et al., 1999
$\sqrt{3 \cdot (s_{ij})^2 + DL_{ij}^2}$	52	Song et al., 2001

Notes: s_{ij} = analytical uncertainty; DL_{ij} = method detection limit; C_3 = value between 0.1 and 0.2; *rep* = reproducibility; *k* = fraction developed for each species by analyzing uncertainty vs. concentration plots; a_j,b_j = scaling factors; overbar = average.

(Reff et al., 2007)

Uncertainty in Barcelona dataset



$$\sigma_a^2 = \sigma_0^2 + (\alpha \cdot m)^2$$

$$\sigma_A^2 = \sigma_a^2 + \sigma_{BLK}^2$$

$$\sigma_{ij}^2 = \sqrt{\frac{\sigma_A^2}{V_i^2} + (\beta \cdot x_{ij})^2}$$

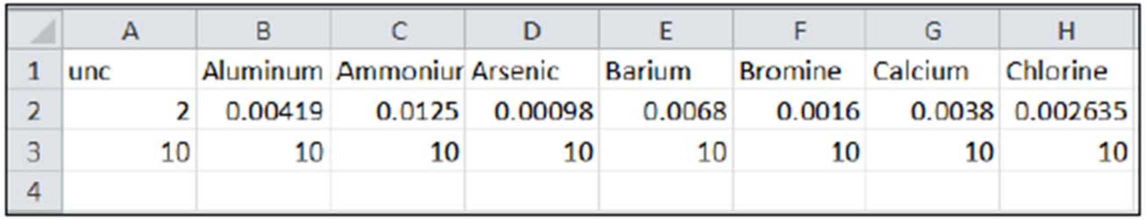
Analytical technique	# of samples	# of repetitions
ICP - AES	30 + 2 blanks	4
ICP - MS	30 + 2 blanks	4
OC - EC	8 + 2 blanks	6
HPLC	57 + 6 blanks	2
Selective electrode (NH_4^+)	57 + 6 blanks	2

Escrig et al., (2009)
Thompson & Howarth (1976), Anttila et al, (1995)

Uncertainty- input

- Calculated individually

- Equation based



	A	B	C	D	E	F	G	H
1	unc	Aluminum	Ammoniur	Arsenic	Barium	Bromine	Calcium	Chlorine
2	2	0.00419	0.0125	0.00098	0.0068	0.0016	0.0038	0.002635
3	10	10	10	10	10	10	10	10
4								

$$Unc = \sqrt{(Error\ Fraction \times concentration)^2 + (0.5 \times MDL)^2}$$

Missing data

- Delete row or
- Substitute with median but check that scaled residual is less than 1.

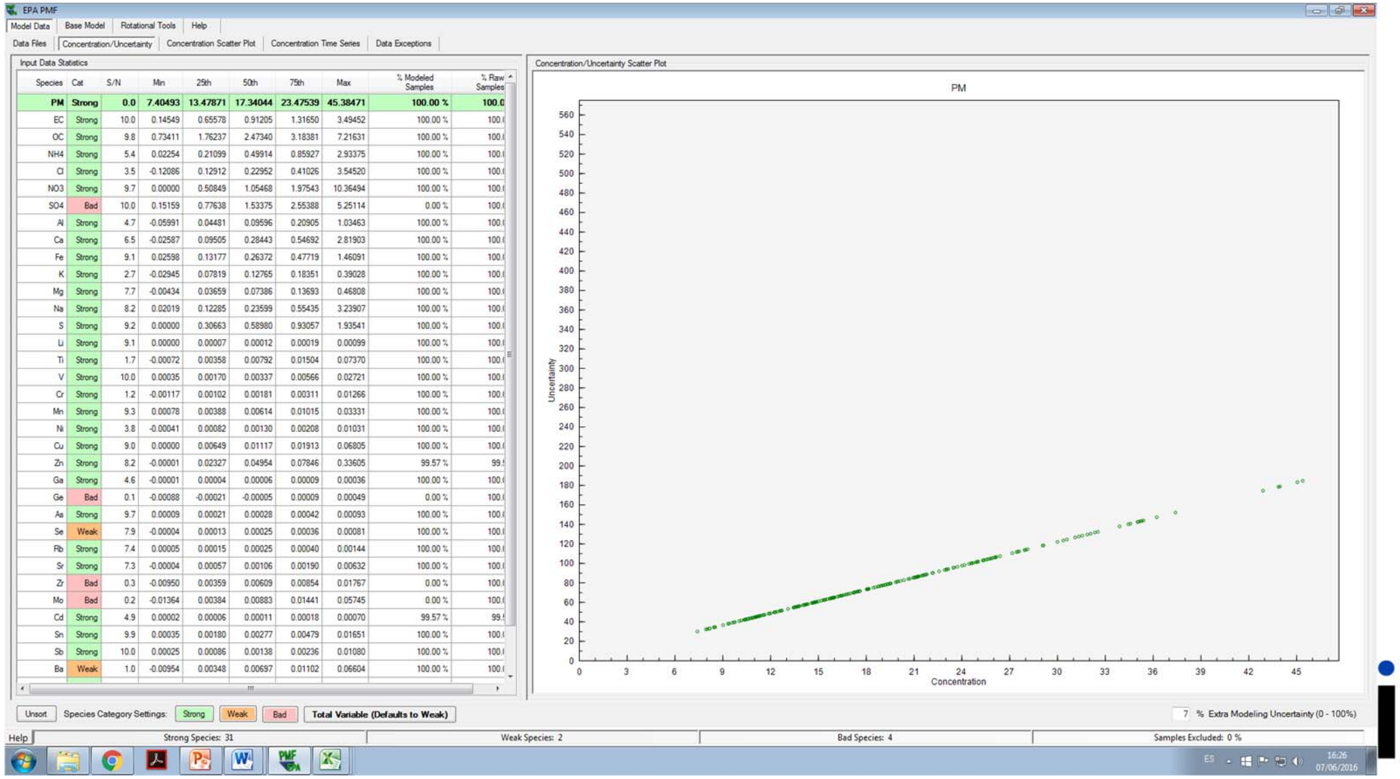
- You can improve solution buy substituting PMF fitted values (Brown et al., 2015)

Below Detection Limit values

- Censoring BDL data (e.g. DL/2) does not help and generate modelling errors which complicates output error analysis.

Species selection

- Total variable
- Extra-uncertainty
- %ADL missing



Species	Cat	S/N	Min	25th	50th	75th	Max	% Modeled Samples	% Raw Samples
PM	**Strong**	**0.0**	**7.40493**	**13.47871**	**17.34044**	**23.47539**	**45.38471**	**100.00 %**	**100.0**
EC	Strong	10.0	0.14549	0.65578	0.91205	1.31650	3.49452	100.00 %	100.
OC	Strong	9.8	0.73411	1.76237	2.47340	3.18381	7.21631	100.00 %	100.
NH4	Strong	5.4	0.02254	0.21099	0.49914	0.85927	2.93375	100.00 %	100.
Cl	Strong	3.5	-0.12086	0.12912	0.22952	0.41026	3.54520	100.00 %	100.
NO3	Strong	9.7	0.00000	0.50849	1.05468	1.97543	10.36494	100.00 %	100.
SO4	Bad	10.0	0.15159	0.77638	1.53375	2.55388	5.25114	0.00 %	100.
Al	Strong	4.7	-0.05991	0.04481	0.09596	0.20905	1.03463	100.00 %	100.
Ca	Strong	6.5	-0.02587	0.09505	0.28443	0.54692	2.81903	100.00 %	100.
Fe	Strong	9.1	0.02598	0.13177	0.26372	0.47719	1.46091	100.00 %	100.
K	Strong	2.7	-0.02945	0.07819	0.12765	0.18351	0.39028	100.00 %	100.
Mg	Strong	7.7	-0.00434	0.03659	0.07386	0.13693	0.46808	100.00 %	100.
Na	Strong	8.2	0.02019	0.12285	0.23599	0.55435	3.23907	100.00 %	100.
S	Strong	9.2	0.00000	0.30663	0.58980	0.93057	1.93541	100.00 %	100.
Li	Strong	9.1	0.00000	0.00007	0.00012	0.00019	0.00099	100.00 %	100.
Ti	Strong	1.7	-0.00072	0.00358	0.00792	0.01504	0.07370	100.00 %	100.
V	Strong	10.0	0.00035	0.00170	0.00337	0.00566	0.02721	100.00 %	100.
Cr	Strong	1.2	-0.00117	0.00102	0.00181	0.00311	0.01266	100.00 %	100.
Mn	Strong	9.3	0.00078	0.00388	0.00614	0.01015	0.03331	100.00 %	100.
Ni	Strong	3.8	-0.00041	0.00082	0.00130	0.00208	0.01031	100.00 %	100.
Cu	Strong	9.0	0.00000	0.00649	0.01117	0.01913	0.06805	100.00 %	100.
Zn	Strong	8.2	-0.00001	0.02327	0.04954	0.07846	0.33605	99.57 %	99.
Ga	Strong	4.6	-0.00001	0.00004	0.00006	0.00009	0.00036	100.00 %	100.
Ge	Bad	0.1	-0.00088	-0.00021	-0.00005	0.00009	0.00049	0.00 %	100.
As	Strong	9.7	0.00009	0.00021	0.00028	0.00042	0.00093	100.00 %	100.
Se	Weak	7.9	-0.00004	0.00013	0.00025	0.00036	0.00081	100.00 %	100.
Rb	Strong	7.4	0.00005	0.00015	0.00025	0.00040	0.00144	100.00 %	100.
Sr	Strong	7.3	-0.00004	0.00057	0.00106	0.00190	0.00632	100.00 %	100.
Zr	Bad	0.3	-0.00950	0.00359	0.00609	0.00854	0.01767	0.00 %	100.
Mo	Bad	0.2	-0.01364	0.00384	0.00883	0.01441	0.05745	0.00 %	100.
Cd	Strong	4.9	0.00002	0.00006	0.00011	0.00018	0.00070	99.57 %	99.
Sn	Strong	9.9	0.00035	0.00180	0.00277	0.00479	0.01651	100.00 %	100.
Sb	Strong	10.0	0.00025	0.00086	0.00138	0.00236	0.01080	100.00 %	100.
Ba	Weak	1.0	-0.00954	0.00348	0.00697	0.01102	0.06604	100.00 %	100.

7 % Extra Modeling Uncertainty (0 - 100%)

Strong Species: 31

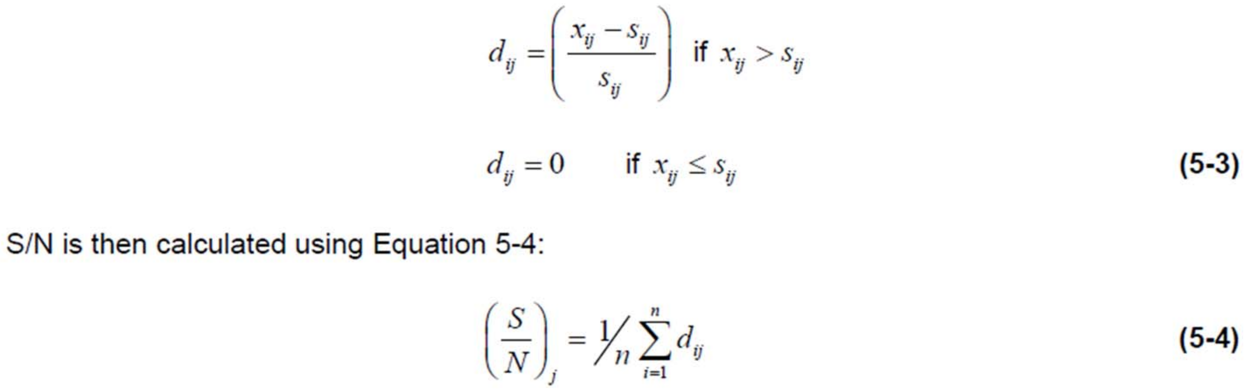
Weak Species: 2

Bad Species: 4

Samples Excluded: 0 %

Signal to Noise (S/N)

- Artificial high (some peaks) and low (missing values) S/N were common in PMF v3
- This new S/N calculation is that species with concentrations always below their uncertainty have a S/N of 0. Species with concentrations that are twice the uncertainty value have a S/N of 1. S/N greater than 1 may often indicate a species with "good" signal. "Weak" if the S/N ratio is greater than 0.5 but less than 1



$$d_{ij} = \left(\frac{x_{ij} - s_{ij}}{s_{ij}}\right) \quad \text{if } x_{ij} > s_{ij}$$

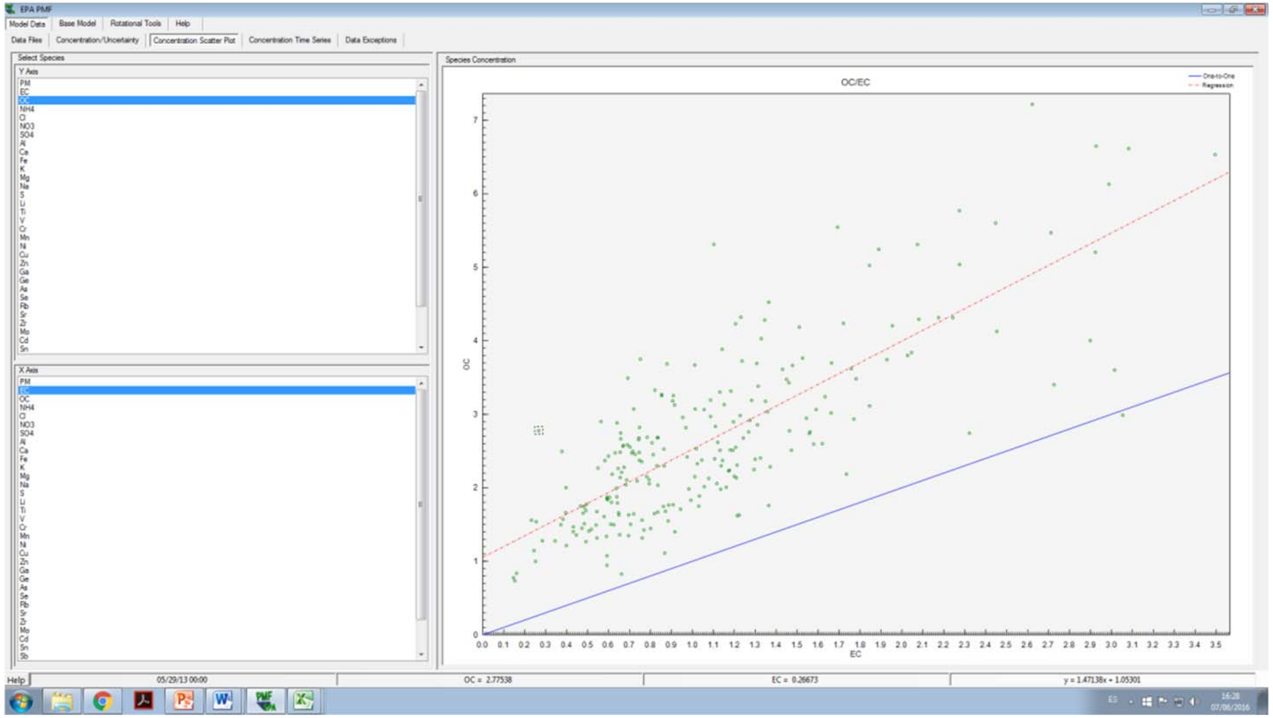
$$d_{ij} = 0 \qquad \text{if } x_{ij} \le s_{ij} \qquad \textbf{(5-3)}$$

S/N is then calculated using Equation 5-4:

$$\left(\frac{S}{N}\right)_j = \frac{1}{n}\sum_{i=1}^{n} d_{ij} \qquad \textbf{(5-4)}$$

Data analysis 1.

- Scatter plots
- Excluding samples…



05/29/13 00:00

OC = 2.77538

EC = 0.26673

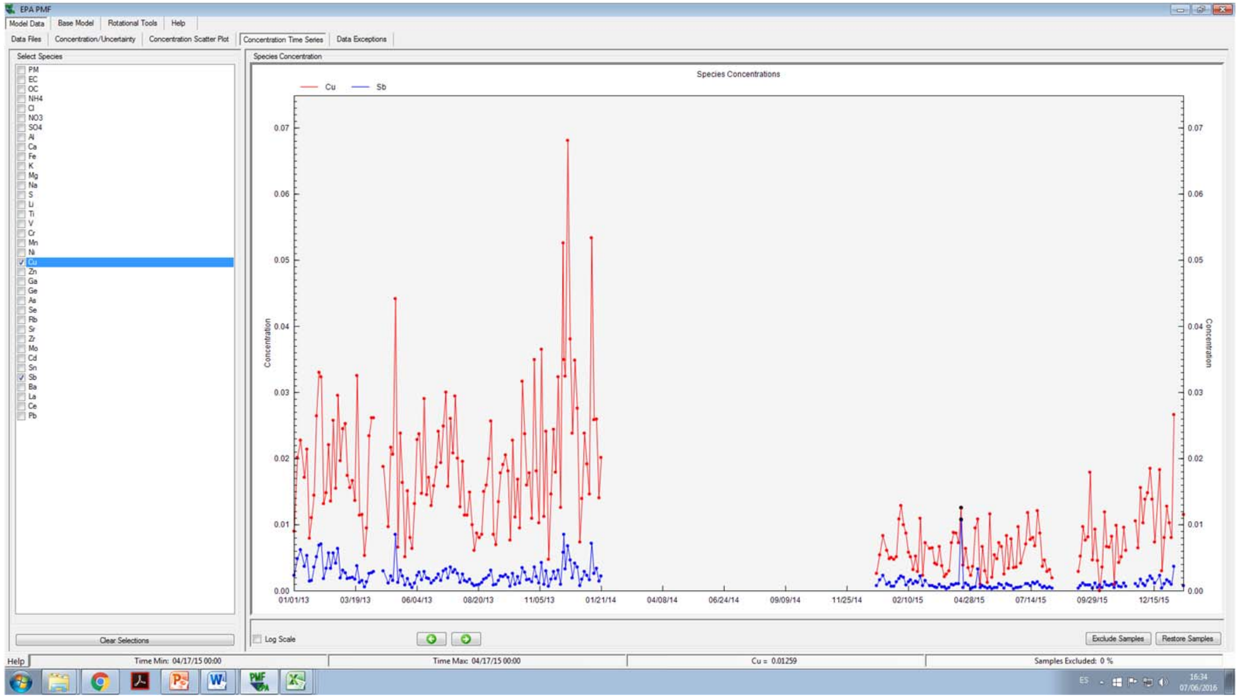
y = 1.47138x + 1.05301

LIFE11
ENV/ES/000584

AIRUSE

Data analysis 2

- Time series
- Summary of data exception



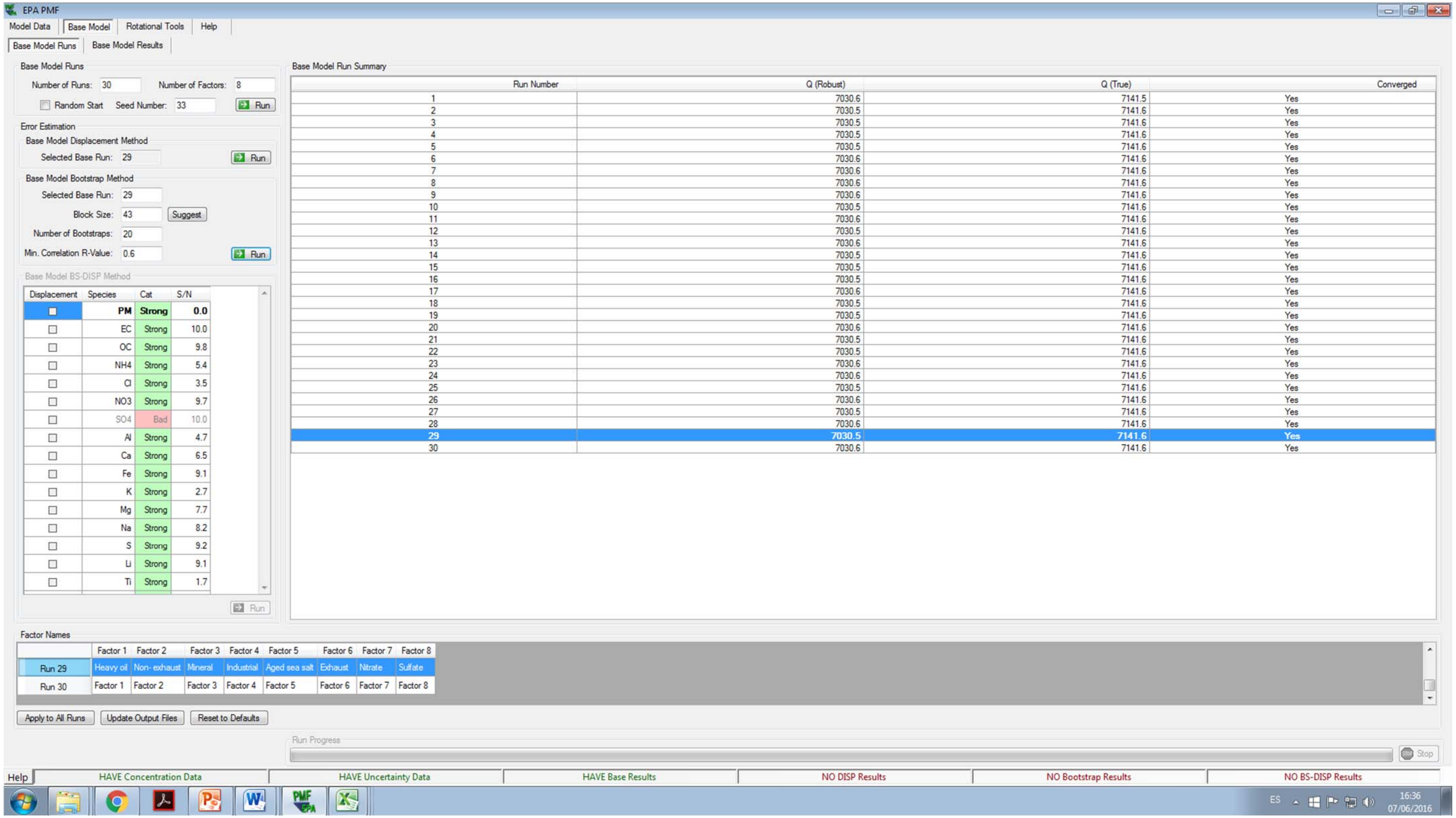
Time Min: 04/17/15 00:00

Time Max: 04/17/15 00:00

Cu = 0.01259

Samples Excluded: 0 %

Run the model



Base Model Runs

Number of Runs: 30
Number of Factors: 8
☐ Random Start
Seed Number: 33

Error Estimation

Base Model Displacement Method
Selected Base Run: 29

Base Model Bootstrap Method
Selected Base Run: 29
Block Size: 43
Number of Bootstraps: 20
Min. Correlation R-Value: 0.6

Base Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	**PM**	**Strong**	**0.0**
☐	EC	Strong	10.0
☐	OC	Strong	9.8
☐	NH4	Strong	5.4
☐	Cl	Strong	3.5
☐	NO3	Strong	9.7
☐	SO4	Bad	10.0
☐	Al	Strong	4.7
☐	Ca	Strong	6.5
☐	Fe	Strong	9.1
☐	K	Strong	2.7
☐	Mg	Strong	7.7
☐	Na	Strong	8.2
☐	S	Strong	9.2
☐	Li	Strong	9.1
☐	Ti	Strong	1.7

Base Model Run Summary

Run Number	Q (Robust)	Q (True)	Converged
1	7030.6	7141.5	Yes
2	7030.5	7141.6	Yes
3	7030.5	7141.6	Yes
4	7030.5	7141.6	Yes
5	7030.5	7141.6	Yes
6	7030.6	7141.6	Yes
7	7030.6	7141.6	Yes
8	7030.6	7141.6	Yes
9	7030.6	7141.6	Yes
10	7030.5	7141.6	Yes
11	7030.6	7141.6	Yes
12	7030.5	7141.6	Yes
13	7030.6	7141.6	Yes
14	7030.5	7141.6	Yes
15	7030.5	7141.6	Yes
16	7030.5	7141.6	Yes
17	7030.6	7141.6	Yes
18	7030.5	7141.6	Yes
19	7030.5	7141.6	Yes
20	7030.6	7141.6	Yes
21	7030.5	7141.6	Yes
22	7030.5	7141.6	Yes
23	7030.6	7141.6	Yes
24	7030.6	7141.6	Yes
25	7030.5	7141.6	Yes
26	7030.6	7141.6	Yes
27	7030.5	7141.6	Yes
28	7030.6	7141.6	Yes
29	7030.5	7141.6	Yes
30	7030.6	7141.6	Yes

Factor Names

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Run 29	Heavy oil	Non- exhaust	Mineral	Industrial	Aged sea salt	Exhaust	Nitrate	Sulfate
Run 30	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8

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

ES
16:36
07/06/2016

Q values

- Min Qrobust is highlighted

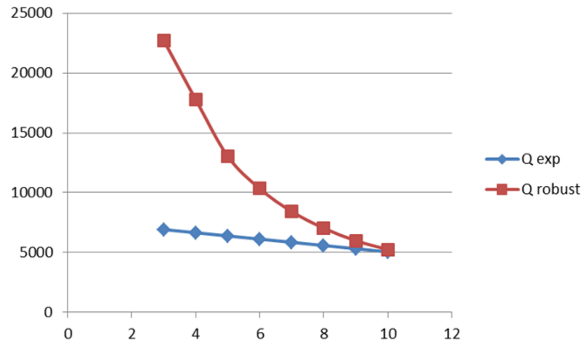
$$Q_{\text{Robust}} = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{e_{ij}}{h_{ij}\sigma_{ij}}\right)^{2} \qquad (4)$$

where

$$h_{ij} = 1 \quad \text{for} \quad |e_{ij}/\sigma_{ij}| \leq \alpha \qquad (5a)$$

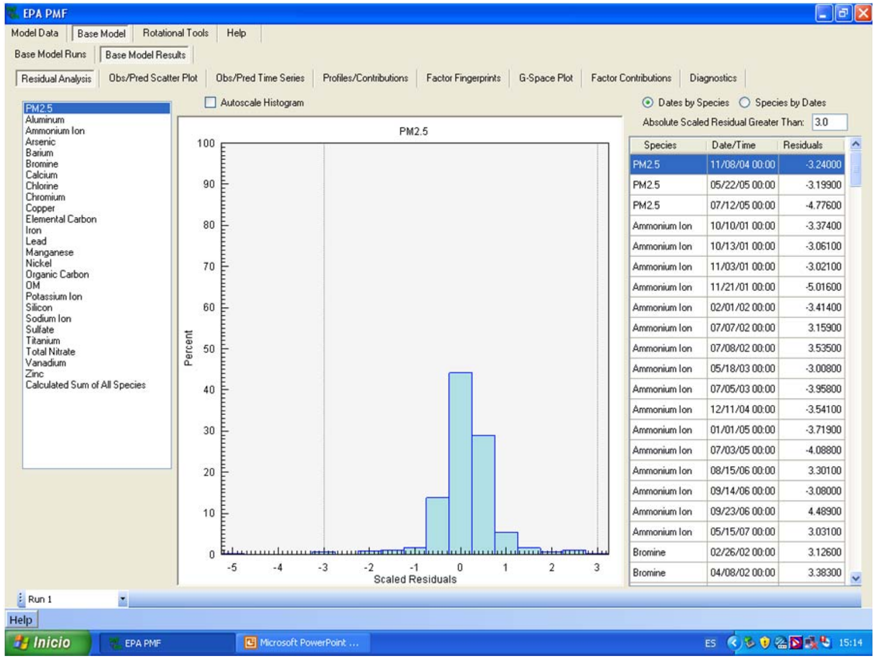
$$h_{ij} = |e_{ij}/\sigma_{ij}|/\alpha \quad \text{for} \quad |e_{ij}/\sigma_{ij}| > \alpha \qquad (5b)$$

In robust mode, the PMF algorithm attempts to minimize Q_{Robust} rather than Q as defined in eq 2 (hereafter, the latter is referred to as Q_{True}).



Results and diagnostics 1

- Residual analysis
- Autoscale
- High residual samples



☐ Autoscale Histogram

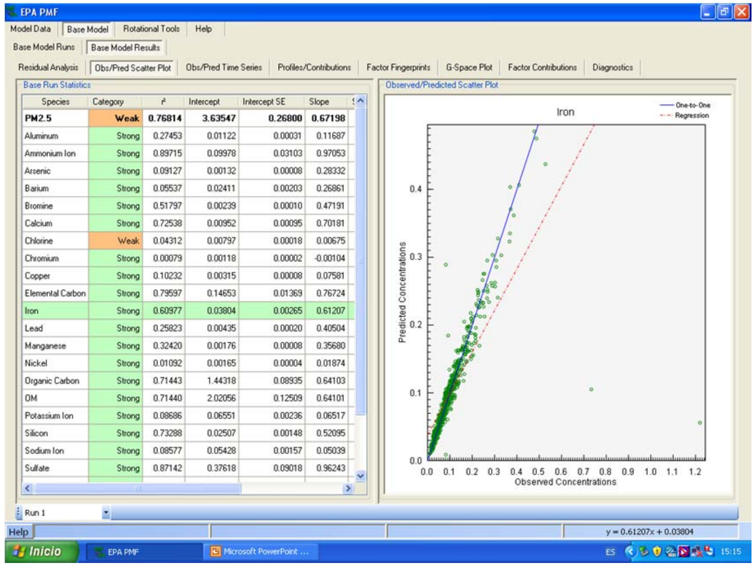
● Dates by Species ○ Species by Dates

Absolute Scaled Residual Greater Than: 3.0

Species	Date/Time	Residuals
PM2.5	11/08/04 00:00	-3.24000
PM2.5	05/22/05 00:00	-3.19900
PM2.5	07/12/05 00:00	-4.77600
Ammonium Ion	10/10/01 00:00	-3.37400
Ammonium Ion	10/13/01 00:00	-3.06100
Ammonium Ion	11/03/01 00:00	-3.02100
Ammonium Ion	11/21/01 00:00	-5.01600
Ammonium Ion	02/01/02 00:00	-3.41400
Ammonium Ion	07/07/02 00:00	3.15900
Ammonium Ion	07/08/02 00:00	3.53500
Ammonium Ion	05/18/03 00:00	-3.00800
Ammonium Ion	07/05/03 00:00	-3.95800
Ammonium Ion	12/11/04 00:00	-3.54100
Ammonium Ion	01/01/05 00:00	-3.71900
Ammonium Ion	07/03/05 00:00	-4.08800
Ammonium Ion	08/15/06 00:00	3.30100
Ammonium Ion	09/14/06 00:00	-3.08000
Ammonium Ion	09/23/06 00:00	4.48900
Ammonium Ion	05/15/07 00:00	3.03100
Bromine	02/26/02 00:00	3.12600
Bromine	04/08/02 00:00	3.38300

Results and diagnostics 2

- Obs/Pred
- Revisit specie selection?



Species	Category	r²	Intercept	Intercept SE	Slope
PM2.5	**Weak**	**0.76814**	**3.63547**	**0.26800**	**0.67198**
Aluminum	Strong	0.27453	0.01122	0.00031	0.11687
Ammonium Ion	Strong	0.89715	0.09978	0.03103	0.97053
Arsenic	Strong	0.09127	0.00132	0.00008	0.28332
Barium	Strong	0.05537	0.02411	0.00203	0.26861
Bromine	Strong	0.51797	0.00239	0.00010	0.47191
Calcium	Strong	0.72538	0.00952	0.00095	0.70181
Chlorine	Weak	0.04312	0.00797	0.00018	0.00675
Chromium	Strong	0.00079	0.00118	0.00002	-0.00104
Copper	Strong	0.10232	0.00315	0.00008	0.07581
Elemental Carbon	Strong	0.79597	0.14653	0.01369	0.76724
Iron	Strong	0.60977	0.03804	0.00265	0.61207
Lead	Strong	0.25823	0.00435	0.00020	0.40504
Manganese	Strong	0.32420	0.00176	0.00008	0.35680
Nickel	Strong	0.01092	0.00165	0.00004	0.01874
Organic Carbon	Strong	0.71443	1.44318	0.08935	0.64103
OM	Strong	0.71440	2.02056	0.12509	0.64101
Potassium Ion	Strong	0.08686	0.06551	0.00236	0.06517
Silicon	Strong	0.73288	0.02507	0.00148	0.52095
Sodium Ion	Strong	0.08577	0.05428	0.00157	0.05039
Sulfate	Strong	0.87142	0.37618	0.09018	0.96243

y = 0.61207x + 0.03804

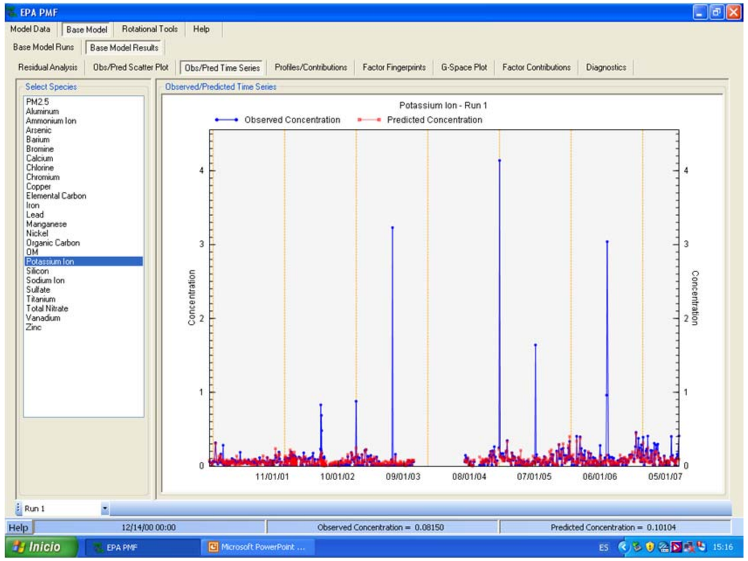
CSIC

LIFE11
ENV/ES/000584

AIRUSE

Results and diagnostics 3

- Obs/Pred time series
- Revisit sample selection?
- Missing sources?



12/14/00 00:00

Observed Concentration = 0.08150

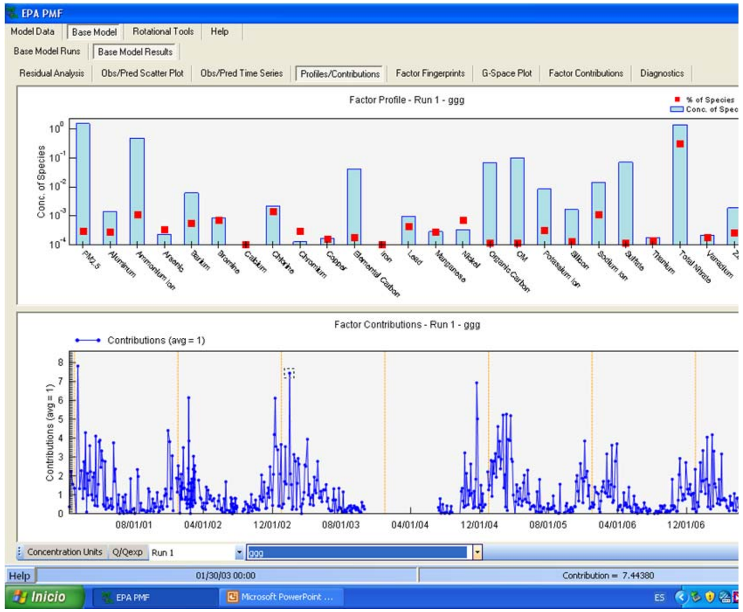
Predicted Concentration = 0.10104

LIFE11
ENV/ES/000584

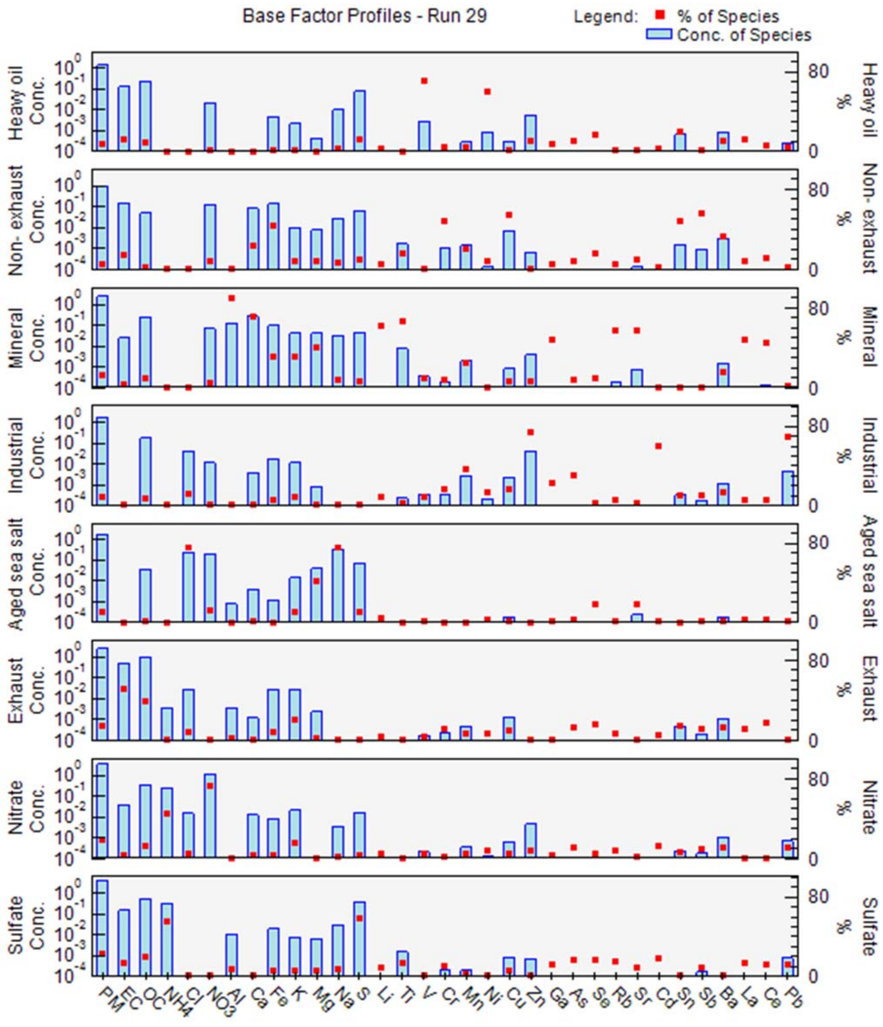
Results and diagnostics 4

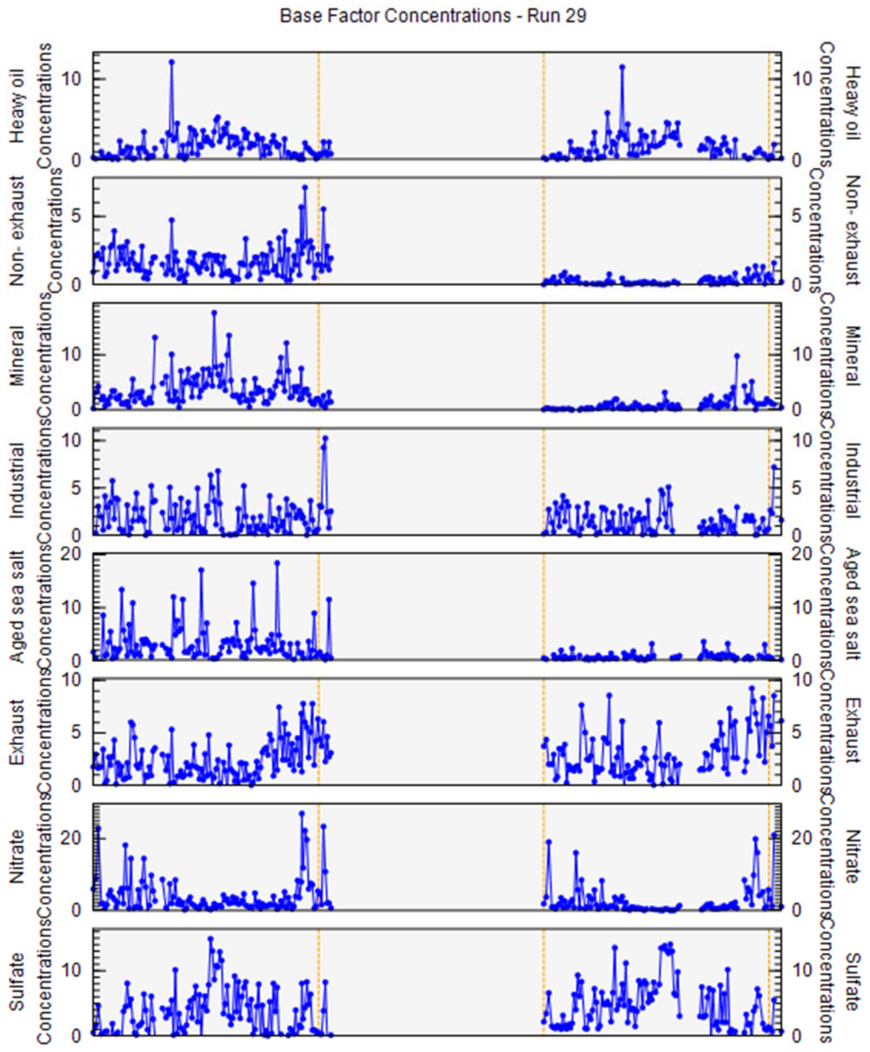
- Concentration unit
- Q/Qexp

- 3 matrices in the output file
  - ☐ µg/m3
  - ☐ % of species
  - ☐ % of total variable



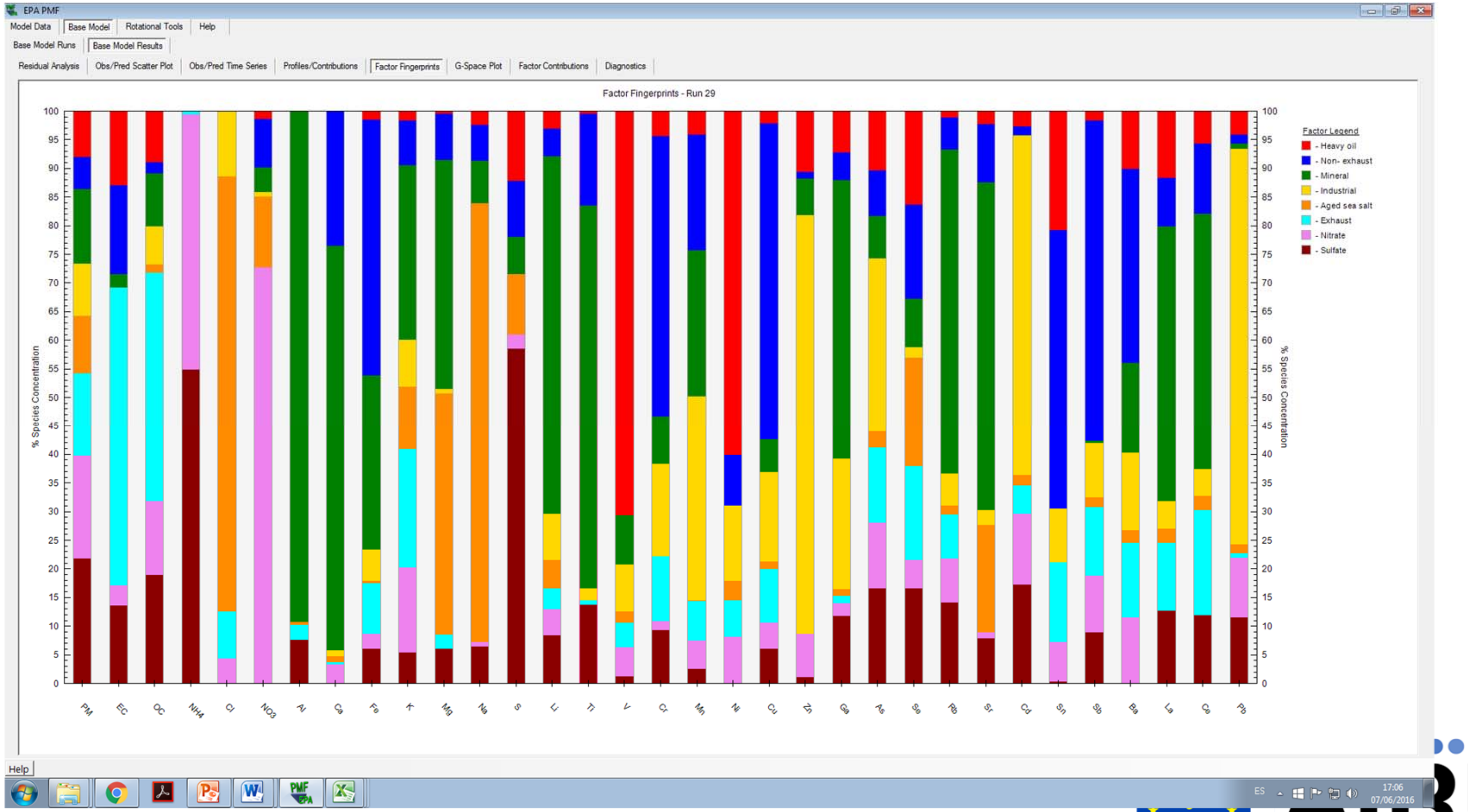
Stacked graphs


Base Factor Profiles - Run 29

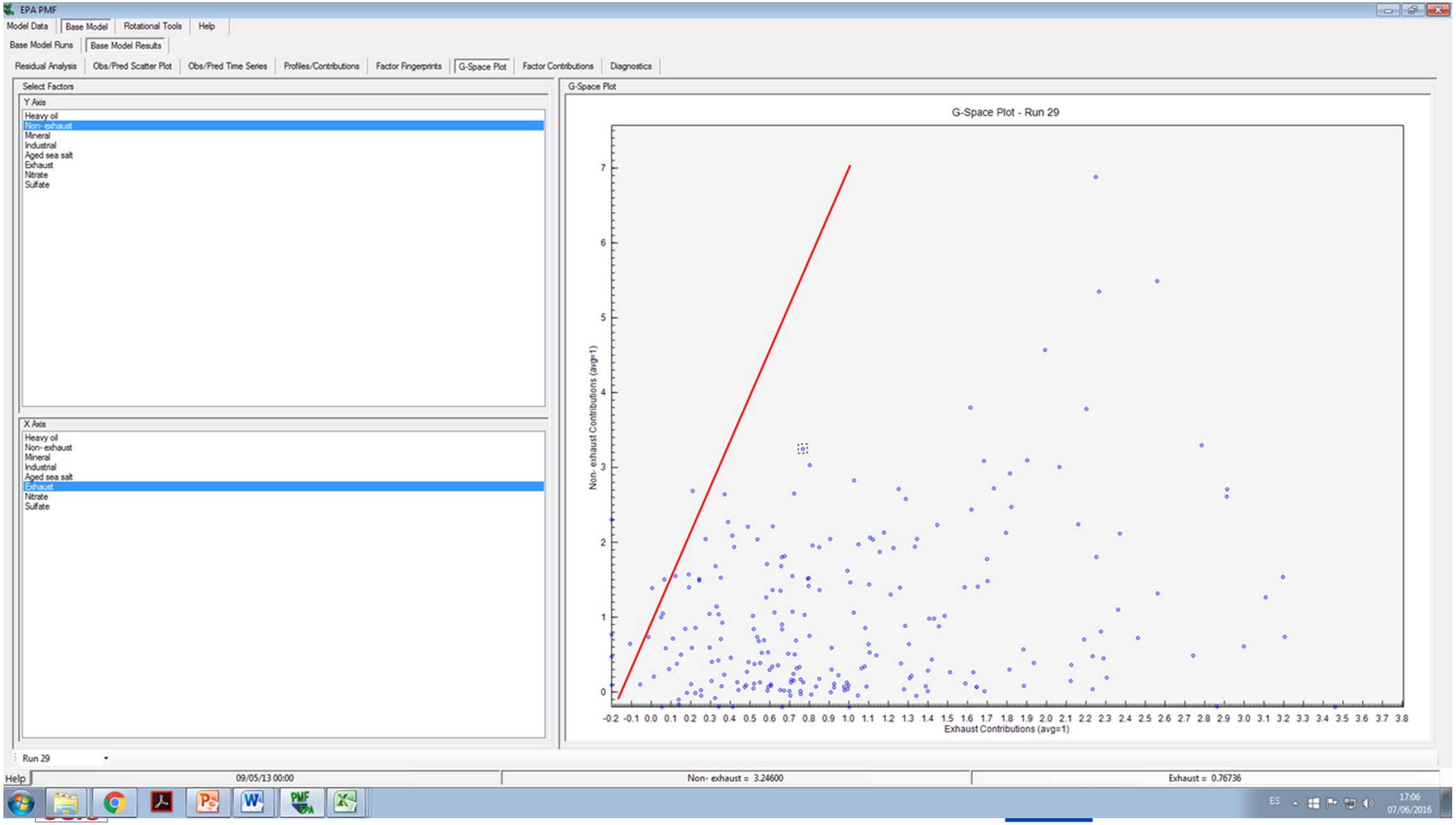

Base Factor Concentrations - Run 29

Factor fingerprint

- Does not include residuals



G-space plots



Y Axis: Non- exhaust

X Axis: Exhaust

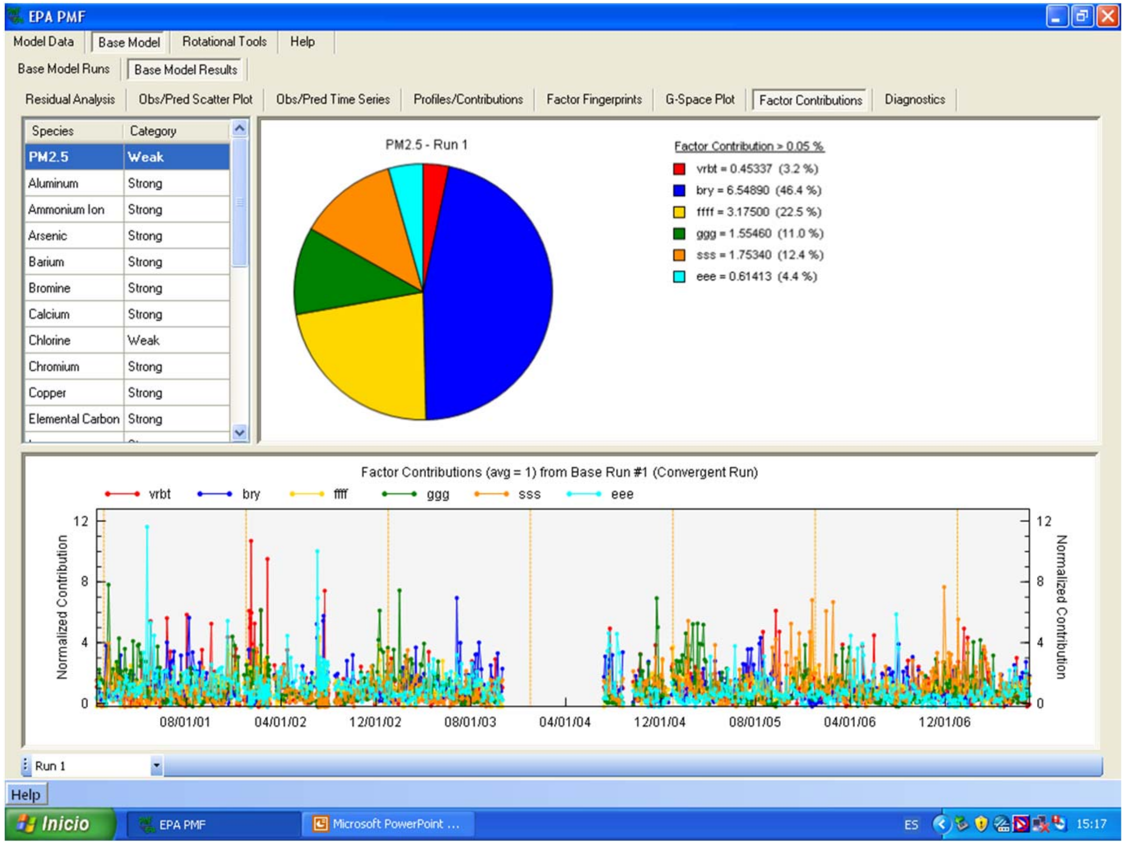
Run 29

09/05/13 00:00

Non- exhaust = 3.24600

Exhaust = 0.76736

Contributions



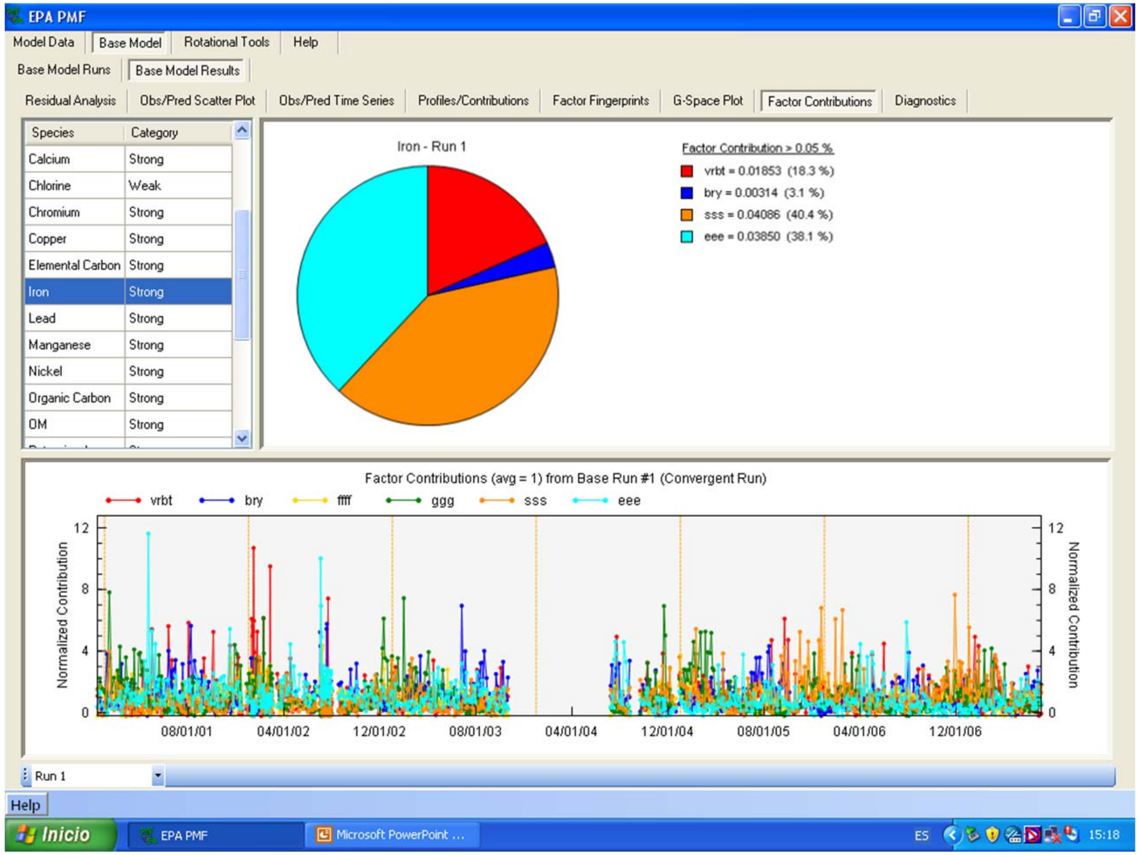
Species	Category
PM2.5	Weak
Aluminum	Strong
Ammonium Ion	Strong
Arsenic	Strong
Barium	Strong
Bromine	Strong
Calcium	Strong
Chlorine	Weak
Chromium	Strong
Copper	Strong
Elemental Carbon	Strong

CSIC

LIFE11
ENV/ES/000584

AIRUSE

For single species



Species	Category
Calcium	Strong
Chlorine	Weak
Chromium	Strong
Copper	Strong
Elemental Carbon	Strong
Iron	Strong
Lead	Strong
Manganese	Strong
Nickel	Strong
Organic Carbon	Strong
OM	Strong

Iron - Run 1

Factor Contribution > 0.05 %

🟥 vrbt = 0.01853 (18.3 %)
🟦 bry = 0.00314 (3.1 %)
🟧 sss = 0.04086 (40.4 %)
🟦 eee = 0.03850 (38.1 %)

Factor Contributions (avg = 1) from Base Run #1 (Convergent Run)

LIFE11
ENV/ES/000584

Output error estimate

1. Random errors (Bootstrapping).
2. Rotational errors (Displacement)
3. Both (BS-DISP)

Modelling errors difficult to estimate, and they are not minimized by large datasets

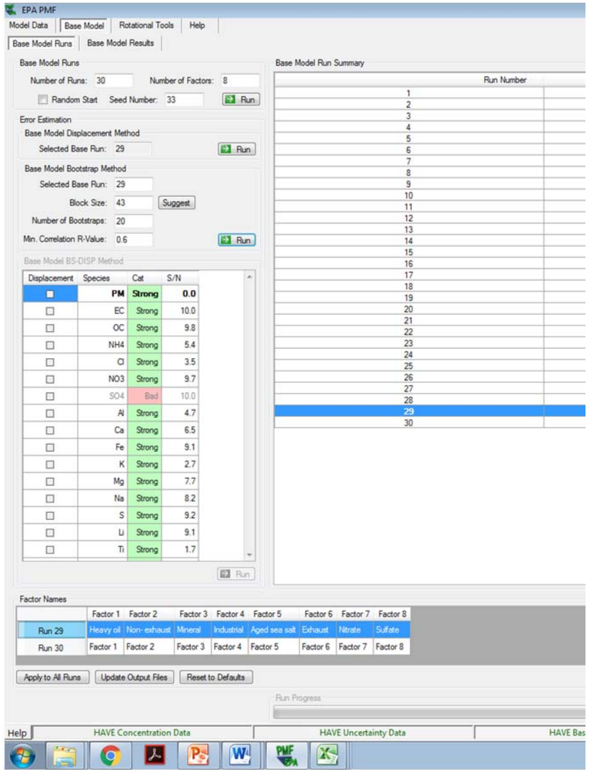
If uncertainties
are not well specified, use only
BS-DISP or BS

Zíková et al (2016) Atmospheric Environment, 134, pp. 84-95.

Brown, et al (2015) Science of the Total Environment, 518-519, pp. 626-635.

Paatero et al (2014) Atmospheric Measurement Techniques, 7 (3), pp. 781-797.

Xie et al., 2012, 2014; Kioumourtzoglou et al., 2014



Base Model Runs

Number of Runs: 30
Number of Factors: 8
☐ Random Start
Seed Number: 33

Error Estimation

Base Model Displacement Method
Selected Base Run: 29

Base Model Bootstrap Method
Selected Base Run: 29
Block Size: 43
Number of Bootstraps: 20
Min. Correlation R-Value: 0.6

Base Model BS-DISP Method

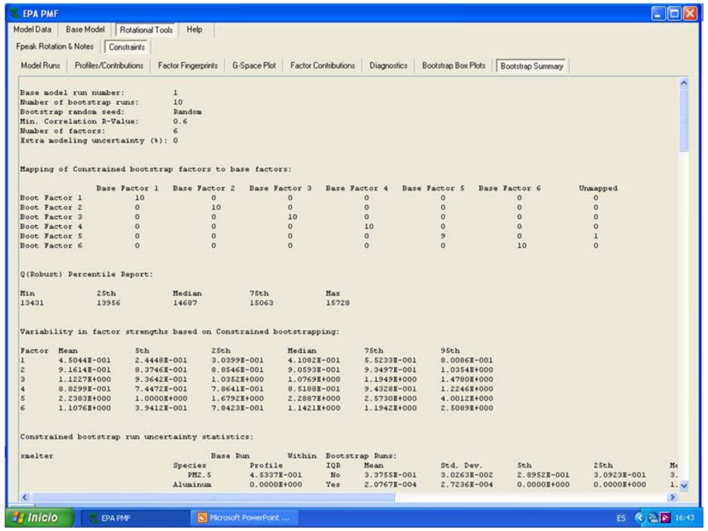
Displacement	Species	Cat	S/N
☐	**PM**	**Strong**	**0.0**
☐	EC	Strong	10.0
☐	OC	Strong	9.8
☐	NH4	Strong	5.4
☐	Cl	Strong	3.5
☐	NO3	Strong	9.7
☐	SO4	Bad	10.0
☐	Al	Strong	4.7
☐	Ca	Strong	6.5
☐	Fe	Strong	9.1
☐	K	Strong	2.7
☐	Mg	Strong	7.7
☐	Na	Strong	8.2
☐	S	Strong	9.2
☐	Li	Strong	9.1
☐	Ti	Strong	1.7

Factor Names

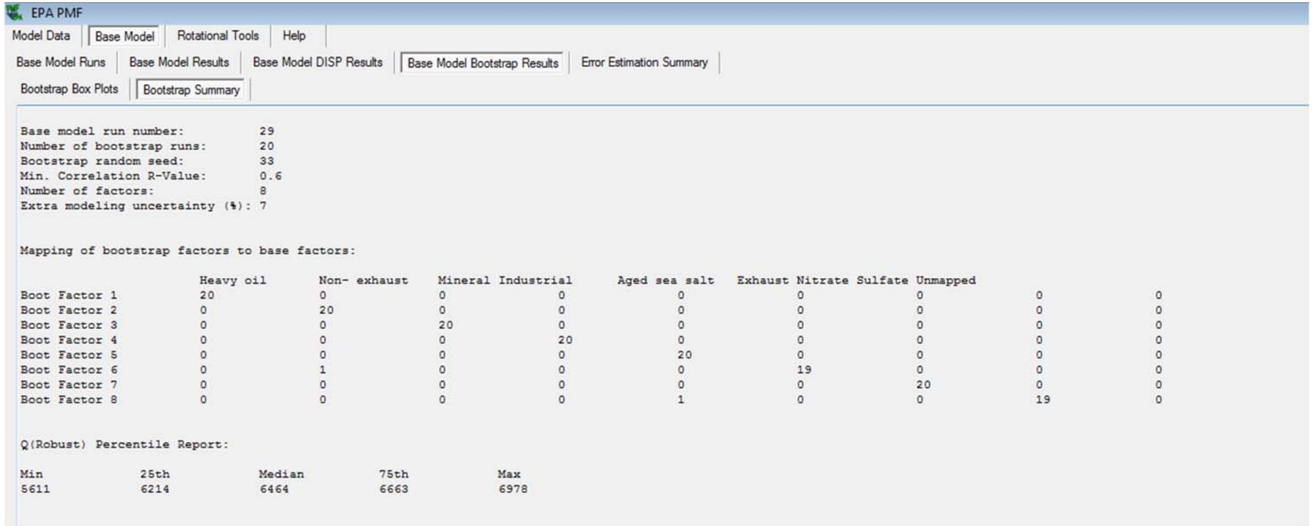
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Run 29	Heavy oil	Non- exhaust	Mineral	Industrial	Aged sea salt	Exhaust	Nitrate	Sulfate
Run 30	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8

Bootstrapping diagnostic

- Random errors, partially evaluates the rotational ambiguity
- not influenced by the user-specified sample uncertainties.
- >80% mapped factors

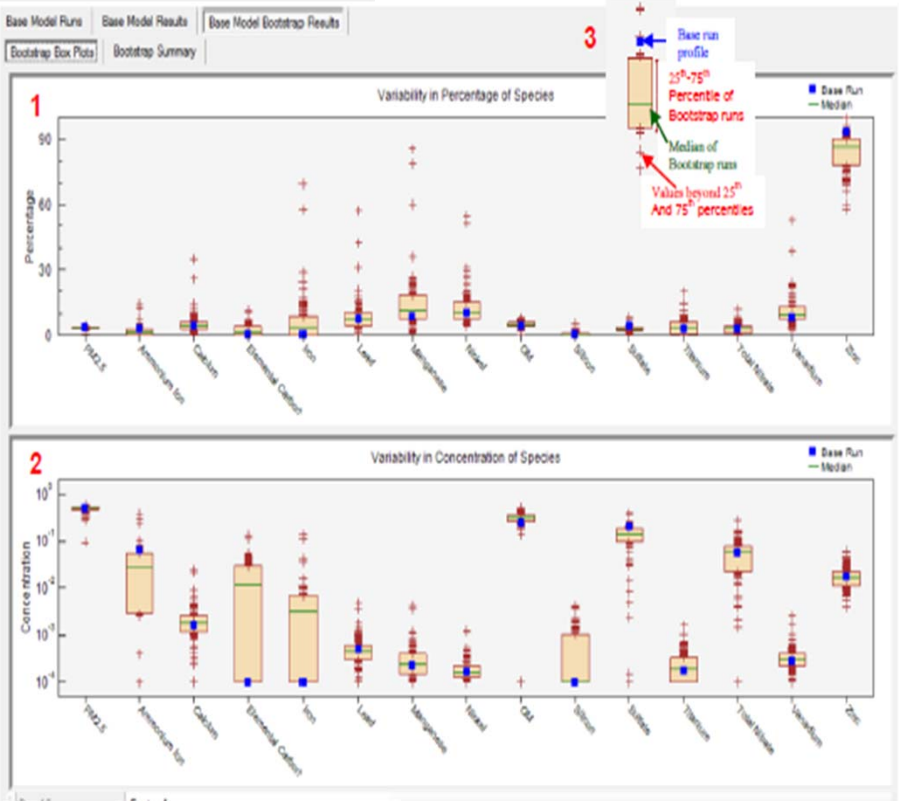


```
Base model run number:         1
Number of bootstrap runs:      10
Bootstrap random seed:         Random
Min. Correlation R-Value:      0.6
Number of factors:             6
Extra modeling uncertainty (%): 0


Mapping of Constrained bootstrap factors to base factors:

                Base Factor 1   Base Factor 2   Base Factor 3   Base Factor 4   Base Factor 5   Base Factor 6        Unmapped
Boot Factor 1        10              0               0               0               0               0               0
Boot Factor 2        0               10              0               0               0               0               0
Boot Factor 3        0               0               10              0               0               0               0
Boot Factor 4        0               0               0               10              0               0               0
Boot Factor 5        0               0               0               0               9               0               1
Boot Factor 6        0               0               0               0               0               10              0


Q(Robust) Percentile Report:

Min             25th            Median          75th            Max
13431           13956           14687           15063           15728


Variability in factor strengths based on Constrained bootstrapping:

Factor  Mean            5th             25th            Median          75th            95th
1       4.5044E-001     2.4448E-001     3.0399E-001     4.1082E-001     5.5233E-001     8.0086E-001
2       9.1614E-001     8.3746E-001     8.8546E-001     9.0593E-001     9.3497E-001     1.0354E+000
3       1.1227E+000     9.3642E-001     1.0352E+000     1.0769E+000     1.1949E+000     1.4780E+000
4       8.8299E-001     7.4472E-001     7.8641E-001     8.5188E-001     9.4328E-001     1.2246E+000
5       2.2383E+000     1.0000E+000     1.6792E+000     2.2887E+000     2.5730E+000     4.0012E+000
6       1.1076E+000     3.9412E-001     7.8423E-001     1.1421E+000     1.1942E+000     2.5089E+000


Constrained bootstrap run uncertainty statistics:

smelter                         Base Run        Within  Bootstrap Runs:
                        Species         Profile         IQR     Mean            Std. Dev.       5th             25th            Me
                           PM2.5        4.5337E-001      No     3.3755E-001     3.0263E-002     2.8952E-001     3.0923E-001     3.
                        Aluminum        0.0000E+000     Yes     2.0767E-004     2.7236E-004     0.0000E+000     0.0000E+000     1.
```



```
Base model run number:          29
Number of bootstrap runs:       20
Bootstrap random seed:          33
Min. Correlation R-Value:       0.6
Number of factors:              8
Extra modeling uncertainty (%): 7


Mapping of bootstrap factors to base factors:

                        Heavy oil       Non- exhaust    Mineral Industrial      Aged sea salt   Exhaust Nitrate Sulfate Unmapped
Boot Factor 1           20              0               0               0               0               0               0               0               0
Boot Factor 2           0               20              0               0               0               0               0               0               0
Boot Factor 3           0               0               20              0               0               0               0               0               0
Boot Factor 4           0               0               0               20              0               0               0               0               0
Boot Factor 5           0               0               0               0               20              0               0               0               0
Boot Factor 6           0               1               0               0               0               19              0               0               0
Boot Factor 7           0               0               0               0               0               0               20              0               0
Boot Factor 8           0               0               0               0               1               0               0               19              0


Q(Robust) Percentile Report:

Min             25th            Median          75th            Max
5611            6214            6464            6663            6978
```

- Species with base run outside the IQR need evaluation



DISP

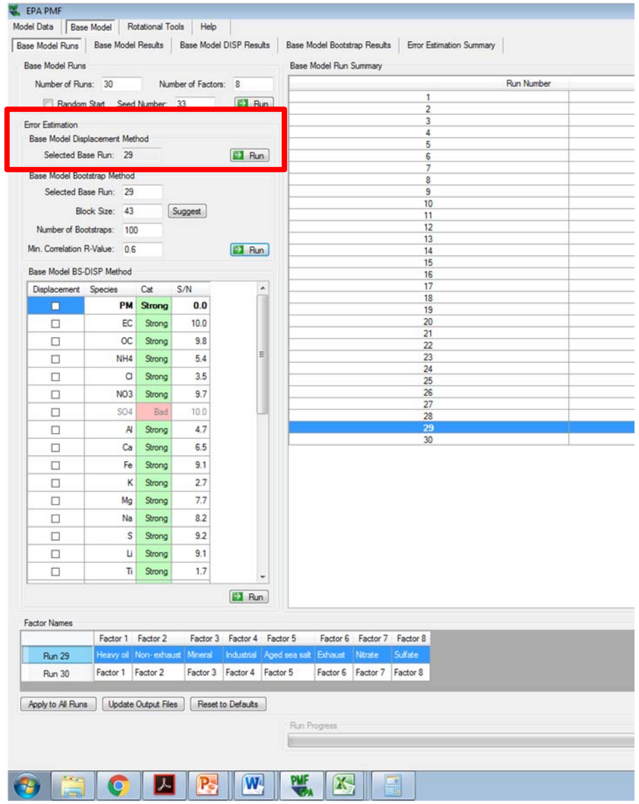
Only on strong/tracer species.

Valid only if data uncertainties are reliable

Invalid if Q/Qexp>10

Diagnostics:

- Error code: 0
- Second value <1%
- Number of swaps



Base Model Runs

Number of Runs: 30
Number of Factors: 8
☐ Random Start
Seed Number: 33

Error Estimation

Base Model Displacement Method

Selected Base Run: 29

Base Model Bootstrap Method

Selected Base Run: 29
Block Size: 43
Number of Bootstraps: 100
Min. Correlation R-Value: 0.6

Base Model BS-DISP Method

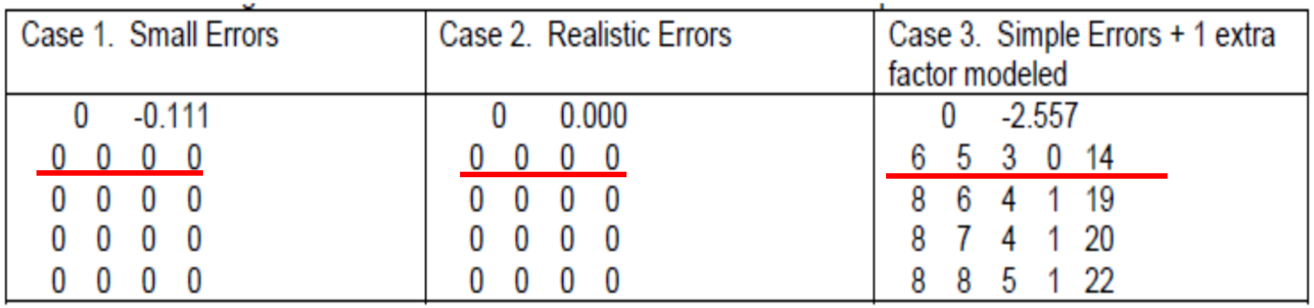
Displacement	Species	Cat	S/N
☐	**PM**	**Strong**	**0.0**
☐	EC	Strong	10.0
☐	OC	Strong	9.8
☐	NH4	Strong	5.4
☐	Cl	Strong	3.5
☐	NO3	Strong	9.7
☐	SO4	Bad	10.0
☐	Al	Strong	4.7
☐	Ca	Strong	6.5
☐	Fe	Strong	9.1
☐	K	Strong	2.7
☐	Mg	Strong	7.7
☐	Na	Strong	8.2
☐	S	Strong	9.2
☐	Li	Strong	9.1
☐	Ti	Strong	1.7

Base Model Run Summary

Run Number
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

Factor Names

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Run 29	Heavy oil	Non- exhaust	Mineral	Industrial	Aged sea salt	Exhaust	Nitrate	Sulfate
Run 30	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8

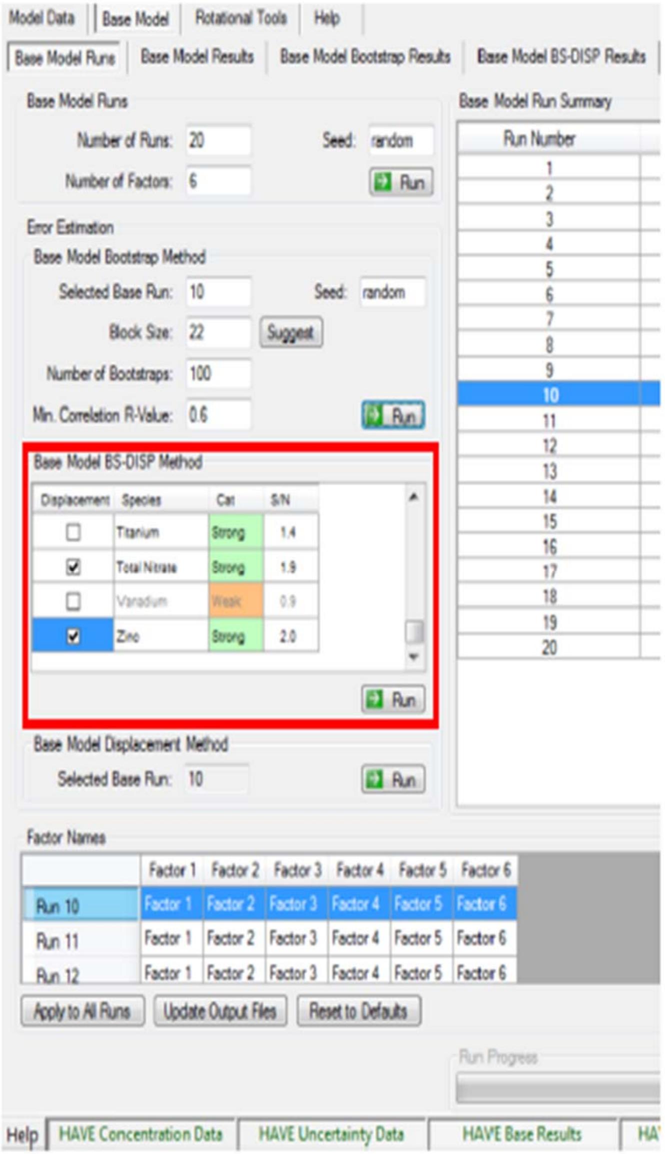


Case 1. Small Errors	Case 2. Realistic Errors	Case 3. Simple Errors + 1 extra factor modeled
0 -0.111	0 0.000	0 -2.557
0 0 0 0	0 0 0 0	6 5 3 0 14
0 0 0 0	0 0 0 0	8 6 4 1 19
0 0 0 0	0 0 0 0	8 7 4 1 20
0 0 0 0	0 0 0 0	8 8 5 1 22

BS-DISP

- Only key and major species
- Long computation times

1. Each DISP defines the span of rotationally accessible space.
2. Each BS resample moves this space around, randomly in different directions.
3. Taken together, all the replications of the rotationally accessible space, in random locations, represent both the random uncertainty and the rotational uncertainty.



Base Model Runs

Number of Runs: 20
Seed: random
Number of Factors: 6

Error Estimation

Base Model Bootstrap Method

Selected Base Run: 10
Seed: random
Block Size: 22
Number of Bootstraps: 100
Min. Correlation R-Value: 0.6

Base Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	Titanium	Strong	1.4
☑	Total Nitrate	Strong	1.9
☐	Vanadium	Weak	0.9
☑	Zinc	Strong	2.0

Base Model Displacement Method

Selected Base Run: 10

Base Model Run Summary

Run Number
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

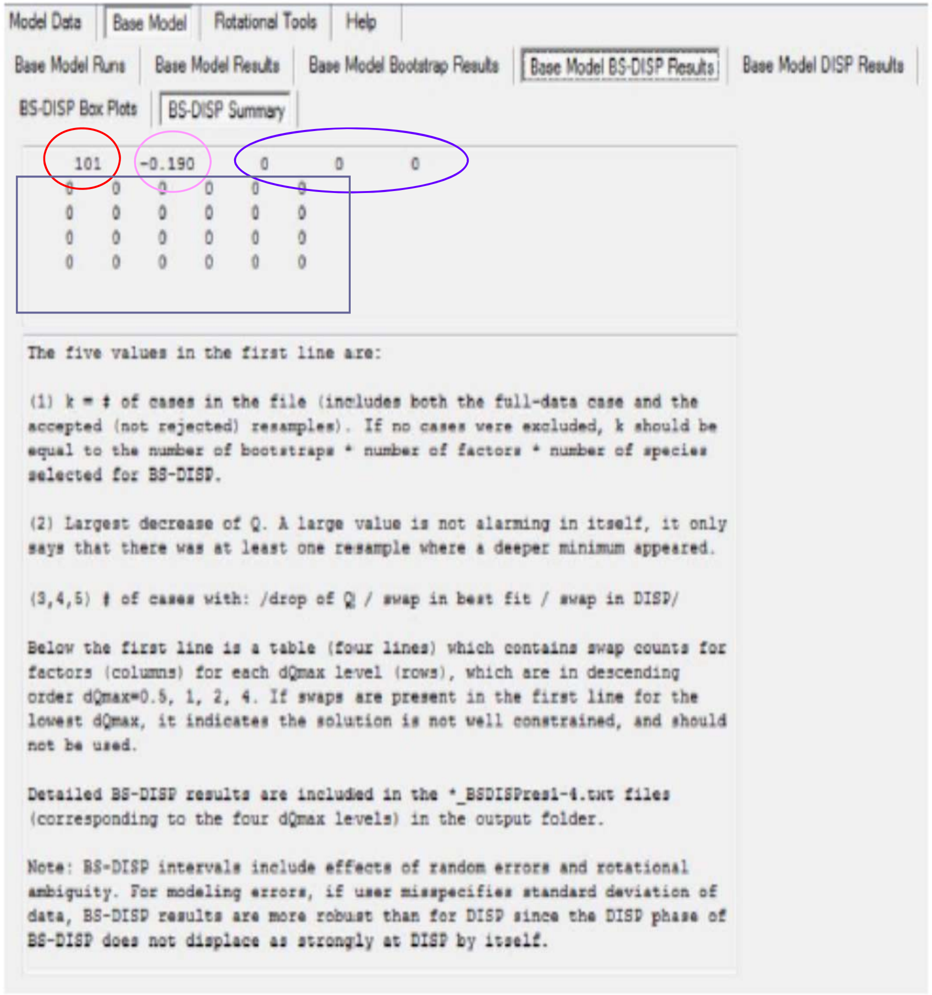
Factor Names

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Run 10	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Run 11	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Run 12	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6

BS-DISP diagnostic

Ideally

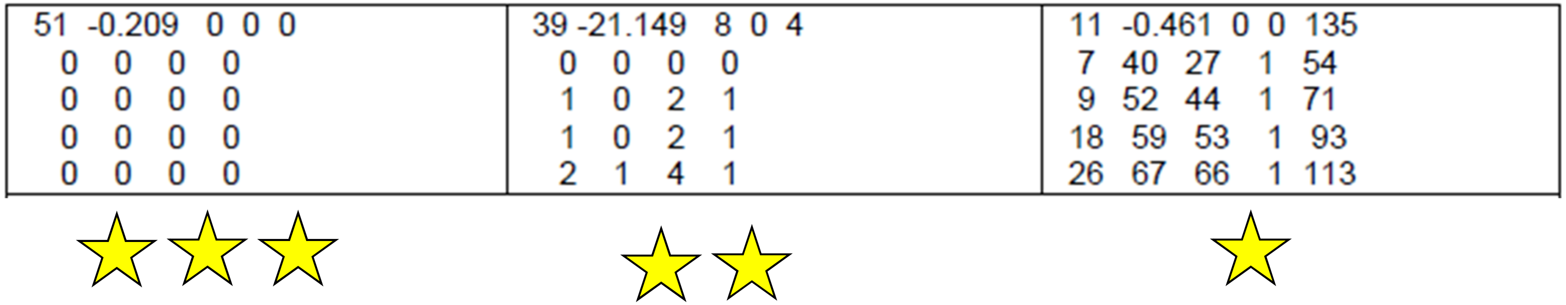
- Nº of runs+1
- <1% (>0.5% could indicate a local minimum)
- Zero values
- 0 in first row



```
   101    -0.190     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) k = # of cases in the file (includes both the full-data case and the
accepted (not rejected) resamples). If no cases were excluded, k should be
equal to the number of bootstraps * number of factors * number of species
selected for BS-DISP.

(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/

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 should
not be used.

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. For modeling errors, if user misspecifies standard deviation of
data, BS-DISP results are more robust than for DISP since the DISP phase of
BS-DISP does not displace as strongly at DISP by itself.
```

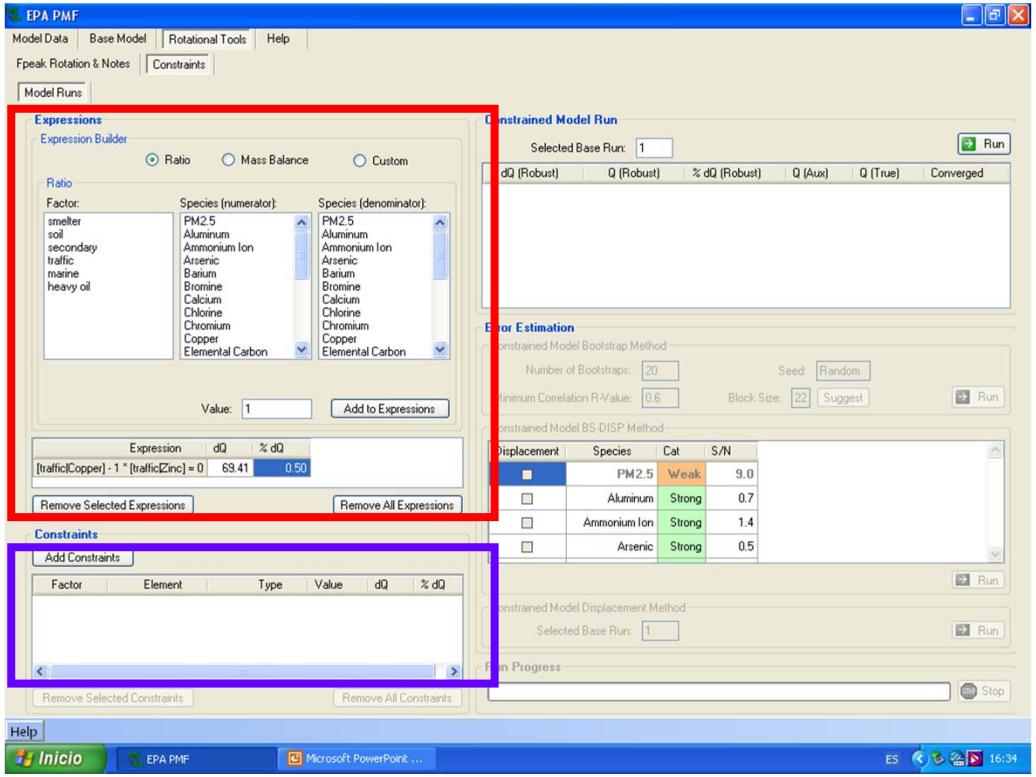
BS-DISP: three examples



51 -0.209 0 0 0	39 -21.149 8 0 4	11 -0.461 0 0 135
0 0 0 0	0 0 0 0	7 40 27 1 54
0 0 0 0	1 0 2 1	9 52 44 1 71
0 0 0 0	1 0 2 1	18 59 53 1 93
0 0 0 0	2 1 4 1	26 67 66 1 113
★★★	★★	★

Rotational tools

- FPEAK (similar to previous version, Paatero 2000)
- Constraints
  - ☐ Expr.builder
  - ☐ From Base run



Value: 1

Expression	dQ	% dQ
[traffic|Copper] - 1 * [traffic|Zinc] = 0	69.41	0.50

Factor	Element	Type	Value	dQ	% dQ

Selected Base Run: 1

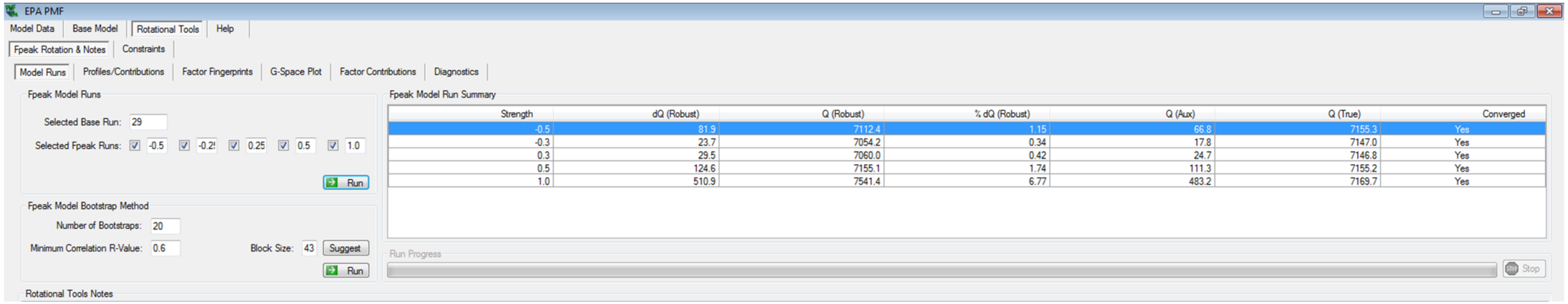
dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged

Number of Bootstraps: 20
Seed: Random
Minimum Correlation R-Value: 0.6
Block Size: 22

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Strong	0.7
☐	Ammonium Ion	Strong	1.4
☐	Arsenic	Strong	0.5

Selected Base Run: 1

Fpeak



Fpeak Model Runs

Selected Base Run: 29

Selected Fpeak Runs: ☑ -0.5 ☑ -0.2! ☑ 0.25 ☑ 0.5 ☑ 1.0

Fpeak Model Bootstrap Method

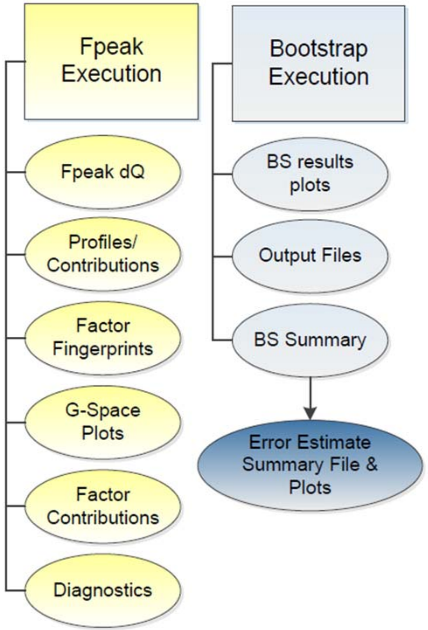
Number of Bootstraps: 20

Minimum Correlation R-Value: 0.6

Block Size: 43

Fpeak Model Run Summary

Strength	dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
-0.5	81.9	7112.4	1.15	66.8	7155.3	Yes
-0.3	23.7	7054.2	0.34	17.8	7147.0	Yes
0.3	29.5	7060.0	0.42	24.7	7146.8	Yes
0.5	124.6	7155.1	1.74	111.3	7155.2	Yes
1.0	510.9	7541.4	6.77	483.2	7169.7	Yes



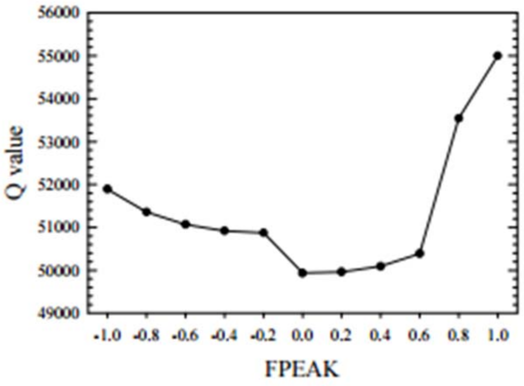
Fpeak Execution

- Fpeak dQ
- Profiles/ Contributions
- Factor Fingerprints
- G-Space Plots
- Factor Contributions
- Diagnostics

Bootstrap Execution

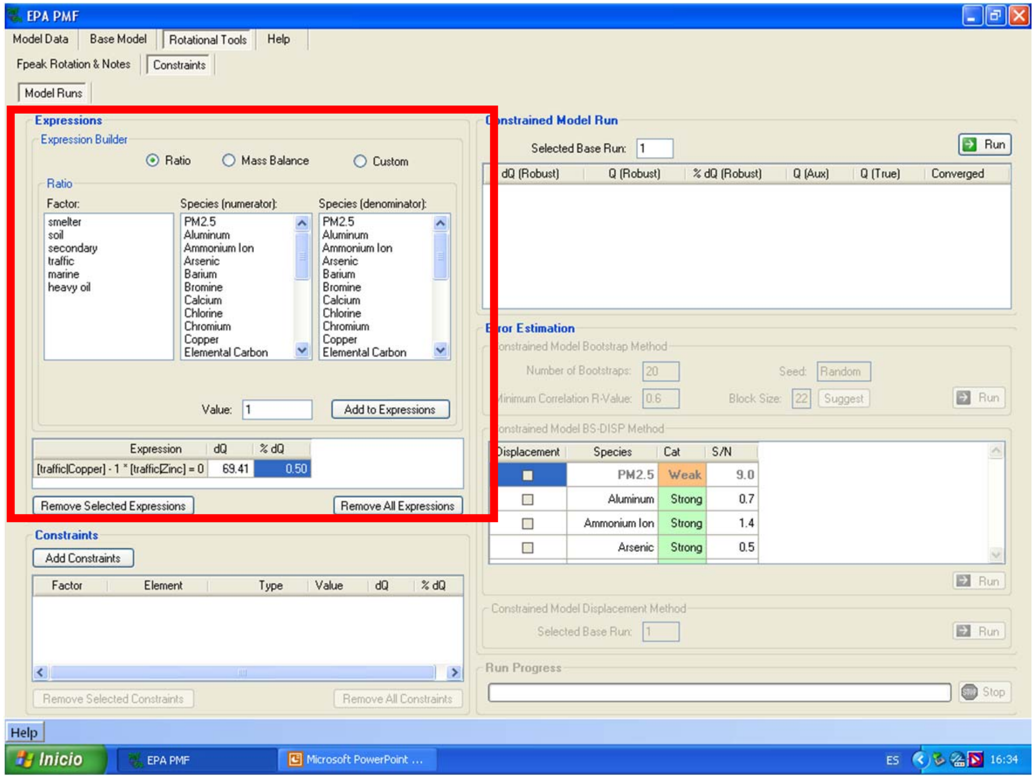
- BS results plots
- Output Files
- BS Summary

Error Estimate Summary File & Plots



https://www3.epa.gov/ttnamti1/files/ambient/pm25/workshop/laymen.pdf

- Expression builder
  - <u>Ratio (elements of same factors)</u>
  - Mass balance (elements of different factors)
  - Custom (also on contributions)



● Ratio ○ Mass Balance ○ Custom

Ratio

Factor:	Species (numerator):	Species (denominator):
smelter	PM2.5	PM2.5
soil	Aluminum	Aluminum
secondary	Ammonium Ion	Ammonium Ion
traffic	Arsenic	Arsenic
marine	Barium	Barium
heavy oil	Bromine	Bromine
	Calcium	Calcium
	Chlorine	Chlorine
	Chromium	Chromium
	Copper	Copper
	Elemental Carbon	Elemental Carbon

Value: 1

Expression	dQ	% dQ
[traffic|Copper] - 1 * [traffic|Zinc] = 0	69.41	0.50

Factor	Element	Type	Value	dQ	% dQ

Selected Base Run: 1

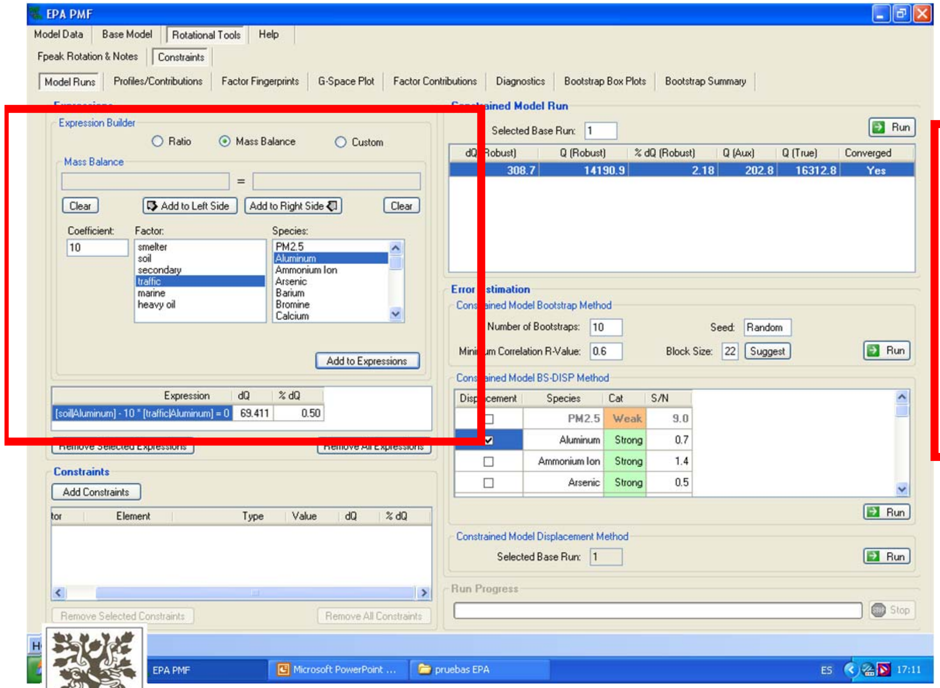
dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged

Number of Bootstraps: 20
Seed: Random
Minimum Correlation R-Value: 0.6
Block Size: 22

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Strong	0.7
☐	Ammonium Ion	Strong	1.4
☐	Arsenic	Strong	0.5

Selected Base Run: 1

- Expression builder
  - ☐ Ratio (elements of same factors)
  - ☐ Mass balance (elements of different factors)
  - ☐ Custom (also on contributions)



○ Ratio ● Mass Balance ○ Custom

Coefficient: 10

Factor: smelter, soil, secondary, **traffic**, marine, heavy oil

Species: PM2.5, **Aluminum**, Ammonium Ion, Arsenic, Barium, Bromine, Calcium

Expression	dQ	% dQ
[soil|Aluminum] - 10 * [traffic|Aluminum] = 0	69.411	0.50

Selected Base Run: 1

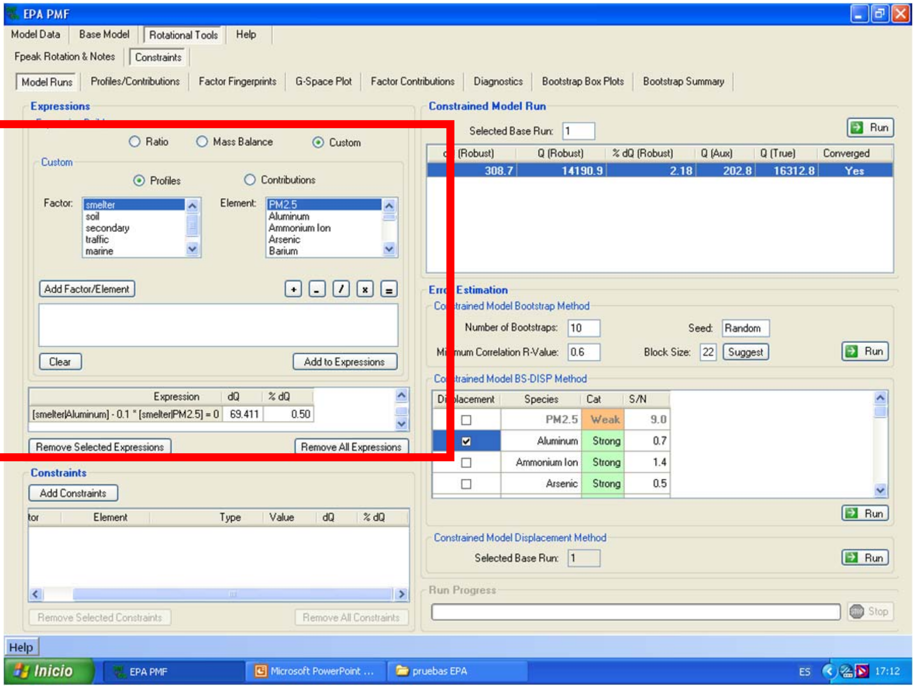
dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
308.7	14190.9	2.18	202.8	16312.8	Yes

Number of Bootstraps: 10
Seed: Random
Minimum Correlation R-Value: 0.6
Block Size: 22

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☑	Aluminum	Strong	0.7
☐	Ammonium Ion	Strong	1.4
☐	Arsenic	Strong	0.5

tor	Element	Type	Value	dQ	% dQ

Selected Base Run: 1



○ Ratio ○ Mass Balance ● Custom

● Profiles ○ Contributions

Factor: **smelter**, soil, secondary, traffic, marine

Element: **PM2.5**, Aluminum, Ammonium Ion, Arsenic, Barium

Expression	dQ	% dQ
[smelter|Aluminum] - 0.1 * [smelter|PM2.5] = 0	69.411	0.50

Selected Base Run: 1

(Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
308.7	14190.9	2.18	202.8	16312.8	Yes

Number of Bootstraps: 10
Seed: Random
Minimum Correlation R-Value: 0.6
Block Size: 22

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☑	Aluminum	Strong	0.7
☐	Ammonium Ion	Strong	1.4
☐	Arsenic	Strong	0.5

tor	Element	Type	Value	dQ	% dQ

Selected Base Run: 1

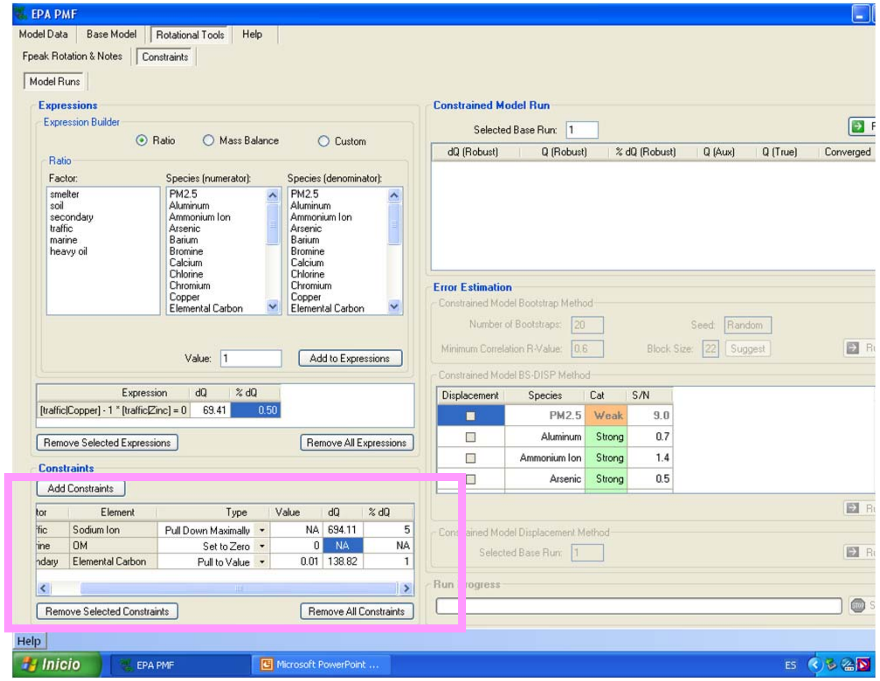
From the base run

- Pulling existing factor profile
- Pulling existing contributions
- Pulling G space plots

Strength of pulling

Soft options (with maximum dQ):
- Pull down maximally
- Pull up maximally
- Pull to value

Hard options (regardless of dQ):
- Forced to zero
- Forced to original value
- Forced to Upper/lower limits



● Ratio ○ Mass Balance ○ Custom

Ratio

Factor:	Species (numerator):	Species (denominator):
smelter soil secondary traffic marine heavy oil	PM2.5 Aluminum Ammonium Ion Arsenic Barium Bromine Calcium Chlorine Chromium Copper Elemental Carbon	PM2.5 Aluminum Ammonium Ion Arsenic Barium Bromine Calcium Chlorine Chromium Copper Elemental Carbon

Value: 1

Expression	dQ	% dQ
[traffic|Copper] - 1 * [traffic|Zinc] = 0	69.41	0.50

tor	Element	Type	Value	dQ	% dQ
ffic	Sodium Ion	Pull Down Maximally	NA	694.11	5
ine	OM	Set to Zero	0	NA	NA
ndary	Elemental Carbon	Pull to Value	0.01	138.82	1

Selected Base Run: 1

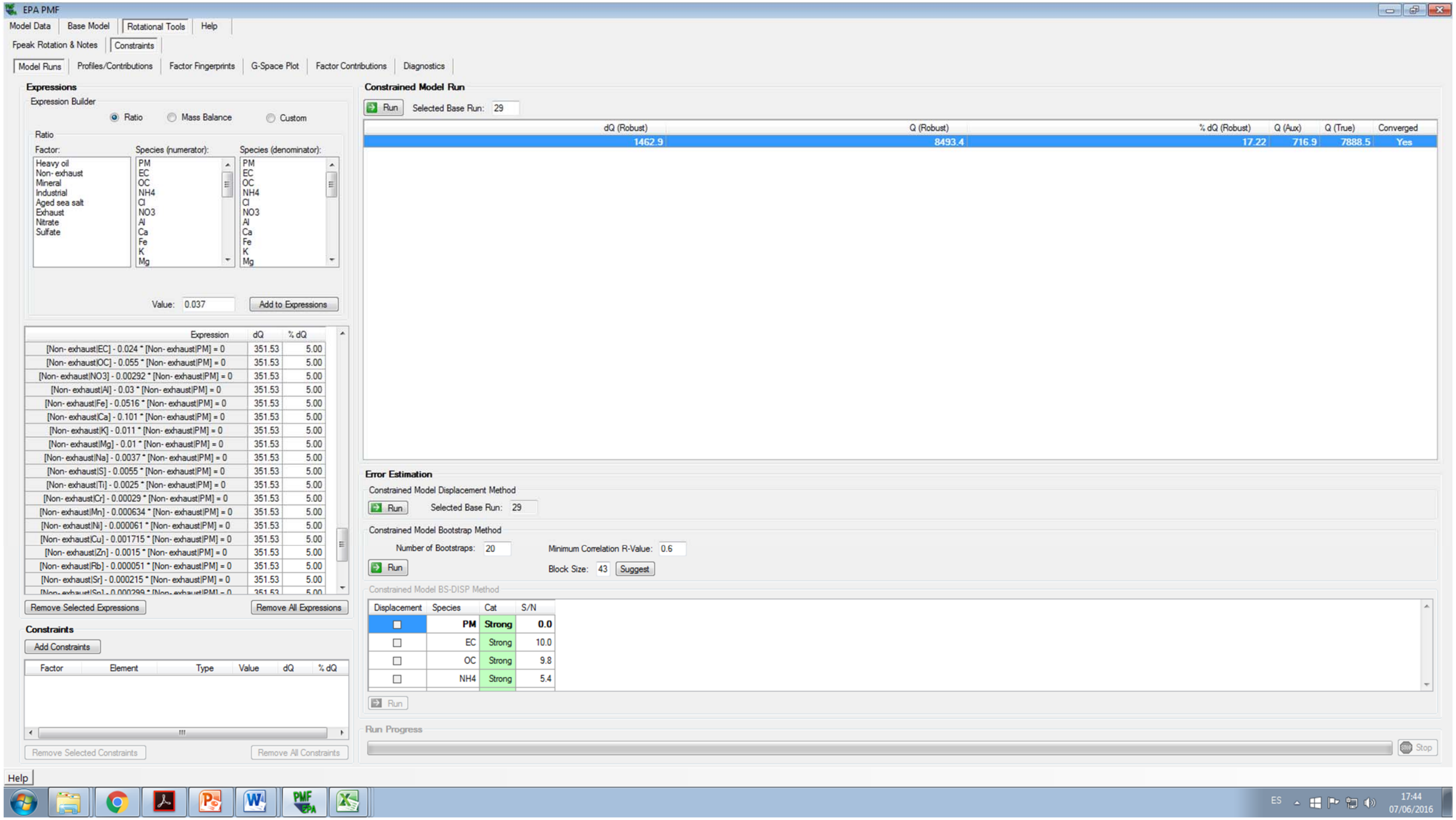
dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged

Number of Bootstraps: 20
Seed: Random
Minimum Correlation R-Value: 0.6
Block Size: 22

Displacement	Species	Cat	S/N
☐	PM2.5	Weak	9.0
☐	Aluminum	Strong	0.7
☐	Ammonium Ion	Strong	1.4
☐	Arsenic	Strong	0.5

Selected Base Run: 1

Run Constraint



Expressions

Expression Builder

● Ratio ○ Mass Balance ○ Custom

Ratio

Factor:	Species (numerator):	Species (denominator):
Heavy oil	PM	PM
Non- exhaust	EC	EC
Mineral	OC	OC
Industrial	NH4	NH4
Aged sea salt	Cl	Cl
Exhaust	NO3	NO3
Nitrate	Al	Al
Sulfate	Ca	Ca
	Fe	Fe
	K	K
	Mg	Mg

Value: 0.037

Expression	dQ	% dQ
[Non- exhaust|EC] - 0.024 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|OC] - 0.055 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|NO3] - 0.00292 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Al] - 0.03 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Fe] - 0.0516 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Ca] - 0.101 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|K] - 0.011 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Mg] - 0.01 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Na] - 0.0037 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|S] - 0.0055 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Ti] - 0.0025 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Cr] - 0.00029 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Mn] - 0.000634 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Ni] - 0.000061 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Cu] - 0.001715 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Zn] - 0.0015 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Rb] - 0.000051 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Sr] - 0.000215 * [Non- exhaust|PM] = 0	351.53	5.00
[Non- exhaust|Sn] - 0.000299 * [Non- exhaust|PM] = 0	351.53	5.00

Constraints

Factor	Element	Type	Value	dQ	% dQ

Constrained Model Run

Selected Base Run: 29

dQ (Robust)	Q (Robust)	% dQ (Robust)	Q (Aux)	Q (True)	Converged
1462.9	8493.4	17.22	716.9	7888.5	Yes

Error Estimation

Constrained Model Displacement Method

Selected Base Run: 29

Constrained Model Bootstrap Method

Number of Bootstraps: 20

Minimum Correlation R-Value: 0.6

Block Size: 43

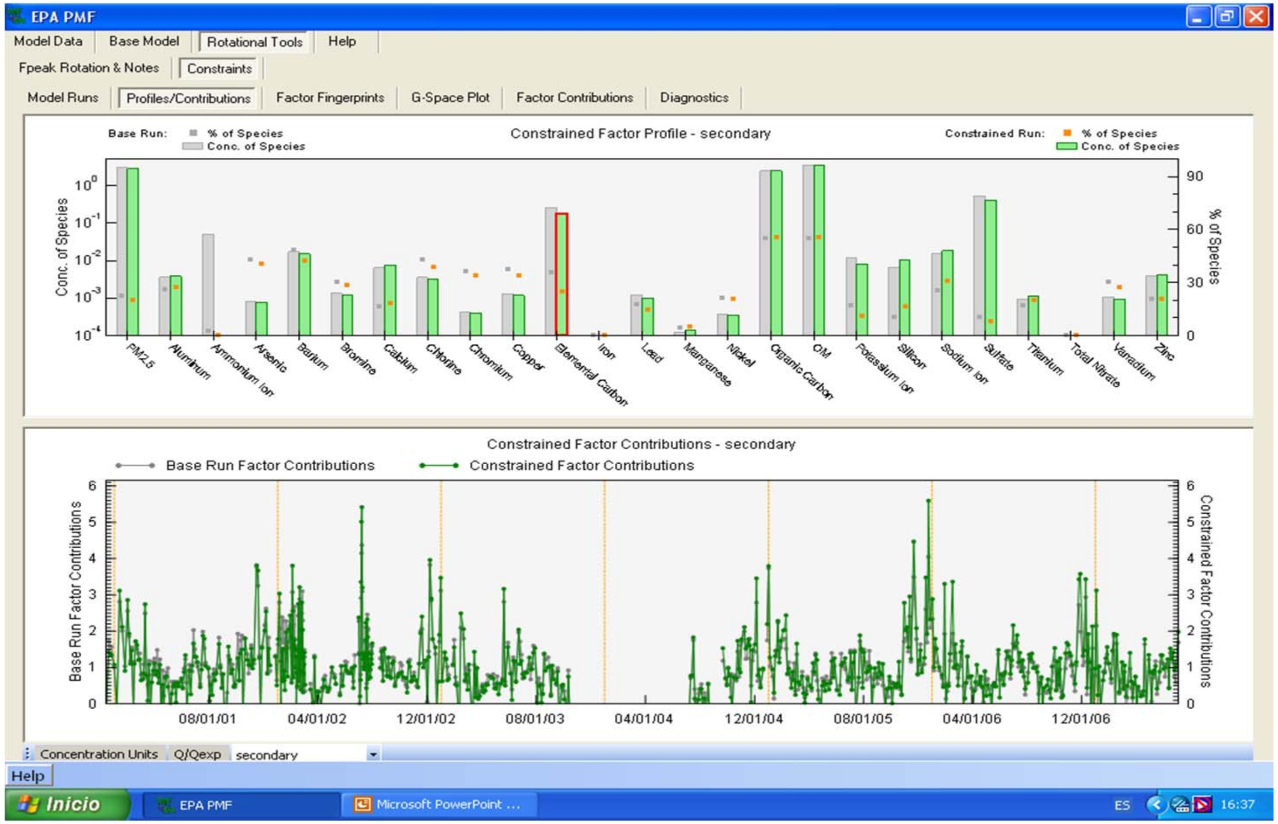
Constrained Model BS-DISP Method

Displacement	Species	Cat	S/N
☐	**PM**	**Strong**	**0.0**
☐	EC	Strong	10.0
☐	OC	Strong	9.8
☐	NH4	Strong	5.4

Run Progress

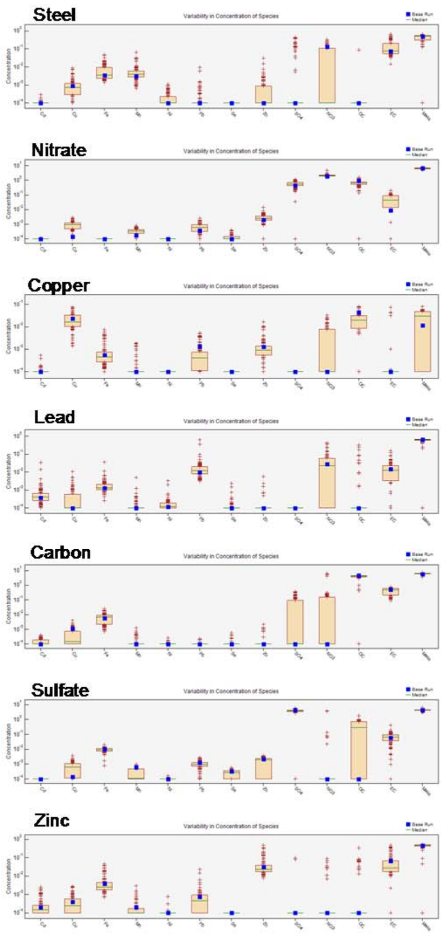
17:44
07/06/2016

Results of constrained PMF

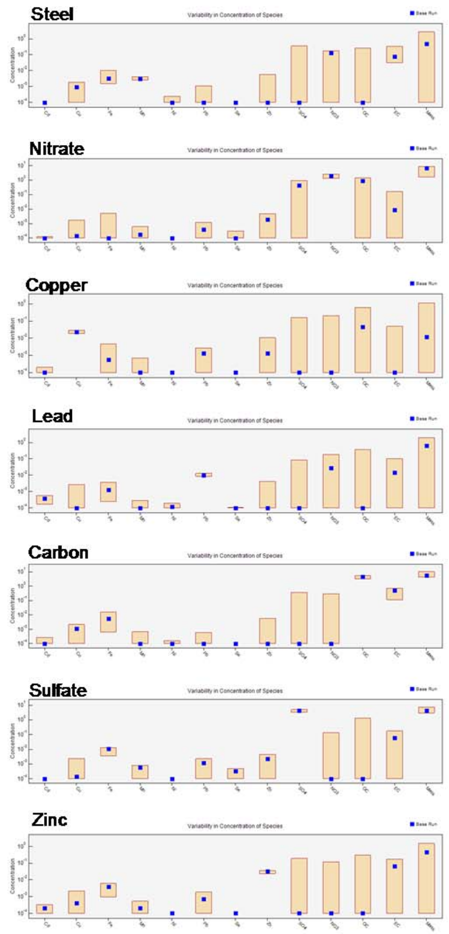


All error estimates can be applied also on Constrained PMF

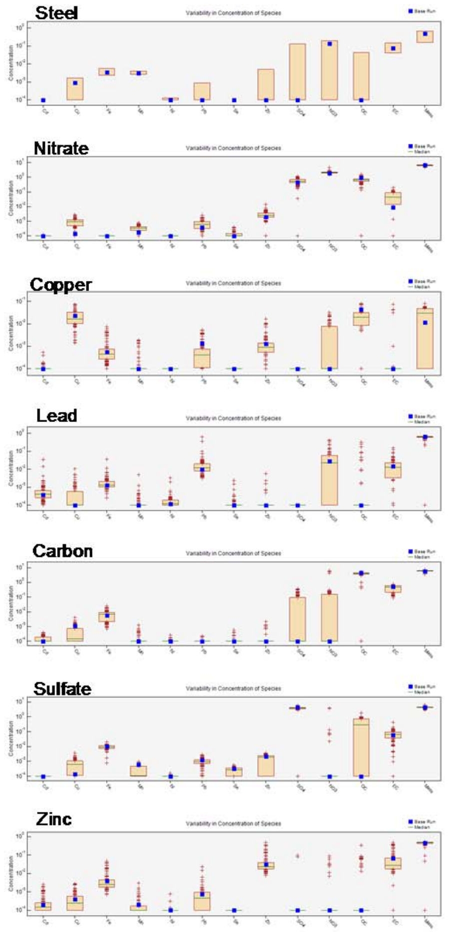
BS



BS-DISP



DISP



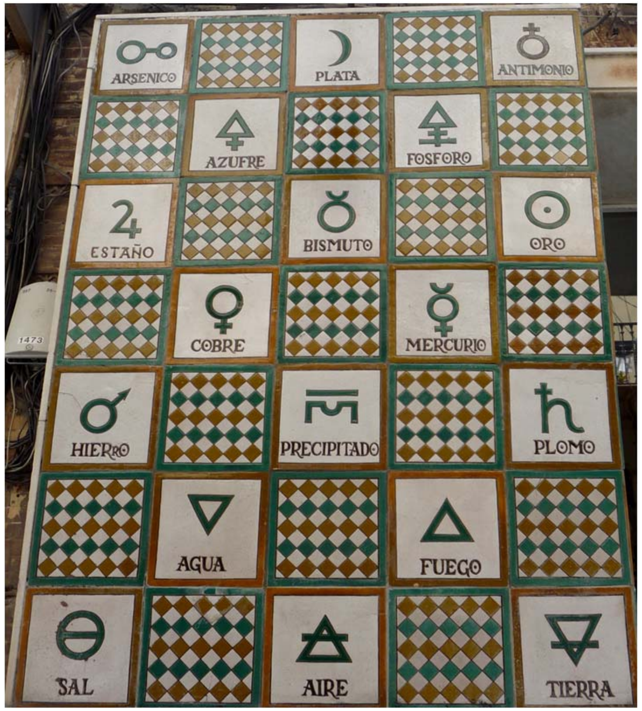
LIFE11
ENV/ES/000584

Thank you for your attention!

…..in some cases a satisfactory analysis cannot be perofrmed with any number of factors

…..we did not reach a full understanding of Error Estimates, more studies are Needed

Brown et al., 2015 (STOTEN)



What to report

- Uncertainty estimate method
- Weak species
- Q values , Q exp
- Missing/BDL values treatment
- BS mapping
- DISP swaps and dQ