REGIONAL TRAINING COURSE ON METHODS AND TOOLS TO IDENTIFY SOURCES OF AIR POLLUTION

IAEA RER 1013: Supporting Air Quality Management

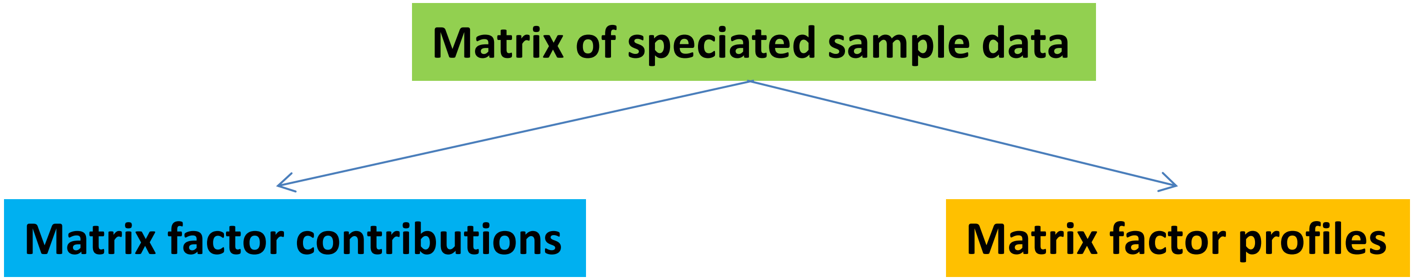
Source apportionment by Positive Matrix Factorization

Marta Almeida
Email: smarta@ctn.ist.utl.pt
Instituto Superior Técnico
Campus Tecnológico e Nuclear
Sacavém, Portugal
02 –06 June, 2014

Concept

Objective:

- Identify number of factors – p
- The species profile of each source – f
- Amount of mass contributed by each factor to each individual sample - g



Matrix of speciated sample data

Matrix factor contributions

Matrix factor profiles

j

i

SAMPLE	Al	As	Br
09-01-2011 12:00	6.4609	0.0008	0.0108
12-01-2011 12:25	7.0842	0.0009	0.0093
14-01-2011 12:45	3.3487	0.0004	0.0053
18-01-2011 13:45	12.2642	0.0017	0.0087
24-01-2011 12:40	0.1984	0.0010	0.0028
26-01-2011 13:20	0.2612	0.0003	0.0023

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

Residual for each sample/species

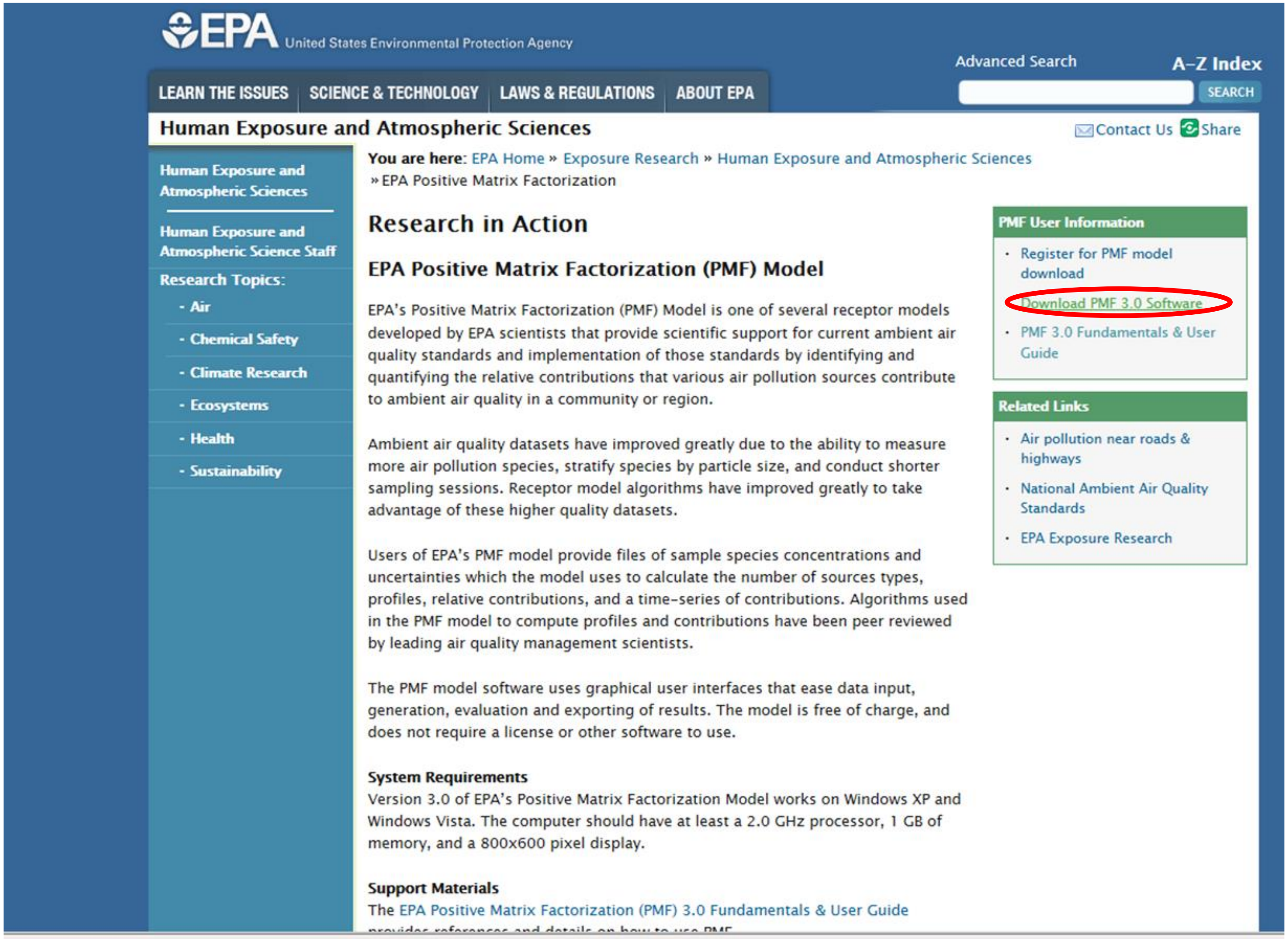
Object function Q

- Results are constrained so that no sample can have negative source contribution;
- PMF allows each data point to be individually weighed;
- The PMF solution minimizes the object function Q, based upon the data uncertainties (*u*).

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

Getting started

Download the programme
http://www.epa.gov/heasd/research/pmf.html



Human Exposure and Atmospheric Sciences

- Human Exposure and Atmospheric Sciences
- Human Exposure and Atmospheric Science Staff
- Research Topics:
  - Air
  - Chemical Safety
  - Climate Research
  - Ecosystems
  - Health
  - Sustainability

You are here: EPA Home » Exposure Research » Human Exposure and Atmospheric Sciences » EPA Positive Matrix Factorization

Research in Action

EPA Positive Matrix Factorization (PMF) Model

EPA's Positive Matrix Factorization (PMF) Model is one of several receptor models developed by EPA scientists that provide scientific support for current ambient air quality standards and implementation of those standards by identifying and quantifying the relative contributions that various air pollution sources contribute to ambient air quality in a community or region.

Ambient air quality datasets have improved greatly due to the ability to measure more air pollution species, stratify species by particle size, and conduct shorter sampling sessions. Receptor model algorithms have improved greatly to take advantage of these higher quality datasets.

Users of EPA's PMF model provide files of sample species concentrations and uncertainties which the model uses to calculate the number of sources types, profiles, relative contributions, and a time-series of contributions. Algorithms used in the PMF model to compute profiles and contributions have been peer reviewed by leading air quality management scientists.

The PMF model software uses graphical user interfaces that ease data input, generation, evaluation and exporting of results. The model is free of charge, and does not require a license or other software to use.

System Requirements
Version 3.0 of EPA's Positive Matrix Factorization Model works on Windows XP and Windows Vista. The computer should have at least a 2.0 GHz processor, 1 GB of memory, and a 800x600 pixel display.

Support Materials
The EPA Positive Matrix Factorization (PMF) 3.0 Fundamentals & User Guide

PMF User Information

- Register for PMF model download
- Download PMF 3.0 Software
- PMF 3.0 Fundamentals & User Guide

Related Links

- Air pollution near roads & highways
- National Ambient Air Quality Standards
- EPA Exposure Research

Getting started

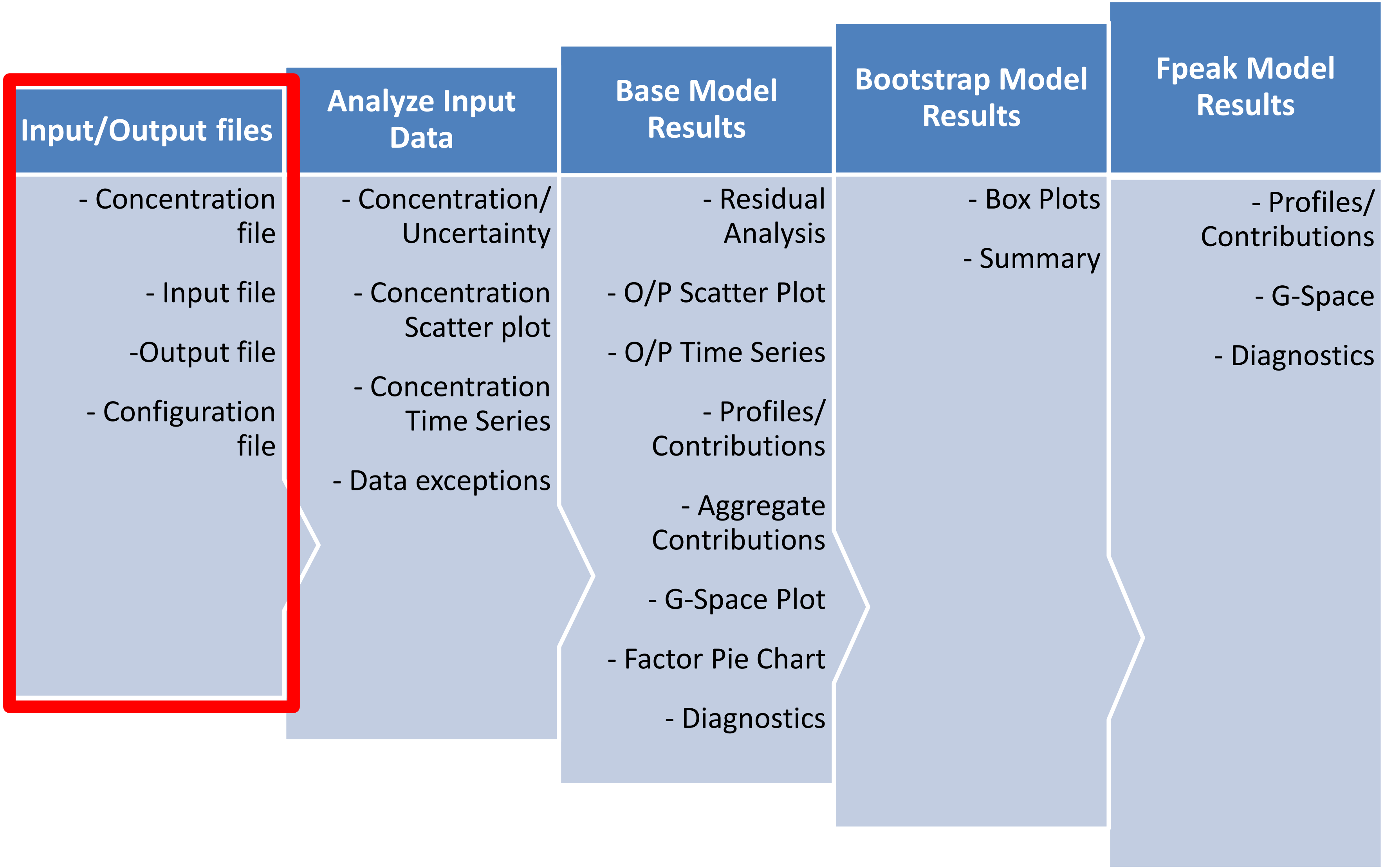
EPA PMF

Version 3.0.2.2

The United States Environmental Protection Agency through its Office of Research and Development funded and collaborated in the research described here under Contract Numbers EP-D-05-004 and 68-W-04-005 to Sonoma Technology, Inc. This software is now being subjected to external peer-review and is for evaluation purposes only. Portions of the code are Copyright©2005-2008 ExoAnalytics Inc. and Copyright©2007-2008 Bytescout.

run as administrator

Order of the operations



Input/Output files

- Concentration file
- Input file
- Output file
- Configuration file

Analyze Input Data

- Concentration/ Uncertainty
- Concentration Scatter plot
- Concentration Time Series
- Data exceptions

Base Model Results

- Residual Analysis
- O/P Scatter Plot
- O/P Time Series
- Profiles/ Contributions
- Aggregate Contributions
- G-Space Plot
- Factor Pie Chart
- Diagnostics

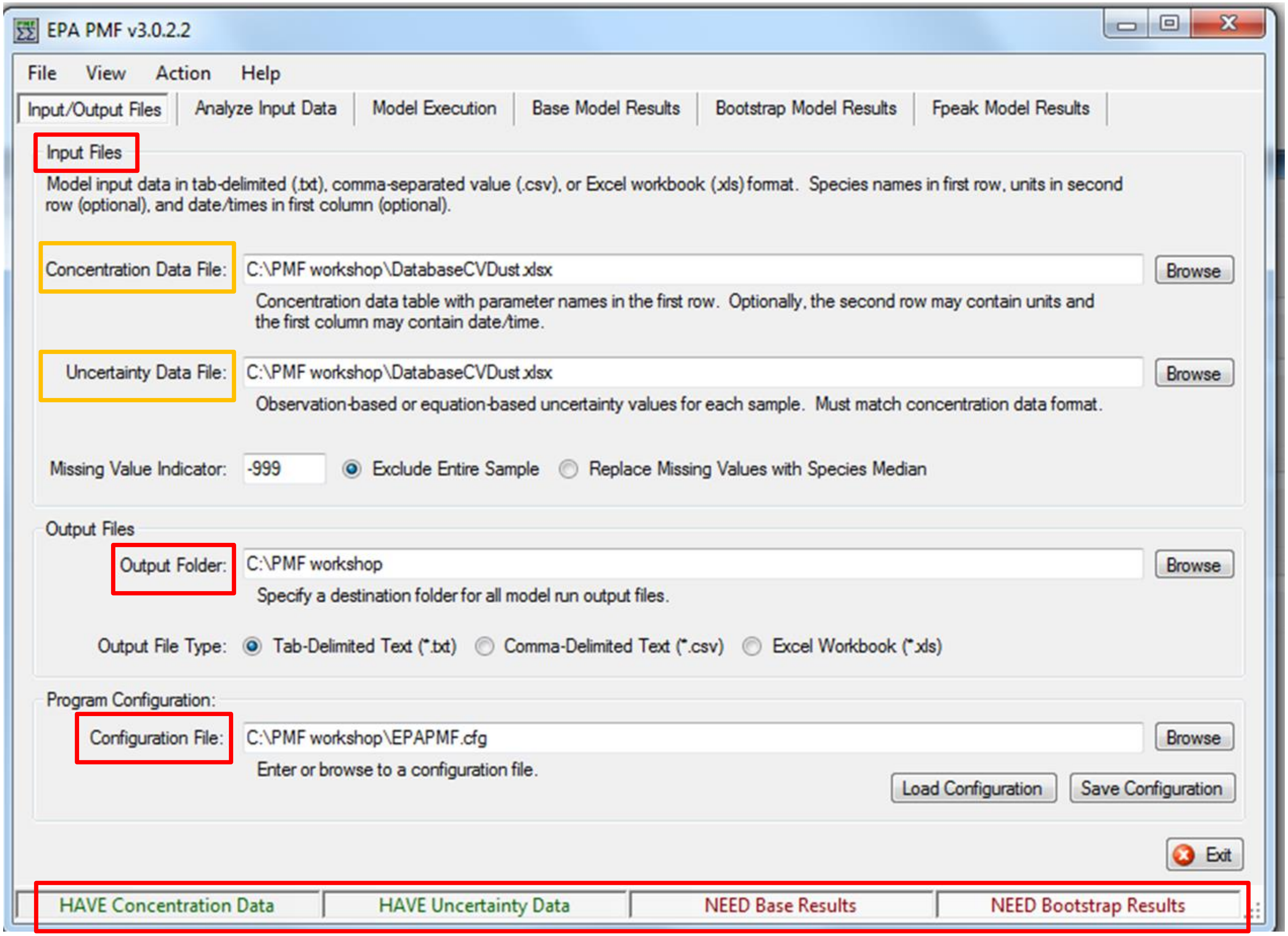
Bootstrap Model Results

- Box Plots
- Summary

Fpeak Model Results

- Profiles/ Contributions
- G-Space
- Diagnostics

Input/Output Files



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).

Concentration Data File: C:\PMF workshop\DatabaseCVDust.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.

Uncertainty Data File: C:\PMF workshop\DatabaseCVDust.xlsx

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

Missing Value Indicator: -999 ● Exclude Entire Sample ○ Replace Missing Values with Species Median

Output Files

Output Folder: C:\PMF workshop

Specify a destination folder for all model run output files.

Output File Type: ● Tab-Delimited Text (*.txt) ○ Comma-Delimited Text (*.csv) ○ Excel Workbook (*.xls)

Program Configuration:

Configuration File: C:\PMF workshop\EPAPMF.cfg

Enter or browse to a configuration file.

HAVE Concentration Data | HAVE Uncertainty Data | NEED Base Results | NEED Bootstrap Results

Input/Output Files

Input files

1. Concentration of chemical species
2. Uncertainties

Accepted formats

1. Tab-delimited (.txt)
2. Comma-separated value (.csv)
3. MS Excel (.xls)

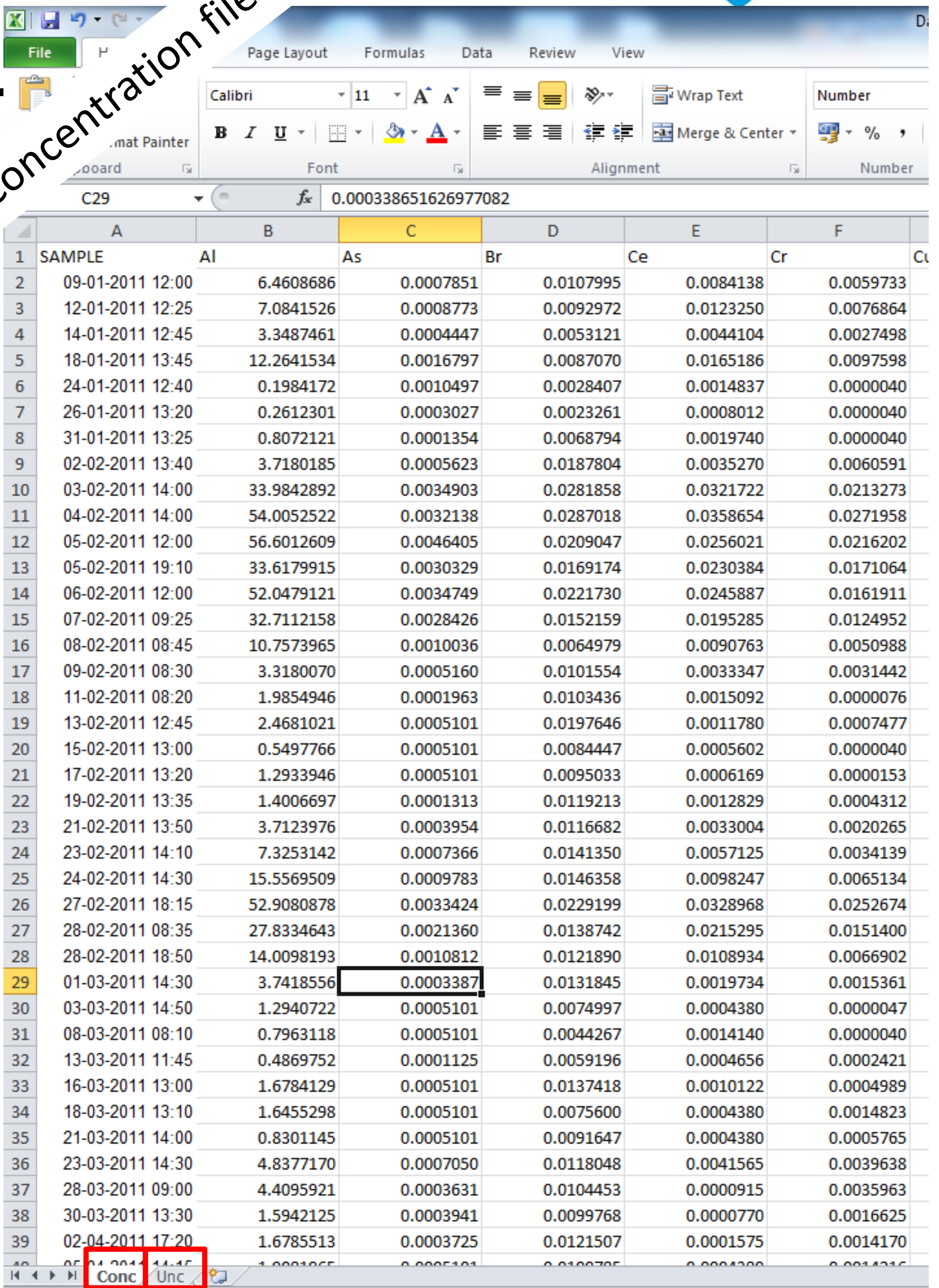
Blank cells are not accepted

Factor analysis technique provides a robust solution when the number of samples minus the number of variable is at least 30 (Henry, 1991)

Parameters

Concentration file

Samples



C29 f_x 0.000338651626977082

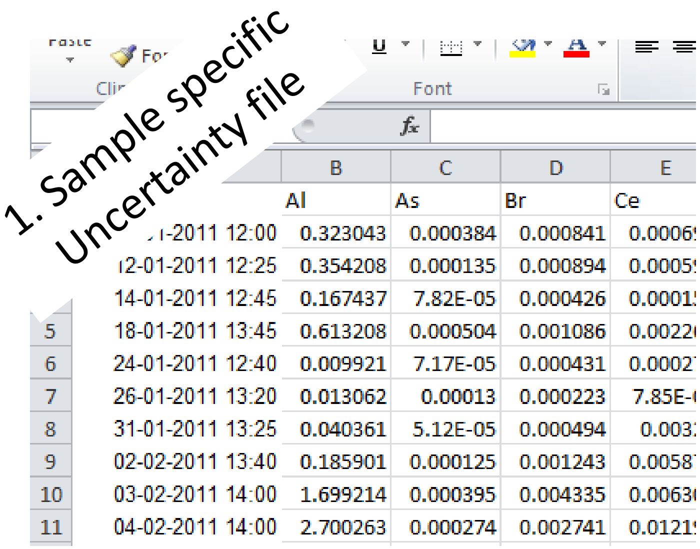
	A	B	C	D	E	F
1	SAMPLE	Al	As	Br	Ce	Cr
2	09-01-2011 12:00	6.4608686	0.0007851	0.0107995	0.0084138	0.0059733
3	12-01-2011 12:25	7.0841526	0.0008773	0.0092972	0.0123250	0.0076864
4	14-01-2011 12:45	3.3487461	0.0004447	0.0053121	0.0044104	0.0027498
5	18-01-2011 13:45	12.2641534	0.0016797	0.0087070	0.0165186	0.0097598
6	24-01-2011 12:40	0.1984172	0.0010497	0.0028407	0.0014837	0.0000040
7	26-01-2011 13:20	0.2612301	0.0003027	0.0023261	0.0008012	0.0000040
8	31-01-2011 13:25	0.8072121	0.0001354	0.0068794	0.0019740	0.0000040
9	02-02-2011 13:40	3.7180185	0.0005623	0.0187804	0.0035270	0.0060591
10	03-02-2011 14:00	33.9842892	0.0034903	0.0281858	0.0321722	0.0213273
11	04-02-2011 14:00	54.0052522	0.0032138	0.0287018	0.0358654	0.0271958
12	05-02-2011 12:00	56.6012609	0.0046405	0.0209047	0.0256021	0.0216202
13	05-02-2011 19:10	33.6179915	0.0030329	0.0169174	0.0230384	0.0171064
14	06-02-2011 12:00	52.0479121	0.0034749	0.0221730	0.0245887	0.0161911
15	07-02-2011 09:25	32.7112158	0.0028426	0.0152159	0.0195285	0.0124952
16	08-02-2011 08:45	10.7573965	0.0010036	0.0064979	0.0090763	0.0050988
17	09-02-2011 08:30	3.3180070	0.0005160	0.0101554	0.0033347	0.0031442
18	11-02-2011 08:20	1.9854946	0.0001963	0.0103436	0.0015092	0.0000076
19	13-02-2011 12:45	2.4681021	0.0005101	0.0197646	0.0011780	0.0007477
20	15-02-2011 13:00	0.5497766	0.0005101	0.0084447	0.0005602	0.0000040
21	17-02-2011 13:20	1.2933946	0.0005101	0.0095033	0.0006169	0.0000153
22	19-02-2011 13:35	1.4006697	0.0001313	0.0119213	0.0012829	0.0004312
23	21-02-2011 13:50	3.7123976	0.0003954	0.0116682	0.0033004	0.0020265
24	23-02-2011 14:10	7.3253142	0.0007366	0.0141350	0.0057125	0.0034139
25	24-02-2011 14:30	15.5569509	0.0009783	0.0146358	0.0098247	0.0065134
26	27-02-2011 18:15	52.9080878	0.0033424	0.0229199	0.0328968	0.0252674
27	28-02-2011 08:35	27.8334643	0.0021360	0.0138742	0.0215295	0.0151400
28	28-02-2011 18:50	14.0098193	0.0010812	0.0121890	0.0108934	0.0066902
29	01-03-2011 14:30	3.7418556	0.0003387	0.0131845	0.0019734	0.0015361
30	03-03-2011 14:50	1.2940722	0.0005101	0.0074997	0.0004380	0.0000047
31	08-03-2011 08:10	0.7963118	0.0005101	0.0044267	0.0014140	0.0000040
32	13-03-2011 11:45	0.4869752	0.0001125	0.0059196	0.0004656	0.0002421
33	16-03-2011 13:00	1.6784129	0.0005101	0.0137418	0.0010122	0.0004989
34	18-03-2011 13:10	1.6455298	0.0005101	0.0075600	0.0004380	0.0014823
35	21-03-2011 14:00	0.8301145	0.0005101	0.0091647	0.0004380	0.0005765
36	23-03-2011 14:30	4.8377170	0.0007050	0.0118048	0.0041565	0.0039638
37	28-03-2011 09:00	4.4095921	0.0003631	0.0104453	0.0000915	0.0035963
38	30-03-2011 13:30	1.5942125	0.0003941	0.0099768	0.0000770	0.0016625
39	02-04-2011 17:20	1.6785513	0.0003725	0.0121507	0.0001575	0.0014170

Input/Output Files

Types of Uncertainties files

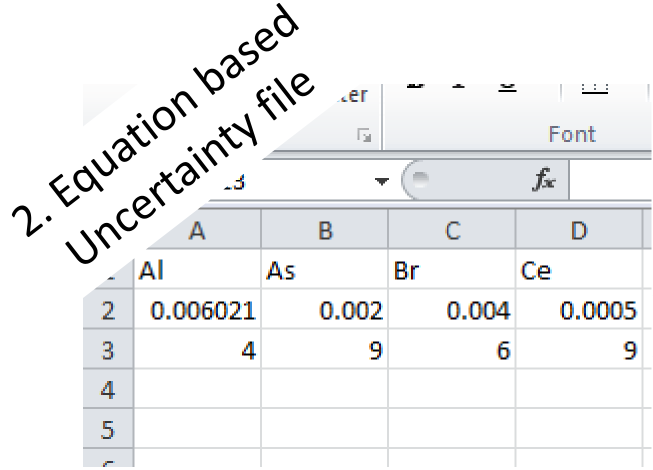
1. Sample specific – Uncertainty file provides na estimate of the uncertainty for each sample of each species;

2. Equation-based – file provides species-specific parameters that PMF uses to calculate uncertainties for each sample.



1. Sample specific Uncertainty file

	A	B	C	D	E
		Al	As	Br	Ce
	1-2011 12:00	0.323043	0.000384	0.000841	0.0006
	12-01-2011 12:25	0.354208	0.000135	0.000894	0.0005
	14-01-2011 12:45	0.167437	7.82E-05	0.000426	0.0001
5	18-01-2011 13:45	0.613208	0.000504	0.001086	0.0022
6	24-01-2011 12:40	0.009921	7.17E-05	0.000431	0.0002
7	26-01-2011 13:20	0.013062	0.00013	0.000223	7.85E-
8	31-01-2011 13:25	0.040361	5.12E-05	0.000494	0.003
9	02-02-2011 13:40	0.185901	0.000125	0.001243	0.0058
10	03-02-2011 14:00	1.699214	0.000395	0.004335	0.0063
11	04-02-2011 14:00	2.700263	0.000274	0.002741	0.0121



2. Equation based Uncertainty file

	A	B	C	D
	Al	As	Br	Ce
2	0.006021	0.002	0.004	0.0005
3	4	9	6	9
4				
5				

Method Detection Limit (MDL)

Error fraction (in percentage)

If concentration ≤ Detection limit => $Unc = \frac{5}{6} \times MDL$

If concentration ≤ Detection limit =>$Unc = \sqrt{((Error\ Fraction \times concentration)^2 + (MDL)^2}$

Input/Output Files

Missing values in the data base

	Concentration	Uncertainty
Concentration < MDL	½ MDL	5/6 MDL
Missing values	Geometric Mean	4 × Geometric Mean

Input/Output Files

Output Files	
*_diag	Contains a record of the **user inputs and model diagnostic** information
*_contrib	Contains the **contributions** for each base run
*_profile	Contains the **profiles** for each base run
*_resid	Contains the **residuals** for each base run
*_strength	Contains the **factor strength** for each nase run
*_profile_boot	Contains the **number of bootstrap runs** mapped to each base run, each **bootstrap profile** that was mapped to the base profile and all **bootstrap statistics**
*_fpeak	Contains the **profiles and contributions of each fpeak** run

Order of the operations

Input/Output files

- Concentration file
- Input file
- Output file
- Configuration file

Analyze Input Data

- Concentration/ Uncertainty
- Concentration Scatter plot
- Concentration Time Series
- Data exceptions

Base Model Results

- Residual Analysis
- O/P Scatter Plot
- O/P Time Series
- Profiles/ Contributions
- Aggregate Contributions
- G-Space Plot
- Factor Pie Chart
- Diagnostics

Bootstrap Model Results

- Box Plots
- Summary

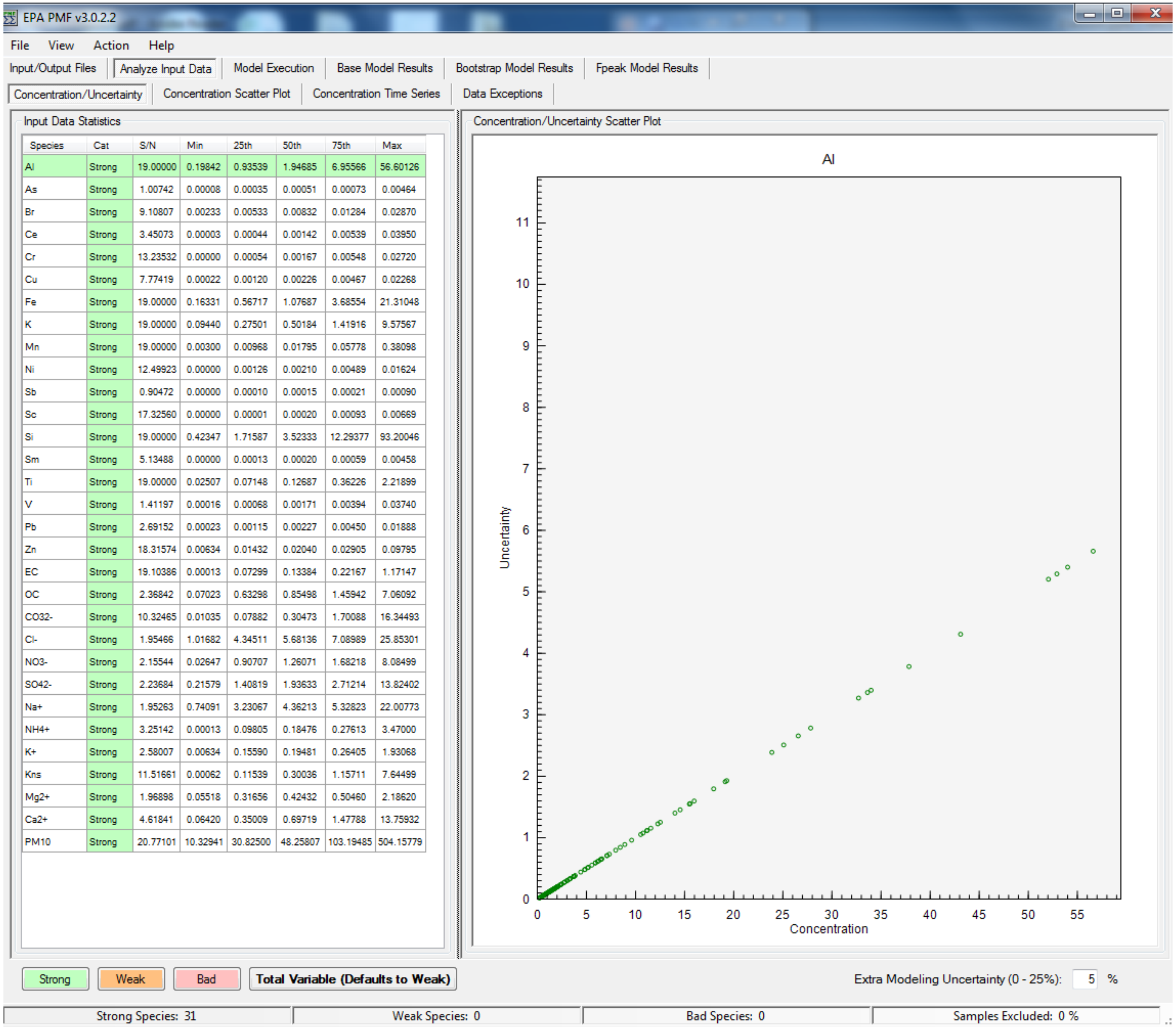
Fpeak Model Results

- Profiles/ Contributions
- G-Space
- Diagnostics

Analyze input data

Objective:

- To **analyze the concentration and uncertainty** data before running the model;
- Help the user decide whether **certain species should be excluded or down-weighted**;
- Help the user decide whether **certain samples should be excluded**.



Species	Cat	S/N	Min	25th	50th	75th	Max
Al	Strong	19.00000	0.19842	0.93539	1.94685	6.95566	56.60126
As	Strong	1.00742	0.00008	0.00035	0.00051	0.00073	0.00464
Br	Strong	9.10807	0.00233	0.00533	0.00832	0.01284	0.02870
Ce	Strong	3.45073	0.00003	0.00044	0.00142	0.00539	0.03950
Cr	Strong	13.23532	0.00000	0.00054	0.00167	0.00548	0.02720
Cu	Strong	7.77419	0.00022	0.00120	0.00226	0.00467	0.02268
Fe	Strong	19.00000	0.16331	0.56717	1.07687	3.68554	21.31048
K	Strong	19.00000	0.09440	0.27501	0.50184	1.41916	9.57567
Mn	Strong	19.00000	0.00300	0.00968	0.01795	0.05778	0.38098
Ni	Strong	12.49923	0.00000	0.00126	0.00210	0.00489	0.01624
Sb	Strong	0.90472	0.00000	0.00010	0.00015	0.00021	0.00090
Sc	Strong	17.32560	0.00000	0.00001	0.00020	0.00093	0.00669
Si	Strong	19.00000	0.42347	1.71587	3.52333	12.29377	93.20046
Sm	Strong	5.13488	0.00000	0.00013	0.00020	0.00059	0.00458
Ti	Strong	19.00000	0.02507	0.07148	0.12687	0.36226	2.21899
V	Strong	1.41197	0.00016	0.00068	0.00171	0.00394	0.03740
Pb	Strong	2.69152	0.00023	0.00115	0.00227	0.00450	0.01888
Zn	Strong	18.31574	0.00634	0.01432	0.02040	0.02905	0.09795
EC	Strong	19.10386	0.00013	0.07299	0.13384	0.22167	1.17147
OC	Strong	2.36842	0.07023	0.63298	0.85498	1.45942	7.06092
CO32-	Strong	10.32465	0.01035	0.07882	0.30473	1.70088	16.34493
Cl-	Strong	1.95466	1.01682	4.34511	5.68136	7.08989	25.85301
NO3-	Strong	2.15544	0.02647	0.90707	1.26071	1.68218	8.08499
SO42-	Strong	2.23684	0.21579	1.40819	1.93633	2.71214	13.82402
Na+	Strong	1.95263	0.74091	3.23067	4.36213	5.32823	22.00773
NH4+	Strong	3.25142	0.00013	0.09805	0.18476	0.27613	3.47000
K+	Strong	2.58007	0.00634	0.15590	0.19481	0.26405	1.93068
Kns	Strong	11.51661	0.00062	0.11539	0.30036	1.15711	7.64499
Mg2+	Strong	1.96898	0.05518	0.31656	0.42432	0.50460	2.18620
Ca2+	Strong	4.61841	0.06420	0.35009	0.69719	1.47788	13.75932
PM10	Strong	20.77101	10.32941	30.82500	48.25807	103.19485	504.15779

Extra Modeling Uncertainty (0 - 25%): 5 %

Strong Species: 31
Weak Species: 0
Bad Species: 0
Samples Excluded: 0 %

Analyze input data

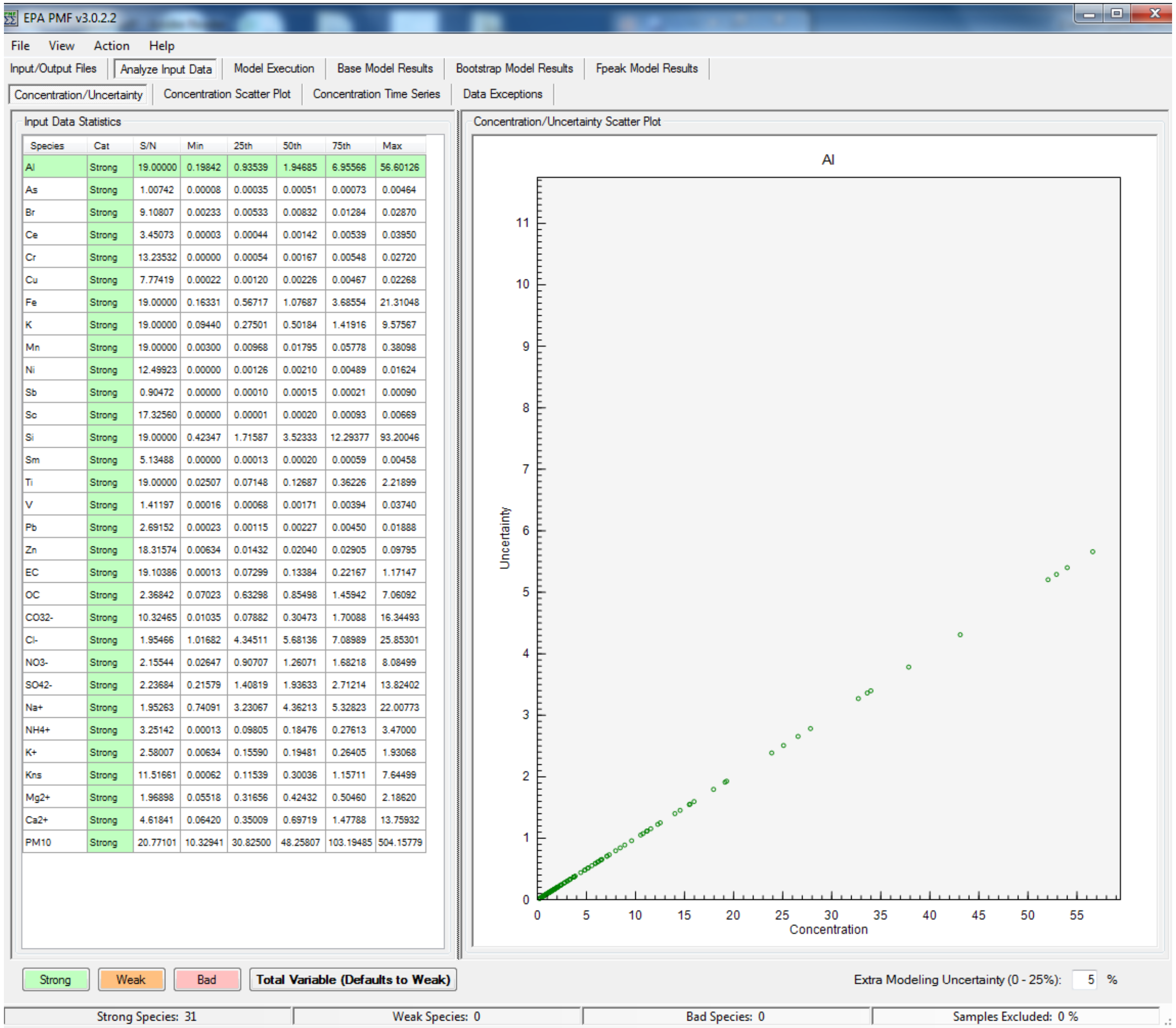
Concentration/Uncertainty

The following statistics are calculated for each species:

- Signal-to-noise ratio (S/N)

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

- Minimum concentration value
- 25th percentile
- Median
- 75th percentile
- Maximum concentration value



Input Data Statistics

Species	Cat	S/N	Min	25th	50th	75th	Max
Al	Strong	19.00000	0.19842	0.93539	1.94685	6.95566	56.60126
As	Strong	1.00742	0.00008	0.00035	0.00051	0.00073	0.00464
Br	Strong	9.10807	0.00233	0.00533	0.00832	0.01284	0.02870
Ce	Strong	3.45073	0.00003	0.00044	0.00142	0.00539	0.03950
Cr	Strong	13.23532	0.00000	0.00054	0.00167	0.00548	0.02720
Cu	Strong	7.77419	0.00022	0.00120	0.00226	0.00467	0.02268
Fe	Strong	19.00000	0.16331	0.56717	1.07687	3.68554	21.31048
K	Strong	19.00000	0.09440	0.27501	0.50184	1.41916	9.57567
Mn	Strong	19.00000	0.00300	0.00968	0.01795	0.05778	0.38098
Ni	Strong	12.49923	0.00000	0.00126	0.00210	0.00489	0.01624
Sb	Strong	0.90472	0.00000	0.00010	0.00015	0.00021	0.00090
Sc	Strong	17.32560	0.00000	0.00001	0.00020	0.00093	0.00669
Si	Strong	19.00000	0.42347	1.71587	3.52333	12.29377	93.20046
Sm	Strong	5.13488	0.00000	0.00013	0.00020	0.00059	0.00458
Ti	Strong	19.00000	0.02507	0.07148	0.12687	0.36226	2.21899
V	Strong	1.41197	0.00016	0.00068	0.00171	0.00394	0.03740
Pb	Strong	2.69152	0.00023	0.00115	0.00227	0.00450	0.01888
Zn	Strong	18.31574	0.00634	0.01432	0.02040	0.02905	0.09795
EC	Strong	19.10386	0.00013	0.07299	0.13384	0.22167	1.17147
OC	Strong	2.36842	0.07023	0.63298	0.85498	1.45942	7.06092
CO32-	Strong	10.32465	0.01035	0.07882	0.30473	1.70088	16.34493
Cl-	Strong	1.95466	1.01682	4.34511	5.68136	7.08989	25.85301
NO3-	Strong	2.15544	0.02647	0.90707	1.26071	1.68218	8.08499
SO42-	Strong	2.23684	0.21579	1.40819	1.93633	2.71214	13.82402
Na+	Strong	1.95263	0.74091	3.23067	4.36213	5.32823	22.00773
NH4+	Strong	3.25142	0.00013	0.09805	0.18476	0.27613	3.47000
K+	Strong	2.58007	0.00634	0.15590	0.19481	0.26405	1.93068
Kns	Strong	11.51661	0.00062	0.11539	0.30036	1.15711	7.64499
Mg2+	Strong	1.96898	0.05518	0.31656	0.42432	0.50460	2.18620
Ca2+	Strong	4.61841	0.06420	0.35009	0.69719	1.47788	13.75932
PM10	Strong	20.77101	10.32941	30.82500	48.25807	103.19485	504.15779

Extra Modeling Uncertainty (0 - 25%): 5 %

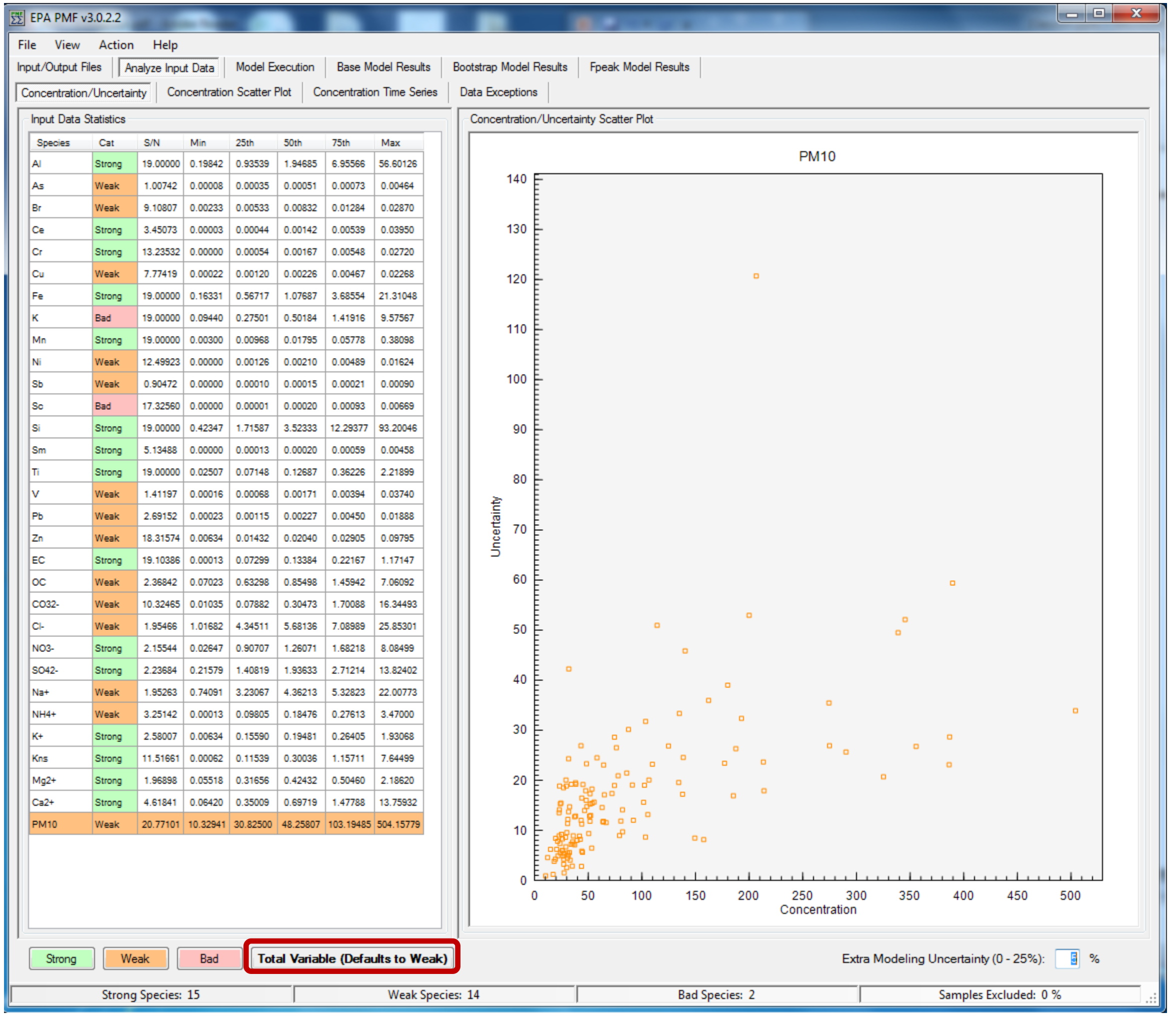
Strong Species: 31
Weak Species: 0
Bad Species: 0
Samples Excluded: 0 %

Analyze input data

Concentration/Uncertainty

Based on these statistics + knowledge of the data set the user can categorise a specie as:

S/N>2	STRONG	-
0.2<S/N<2	WEEK	uncertainty is tripled
S/N<0.2	BAD	Specie is excluded



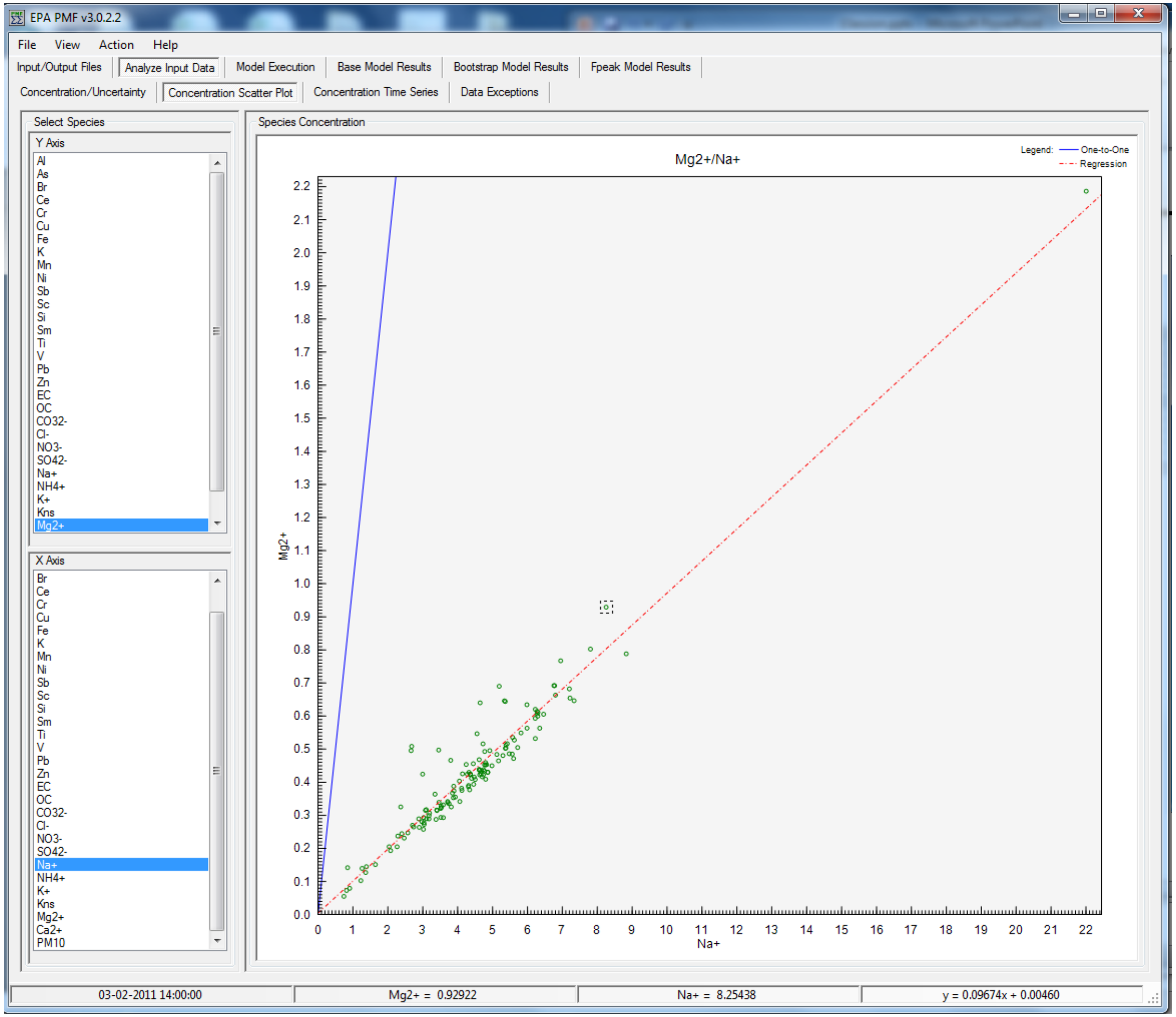
Species	Cat	S/N	Min	25th	50th	75th	Max
Al	Strong	19.00000	0.19842	0.93539	1.94685	6.95566	56.60126
As	Weak	1.00742	0.00008	0.00035	0.00051	0.00073	0.00464
Br	Weak	9.10807	0.00233	0.00533	0.00832	0.01284	0.02870
Ce	Strong	3.45073	0.00003	0.00044	0.00142	0.00539	0.03950
Cr	Strong	13.23532	0.00000	0.00054	0.00167	0.00548	0.02720
Cu	Weak	7.77419	0.00022	0.00120	0.00226	0.00467	0.02268
Fe	Strong	19.00000	0.16331	0.56717	1.07687	3.68554	21.31048
K	Bad	19.00000	0.09440	0.27501	0.50184	1.41916	9.57567
Mn	Strong	19.00000	0.00300	0.00968	0.01795	0.05778	0.38098
Ni	Weak	12.49923	0.00000	0.00126	0.00210	0.00489	0.01624
Sb	Weak	0.90472	0.00000	0.00010	0.00015	0.00021	0.00090
Sc	Bad	17.32560	0.00000	0.00001	0.00020	0.00093	0.00669
Si	Strong	19.00000	0.42347	1.71587	3.52333	12.29377	93.20046
Sm	Strong	5.13488	0.00000	0.00013	0.00020	0.00059	0.00458
Ti	Strong	19.00000	0.02507	0.07148	0.12687	0.36226	2.21899
V	Weak	1.41197	0.00016	0.00068	0.00171	0.00394	0.03740
Pb	Weak	2.69152	0.00023	0.00115	0.00227	0.00450	0.01888
Zn	Weak	18.31574	0.00634	0.01432	0.02040	0.02905	0.09795
EC	Strong	19.10386	0.00013	0.07299	0.13384	0.22167	1.17147
OC	Weak	2.36842	0.07023	0.63298	0.85498	1.45942	7.06092
CO32-	Weak	10.32465	0.01035	0.07882	0.30473	1.70088	16.34493
Cl-	Weak	1.95466	1.01682	4.34511	5.68136	7.08989	25.85301
NO3-	Strong	2.15544	0.02647	0.90707	1.26071	1.68218	8.08499
SO42-	Strong	2.23684	0.21579	1.40819	1.93633	2.71214	13.82402
Na+	Weak	1.95263	0.74091	3.23067	4.36213	5.32823	22.00773
NH4+	Weak	3.25142	0.00013	0.09805	0.18476	0.27613	3.47000
K+	Strong	2.58007	0.00634	0.15590	0.19481	0.26405	1.93068
Kns	Strong	11.51661	0.00062	0.11539	0.30036	1.15711	7.64499
Mg2+	Strong	1.96898	0.05518	0.31656	0.42432	0.50460	2.18620
Ca2+	Strong	4.61841	0.06420	0.35009	0.69719	1.47788	13.75932
PM10	Weak	20.77101	10.32941	30.82500	48.25807	103.19485	504.15779

Strong Species: 15
Weak Species: 14
Bad Species: 2
Samples Excluded: 0 %

Analyze input data

Scatter plots

- Useful pre-PMF analysis tool;
- The user should examine scatter plots to look for expected relations



03-02-2011 14:00:00

Mg2+ = 0.92922

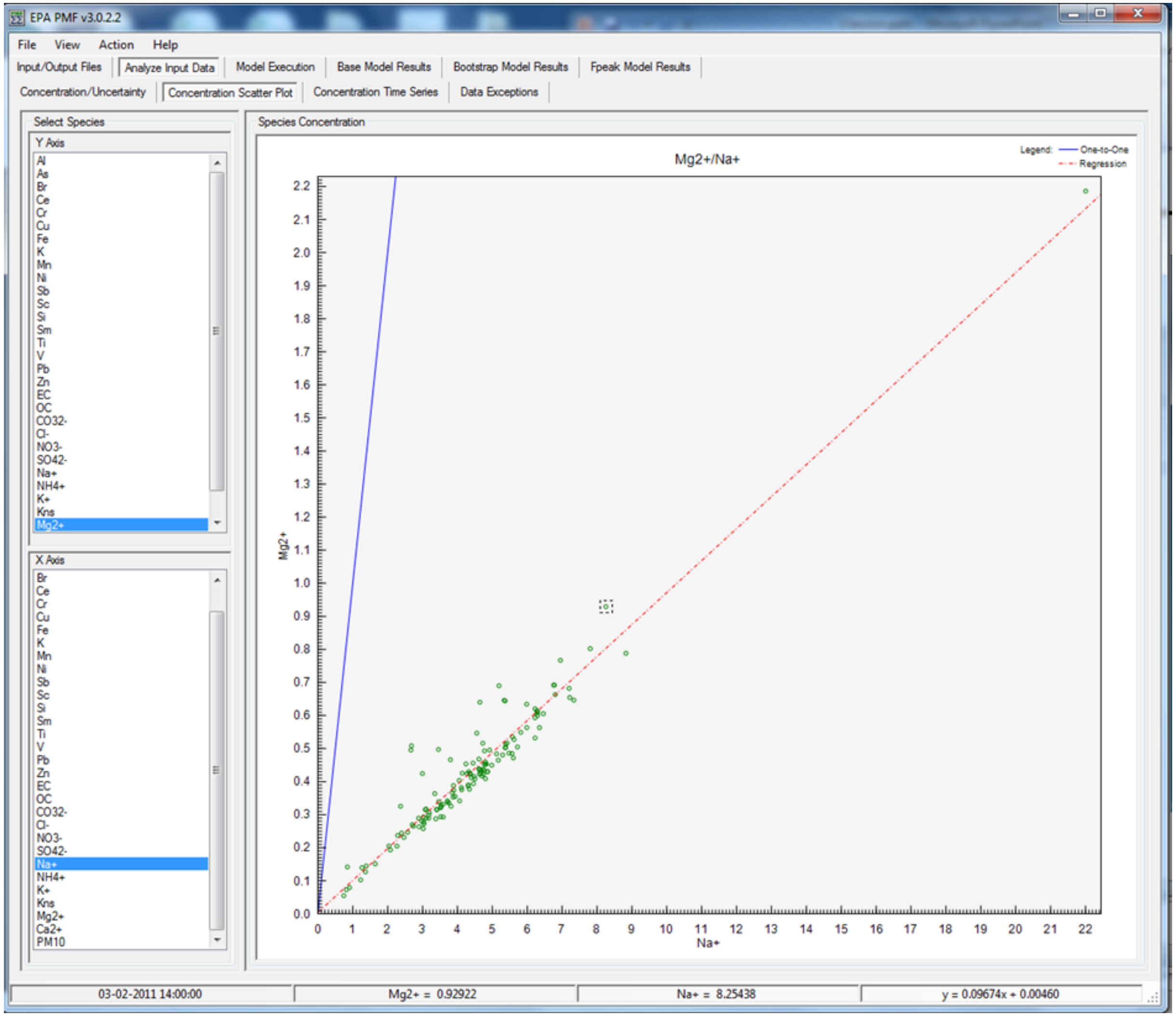
Na+ = 8.25438

y = 0.09674x + 0.00460

Analyze input data

Scatter plots

- Useful pre-PMF analysis tool;
- The user should examine scatter plots to look for expected relations

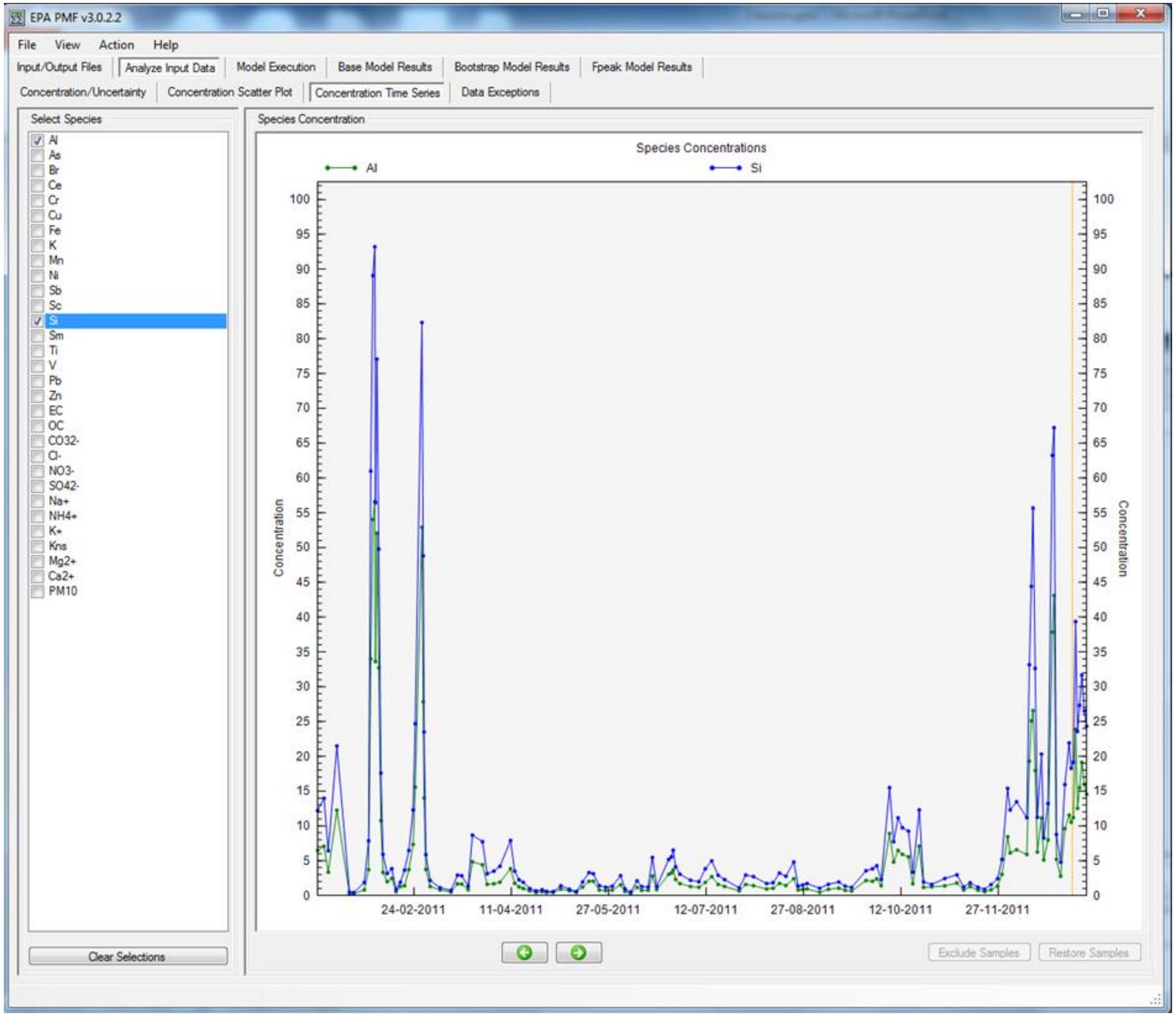


03-02-2011 14:00:00 | Mg2+ = 0.92922 | Na+ = 8.25438 | y = 0.09674x + 0.00460

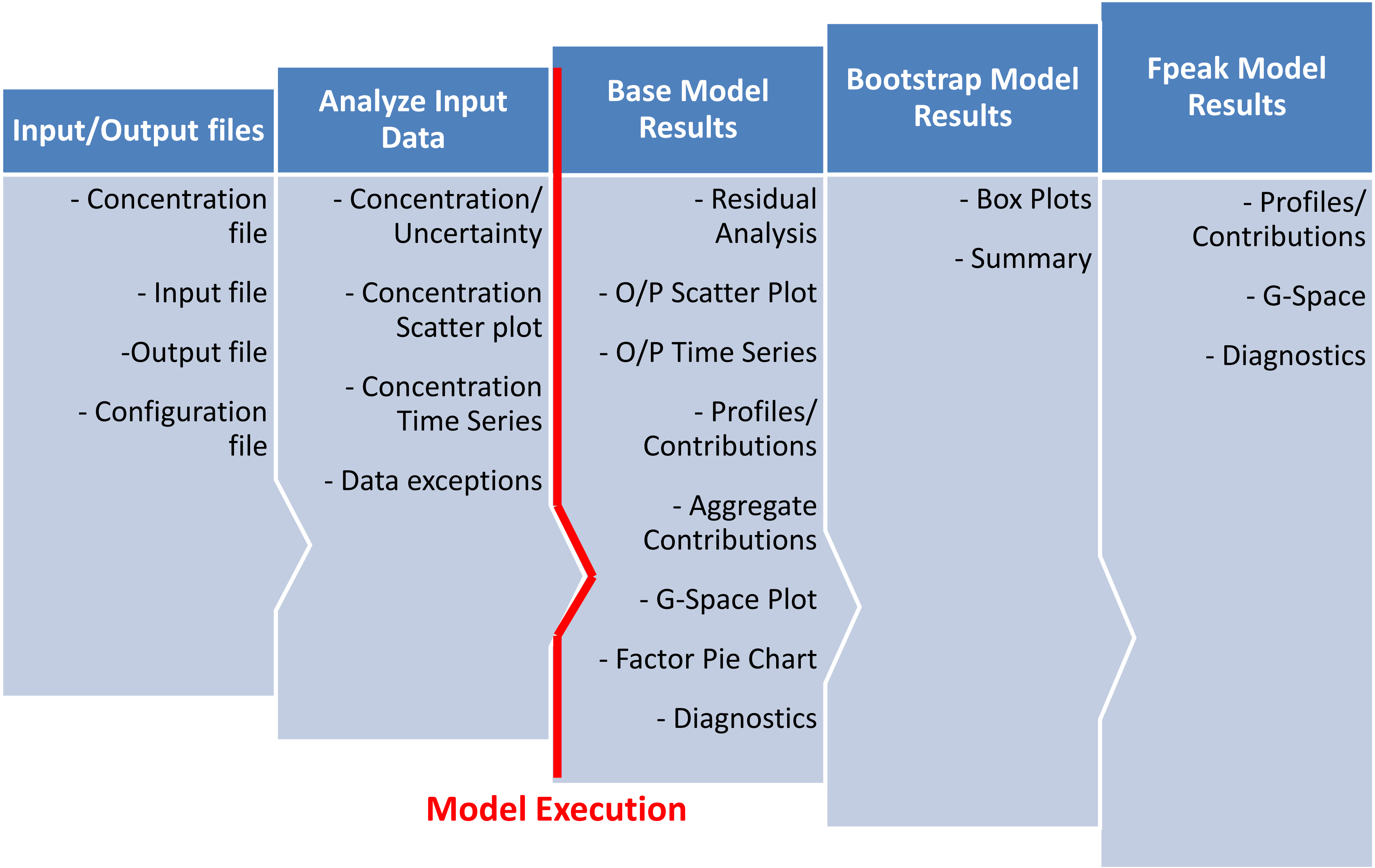
Analyze input data

Concentration Time Series

- Useful to detect unusual events;
- Help to exclude extreme events from the model.



Order of the operations



Input/Output files

- Concentration file
- Input file
- Output file
- Configuration file

Analyze Input Data

- Concentration/ Uncertainty
- Concentration Scatter plot
- Concentration Time Series
- Data exceptions

Base Model Results

- Residual Analysis
- O/P Scatter Plot
- O/P Time Series
- Profiles/ Contributions
- Aggregate Contributions
- G-Space Plot
- Factor Pie Chart
- Diagnostics

Bootstrap Model Results

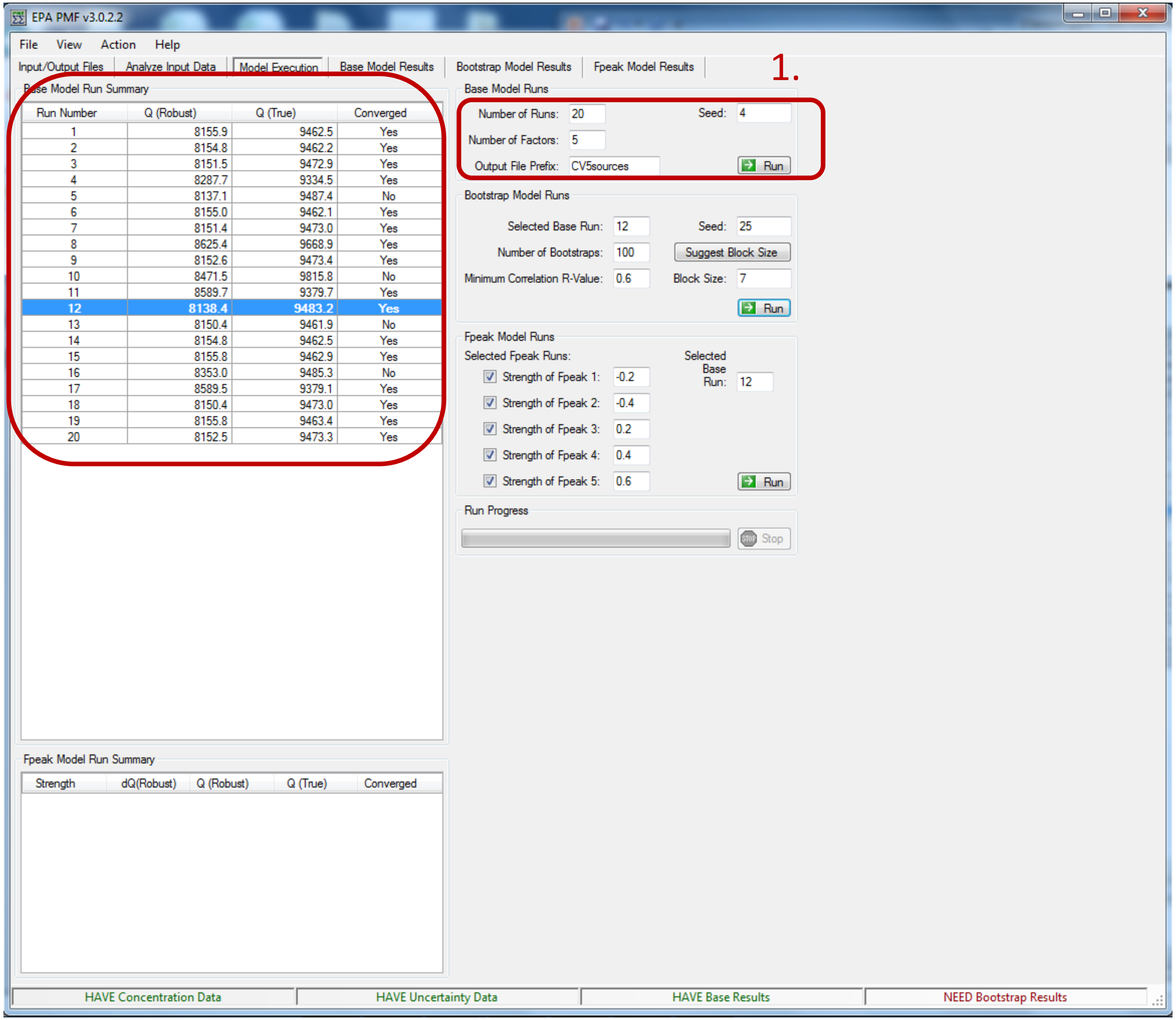
- Box Plots
- Summary

Fpeak Model Results

- Profiles/ Contributions
- G-Space
- Diagnostics

Model Execution

Model execution



2.

Base Model Run Summary

Run Number	Q (Robust)	Q (True)	Converged
1	8155.9	9462.5	Yes
2	8154.8	9462.2	Yes
3	8151.5	9472.9	Yes
4	8287.7	9334.5	Yes
5	8137.1	9487.4	No
6	8155.0	9462.1	Yes
7	8151.4	9473.0	Yes
8	8625.4	9668.9	Yes
9	8152.6	9473.4	Yes
10	8471.5	9815.8	No
11	8589.7	9379.7	Yes
12	8138.4	9483.2	Yes
13	8150.4	9461.9	No
14	8154.8	9462.5	Yes
15	8155.8	9462.9	Yes
16	8353.0	9485.3	No
17	8589.5	9379.1	Yes
18	8150.4	9473.0	Yes
19	8155.8	9463.4	Yes
20	8152.5	9473.3	Yes

Fpeak Model Run Summary

Strength	dQ(Robust)	Q (Robust)	Q (True)	Converged

1.

Base Model Runs

Number of Runs: 20
Seed: 4
Number of Factors: 5
Output File Prefix: CV5sources

Bootstrap Model Runs

Selected Base Run: 12
Seed: 25
Number of Bootstraps: 100
Minimum Correlation R-Value: 0.6
Block Size: 7

Fpeak Model Runs

Selected Fpeak Runs:
☑ Strength of Fpeak 1: -0.2
☑ Strength of Fpeak 2: -0.4
☑ Strength of Fpeak 3: 0.2
☑ Strength of Fpeak 4: 0.4
☑ Strength of Fpeak 5: 0.6
Selected Base Run: 12

HAVE Concentration Data
HAVE Uncertainty Data
HAVE Base Results
NEED Bootstrap Results

Model execution

Base run summary

Q values - assessment of how well the model fit the input data

- Q (robust) – calculated excluding outliers
- Q (true) – calculated including all points
- Theoretical Q ≈ nm-p(n+m)   (n - no. of species; m – no. of samples; p - no. of factors)

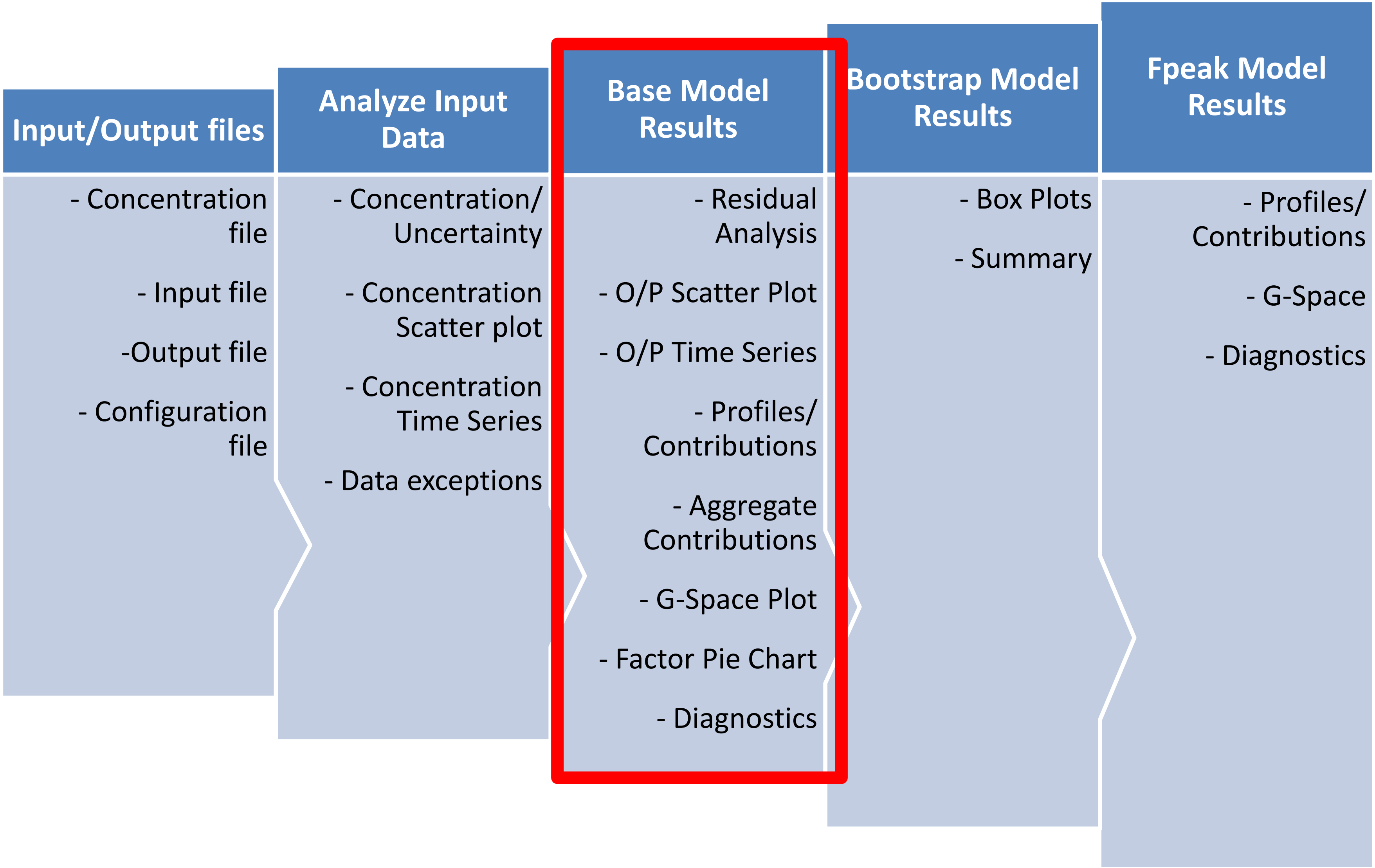
✓ Q (true)>1.5 Q (robust) => indicate that peak events may be disproportionately influencing the model ;

✓ Theoretical Q ≈ Q(true) and Q (robust)

Only converged solutions should be investigated further

(Causes for non-convergence - Uncertainties too low, specified incorrectly or inappropriate input parameters)

Order of the operations



Input/Output files
- Concentration file
- Input file
- Output file
- Configuration file

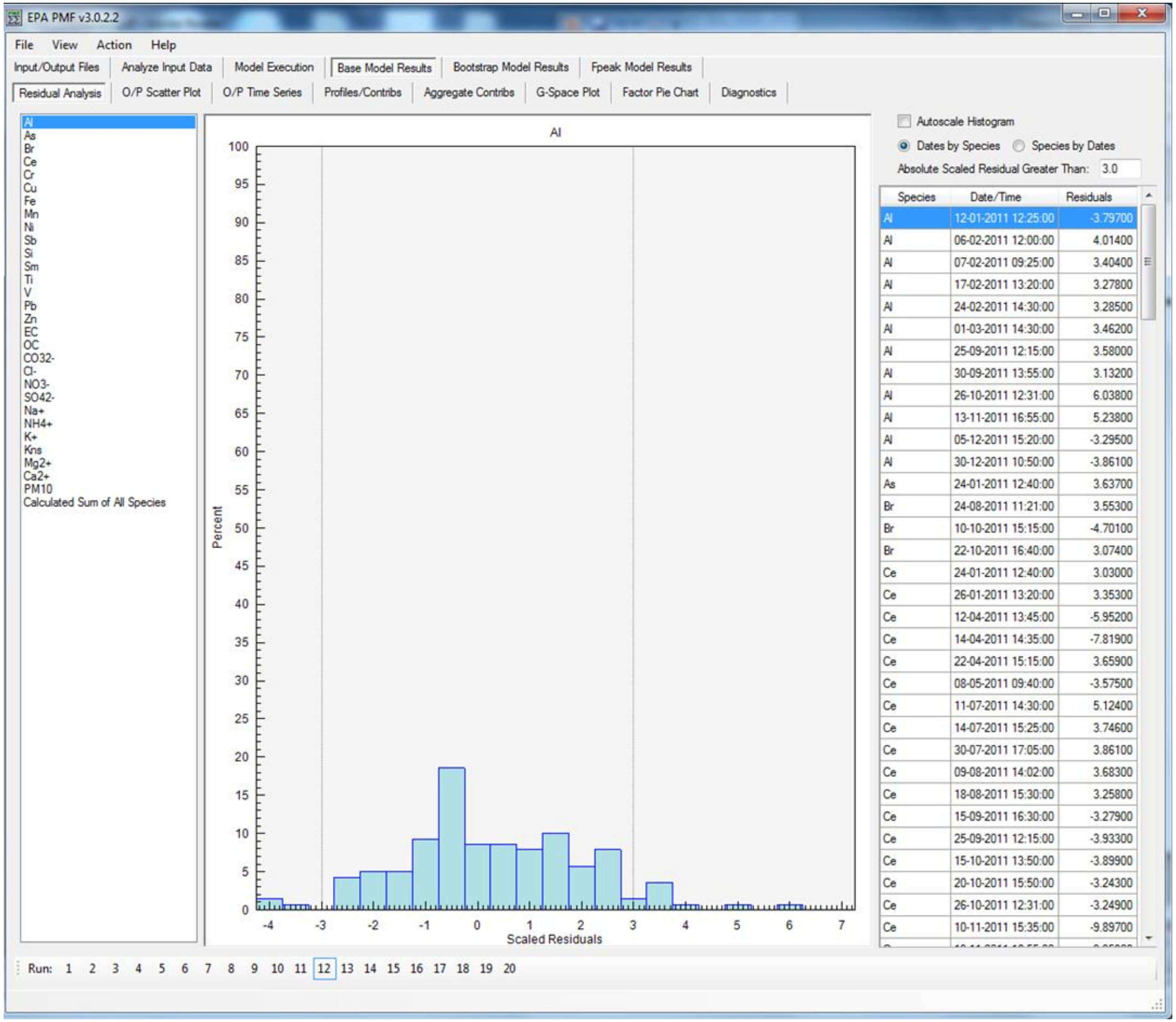
Analyze Input Data
- Concentration/ Uncertainty
- Concentration Scatter plot
- Concentration Time Series
- Data exceptions

Base Model Results
- Residual Analysis
- O/P Scatter Plot
- O/P Time Series
- Profiles/ Contributions
- Aggregate Contributions
- G-Space Plot
- Factor Pie Chart
- Diagnostics

Bootstrap Model Results
- Box Plots
- Summary

Fpeak Model Results
- Profiles/ Contributions
- G-Space
- Diagnostics

Base model results



☐ Autoscale Histogram

● Dates by Species ○ Species by Dates

Absolute Scaled Residual Greater Than: 3.0

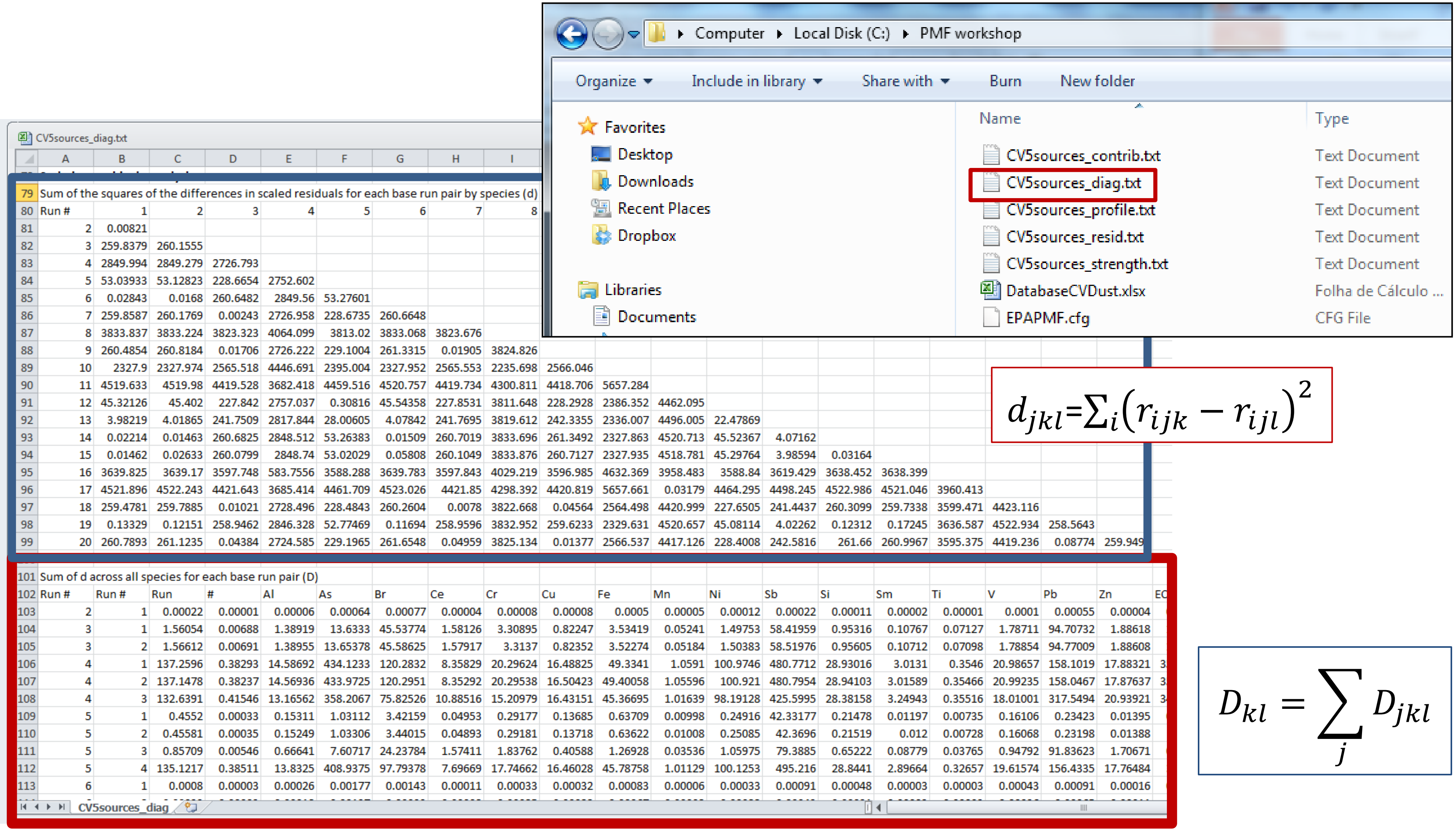
Species	Date/Time	Residuals
Al	12-01-2011 12:25:00	-3.79700
Al	06-02-2011 12:00:00	4.01400
Al	07-02-2011 09:25:00	3.40400
Al	17-02-2011 13:20:00	3.27800
Al	24-02-2011 14:30:00	3.28500
Al	01-03-2011 14:30:00	3.46200
Al	25-09-2011 12:15:00	3.58000
Al	30-09-2011 13:55:00	3.13200
Al	26-10-2011 12:31:00	6.03800
Al	13-11-2011 16:55:00	5.23800
Al	05-12-2011 15:20:00	-3.29500
Al	30-12-2011 10:50:00	-3.86100
As	24-01-2011 12:40:00	3.63700
Br	24-08-2011 11:21:00	3.55300
Br	10-10-2011 15:15:00	-4.70100
Br	22-10-2011 16:40:00	3.07400
Ce	24-01-2011 12:40:00	3.03000
Ce	26-01-2011 13:20:00	3.35300
Ce	12-04-2011 13:45:00	-5.95200
Ce	14-04-2011 14:35:00	-7.81900
Ce	22-04-2011 15:15:00	3.65900
Ce	08-05-2011 09:40:00	-3.57500
Ce	11-07-2011 14:30:00	5.12400
Ce	14-07-2011 15:25:00	3.74600
Ce	30-07-2011 17:05:00	3.86100
Ce	09-08-2011 14:02:00	3.68300
Ce	18-08-2011 15:30:00	3.25800
Ce	15-09-2011 16:30:00	-3.27900
Ce	25-09-2011 12:15:00	-3.93300
Ce	15-10-2011 13:50:00	-3.89900
Ce	20-10-2011 15:50:00	-3.24300
Ce	26-10-2011 12:31:00	-3.24900
Ce	10-11-2011 15:35:00	-9.89700

Residuals

Objective:

- To determine how well the model fit each species
- Poor fit when:
    - large residuals (|r|>3)
    - Non-normal curve

Base model results



Name	Type
CV5sources_contrib.txt	Text Document
CV5sources_diag.txt	Text Document
CV5sources_profile.txt	Text Document
CV5sources_resid.txt	Text Document
CV5sources_strength.txt	Text Document
DatabaseCVDust.xlsx	Folha de Cálculo ...
EPAPMF.cfg	CFG File

CV5sources_diag.txt

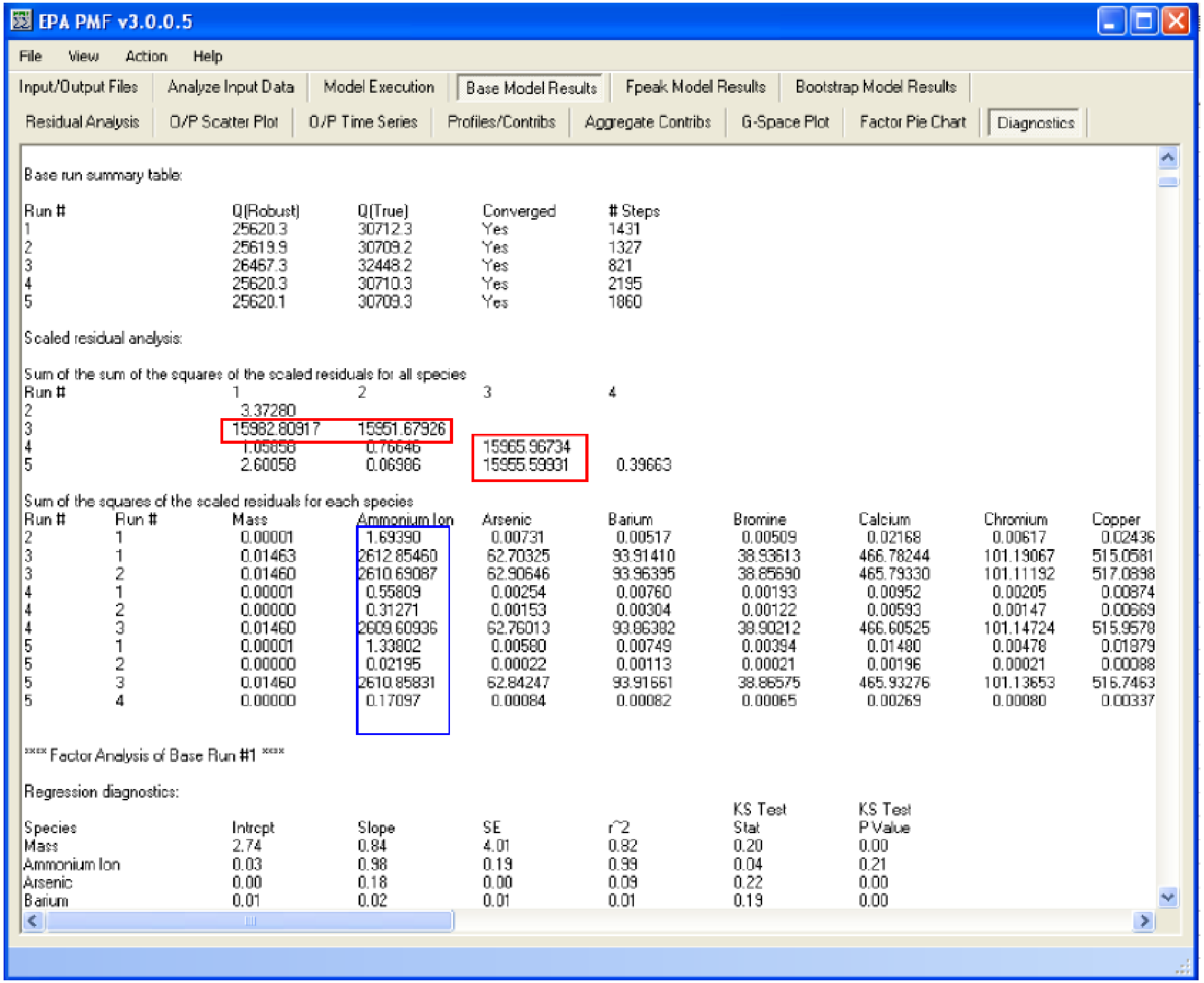
	A	B	C	D	E	F	G	H	I											
79	Sum of the squares of the differences in scaled residuals for each base run pair by species (d)																			
80	Run #	1	2	3	4	5	6	7	8											
81	2	0.00821																		
82	3	259.8379	260.1555																	
83	4	2849.994	2849.279	2726.793																
84	5	53.03933	53.12823	228.6654	2752.602															
85	6	0.02843	0.0168	260.6482	2849.56	53.27601														
86	7	259.8587	260.1769	0.00243	2726.958	228.6735	260.6648													
87	8	3833.837	3833.224	3823.323	4064.099	3813.02	3833.068	3823.676												
88	9	260.4854	260.8184	0.01706	2726.222	229.1004	261.3315	0.01905	3824.826											
89	10	2327.9	2327.974	2565.518	4446.691	2395.004	2327.952	2565.553	2235.698	2566.046										
90	11	4519.633	4519.98	4419.528	3682.418	4459.516	4520.757	4419.734	4300.811	4418.706	5657.284									
91	12	45.32126	45.402	227.842	2757.037	0.30816	45.54358	227.8531	3811.648	228.2928	2386.352	4462.095								
92	13	3.98219	4.01865	241.7509	2817.844	28.00605	4.07842	241.7695	3819.612	242.3355	2336.007	4496.005	22.47869							
93	14	0.02214	0.01463	260.6825	2848.512	53.26383	0.01509	260.7019	3833.696	261.3492	2327.863	4520.713	45.52367	4.07162						
94	15	0.01462	0.02633	260.0799	2848.74	53.02029	0.05808	260.1049	3833.876	260.7127	2327.935	4518.781	45.29764	3.98594	0.03164					
95	16	3639.825	3639.17	3597.748	583.7556	3588.288	3639.783	3597.843	4029.219	3596.985	4632.369	3958.483	3588.84	3619.429	3638.452	3638.399				
96	17	4521.896	4522.243	4421.643	3685.414	4461.709	4523.026	4421.85	4298.392	4420.819	5657.661	0.03179	4464.295	4498.245	4522.986	4521.046	3960.413			
97	18	259.4781	259.7885	0.01021	2728.496	228.4843	260.2604	0.0078	3822.668	0.04564	2564.498	4420.999	227.6505	241.4437	260.3099	259.7338	3599.471	4423.116		
98	19	0.13329	0.12151	258.9462	2846.328	52.77469	0.11694	258.9596	3832.952	259.6233	2329.631	4520.657	45.08114	4.02262	0.12312	0.17245	3636.587	4522.934	258.5643	
99	20	260.7893	261.1235	0.04384	2724.585	229.1965	261.6548	0.04959	3825.134	0.01377	2566.537	4417.126	228.4008	242.5816	261.66	260.9967	3595.375	4419.236	0.08774	259.949

101	Sum of d across all species for each base run pair (D)																				
102	Run #	Run #	Run	#	Al	As	Br	Ce	Cr	Cu	Fe	Mn	Ni	Sb	Si	Sm	Ti	V	Pb	Zn	EC
103	2	1	0.00022	0.00001	0.00006	0.00064	0.00077	0.00004	0.00008	0.00008	0.0005	0.00005	0.00012	0.00022	0.00011	0.00002	0.00001	0.0001	0.00055	0.00004	
104	3	1	1.56054	0.00688	1.38919	13.6333	45.53774	1.58126	3.30895	0.82247	3.53419	0.05241	1.49753	58.41959	0.95316	0.10767	0.07127	1.78711	94.70732	1.88618	
105	3	2	1.56612	0.00691	1.38955	13.65378	45.58625	1.57917	3.3137	0.82352	3.52274	0.05184	1.50383	58.51976	0.95605	0.10712	0.07098	1.78854	94.77009	1.88608	
106	4	1	137.2596	0.38293	14.58692	434.1233	120.2832	8.35829	20.29624	16.48825	49.3341	1.0591	100.9746	480.7712	28.93016	3.0131	0.3546	20.98657	158.1019	17.88321	
107	4	2	137.1478	0.38237	14.56936	433.9725	120.2951	8.35292	20.29538	16.50423	49.40058	1.05596	100.921	480.7954	28.94103	3.01589	0.35466	20.99235	158.0467	17.87637	
108	4	3	132.6391	0.41546	13.16562	358.2067	75.82526	10.88516	15.20979	16.43151	45.36695	1.01639	98.19128	425.5995	28.38158	3.24943	0.35516	18.01001	317.5494	20.93921	
109	5	1	0.4552	0.00033	0.15311	1.03112	3.42159	0.04953	0.29177	0.13685	0.63709	0.00998	0.24916	42.33177	0.21478	0.01197	0.00735	0.16106	0.23423	0.01395	
110	5	2	0.45581	0.00035	0.15249	1.03306	3.44015	0.04893	0.29181	0.13718	0.63622	0.01008	0.25085	42.3696	0.21519	0.012	0.00728	0.16068	0.23198	0.01388	
111	5	3	0.85709	0.00546	0.66641	7.60717	24.23784	1.57411	1.83762	0.40588	1.26928	0.03536	1.05975	79.3885	0.65222	0.08779	0.03765	0.94792	91.83623	1.70671	
112	5	4	135.1217	0.38511	13.8325	408.9375	97.79378	7.69669	17.74662	16.46028	45.78758	1.01129	100.1253	495.216	28.8441	2.89664	0.32657	19.61574	156.4335	17.76484	
113	6	1	0.0008	0.00003	0.00026	0.00177	0.00143	0.00011	0.00033	0.00032	0.00083	0.00006	0.00033	0.00091	0.00048	0.00003	0.00003	0.00043	0.00091	0.00016	

$$d_{jkl}=\sum_i\left(r_{ijk}-r_{ijl}\right)^2$$

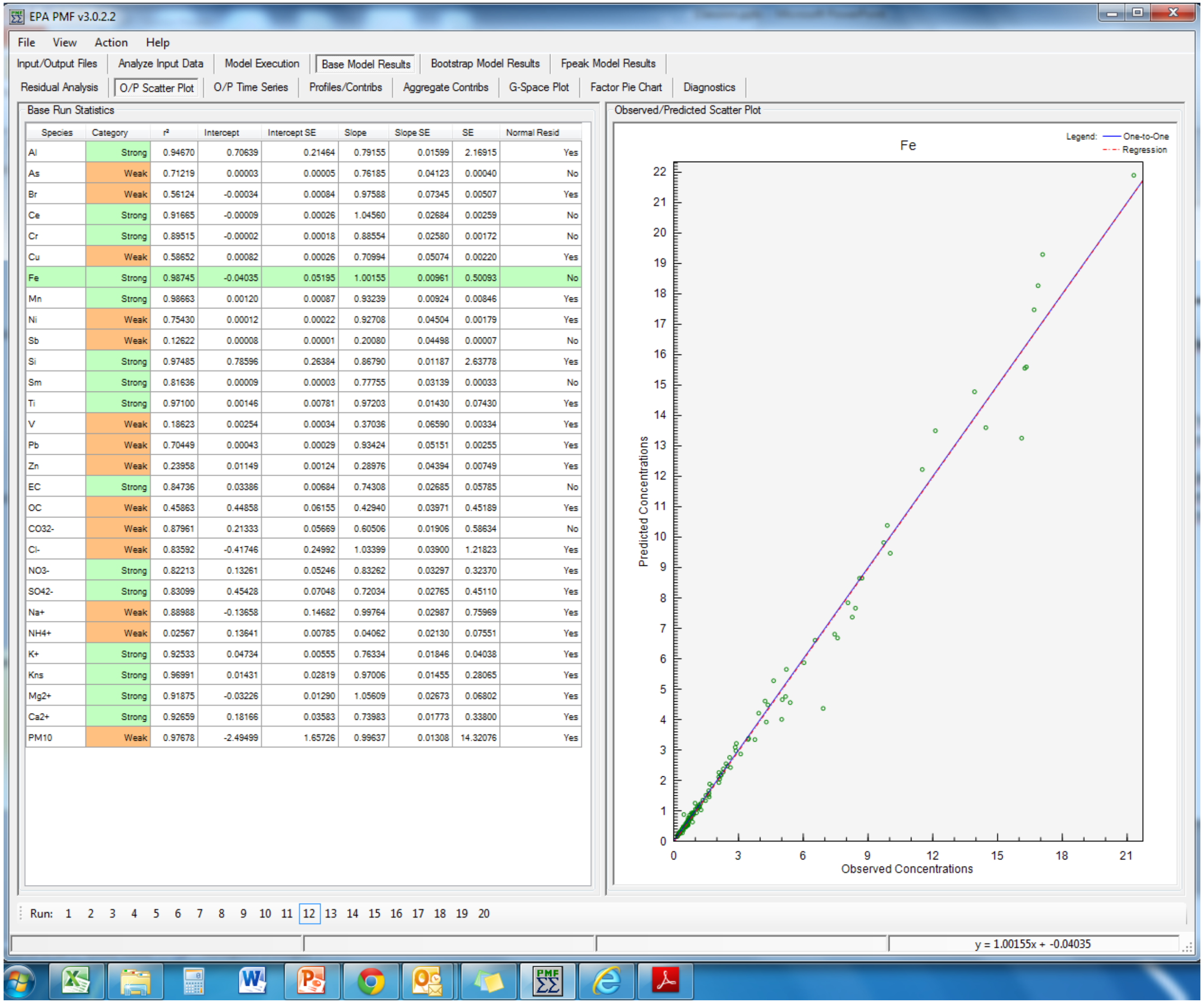
$$D_{kl}=\sum_j D_{jkl}$$

Base model results



```
Base run summary table:

Run #               Q(Robust)     Q(True)      Converged     # Steps
1                   25620.3       30712.3      Yes           1431
2                   25619.9       30709.2      Yes           1327
3                   26467.3       32448.2      Yes           821
4                   25620.3       30710.3      Yes           2195
5                   25620.1       30709.3      Yes           1860

Scaled residual analysis:

Sum of the sum of the squares of the scaled residuals for all species
Run #               1             2            3             4
2                    3.37280
3                   15982.80917   15951.67926
4                    1.05858       0.76646     15965.96734
5                    2.60058       0.06986     15955.59931    0.39663

Sum of the squares of the scaled residuals for each species
Run #    Run #      Mass          Ammonium Ion  Arsenic      Barium       Bromine      Calcium      Chromium     Copper
2        1           0.00001       1.69390       0.00731      0.00517      0.00509      0.02168      0.00617      0.02436
3        1           0.01463      2612.85460    62.70325     93.91410     38.93613     466.78244    101.19067    515.0581
3        2           0.01460      2610.69087    62.90646     93.96395     38.85690     465.79330    101.11192    517.0898
4        1           0.00001       0.55809       0.00254      0.00760      0.00193      0.00952      0.00205      0.00874
4        2           0.00000       0.31271       0.00153      0.00304      0.00122      0.00593      0.00147      0.00669
4        3           0.01460      2609.60936    62.76013     93.86382     38.90212     466.60525    101.14724    515.9578
5        1           0.00001       1.33802       0.00580      0.00749      0.00394      0.01480      0.00478      0.01879
5        2           0.00000       0.02195       0.00022      0.00113      0.00021      0.00196      0.00021      0.00088
5        3           0.01460      2610.85831    62.84247     93.91661     38.86575     465.93276    101.13653    516.7463
5        4           0.00000       0.17097       0.00084      0.00082      0.00065      0.00269      0.00080      0.00337


**** Factor Analysis of Base Run #1 ****

Regression diagnostics:

                                                                        KS Test      KS Test
Species             Intrcpt       Slope        SE            r^2        Stat         P Value
Mass                2.74          0.84         4.01          0.82       0.20         0.00
Ammonium Ion        0.03          0.98         0.19          0.99       0.04         0.21
Arsenic             0.00          0.18         0.00          0.09       0.22         0.00
Barium              0.01          0.02         0.01          0.01       0.19         0.00
```

Base model results



Base Run Statistics

Species	Category	r²	Intercept	Intercept SE	Slope	Slope SE	SE	Normal Resid
Al	Strong	0.94670	0.70639	0.21464	0.79155	0.01599	2.16915	Yes
As	Weak	0.71219	0.00003	0.00005	0.76185	0.04123	0.00040	No
Br	Weak	0.56124	-0.00034	0.00084	0.97588	0.07345	0.00507	Yes
Ce	Strong	0.91665	-0.00009	0.00026	1.04560	0.02684	0.00259	No
Cr	Strong	0.89515	-0.00002	0.00018	0.88554	0.02580	0.00172	No
Cu	Weak	0.58652	0.00082	0.00026	0.70994	0.05074	0.00220	Yes
Fe	Strong	0.98745	-0.04035	0.05195	1.00155	0.00961	0.50093	No
Mn	Strong	0.98663	0.00120	0.00087	0.93239	0.00924	0.00846	Yes
Ni	Weak	0.75430	0.00012	0.00022	0.92708	0.04504	0.00179	Yes
Sb	Weak	0.12622	0.00008	0.00001	0.20080	0.04498	0.00007	No
Si	Strong	0.97485	0.78596	0.26384	0.86790	0.01187	2.63778	Yes
Sm	Strong	0.81636	0.00009	0.00003	0.77755	0.03139	0.00033	No
Ti	Strong	0.97100	0.00146	0.00781	0.97203	0.01430	0.07430	Yes
V	Weak	0.18623	0.00254	0.00034	0.37036	0.06590	0.00334	Yes
Pb	Weak	0.70449	0.00043	0.00029	0.93424	0.05151	0.00255	Yes
Zn	Weak	0.23958	0.01149	0.00124	0.28976	0.04394	0.00749	Yes
EC	Strong	0.84736	0.03386	0.00684	0.74308	0.02685	0.05785	No
OC	Weak	0.45863	0.44858	0.06155	0.42940	0.03971	0.45189	Yes
CO32-	Weak	0.87961	0.21333	0.05669	0.60506	0.01906	0.58634	No
Cl-	Weak	0.83592	-0.41746	0.24992	1.03399	0.03900	1.21823	Yes
NO3-	Strong	0.82213	0.13261	0.05246	0.83262	0.03297	0.32370	Yes
SO42-	Strong	0.83099	0.45428	0.07048	0.72034	0.02765	0.45110	Yes
Na+	Weak	0.88988	-0.13658	0.14682	0.99764	0.02987	0.75969	Yes
NH4+	Weak	0.02567	0.13641	0.00785	0.04062	0.02130	0.07551	Yes
K+	Strong	0.92533	0.04734	0.00555	0.76334	0.01846	0.04038	Yes
Kns	Strong	0.96991	0.01431	0.02819	0.97006	0.01455	0.28065	Yes
Mg2+	Strong	0.91875	-0.03226	0.01290	1.05609	0.02673	0.06802	Yes
Ca2+	Strong	0.92659	0.18166	0.03583	0.73983	0.01773	0.33800	Yes
PM10	Weak	0.97678	-2.49499	1.65726	0.99637	0.01308	14.32076	Yes

y = 1.00155x + -0.04035

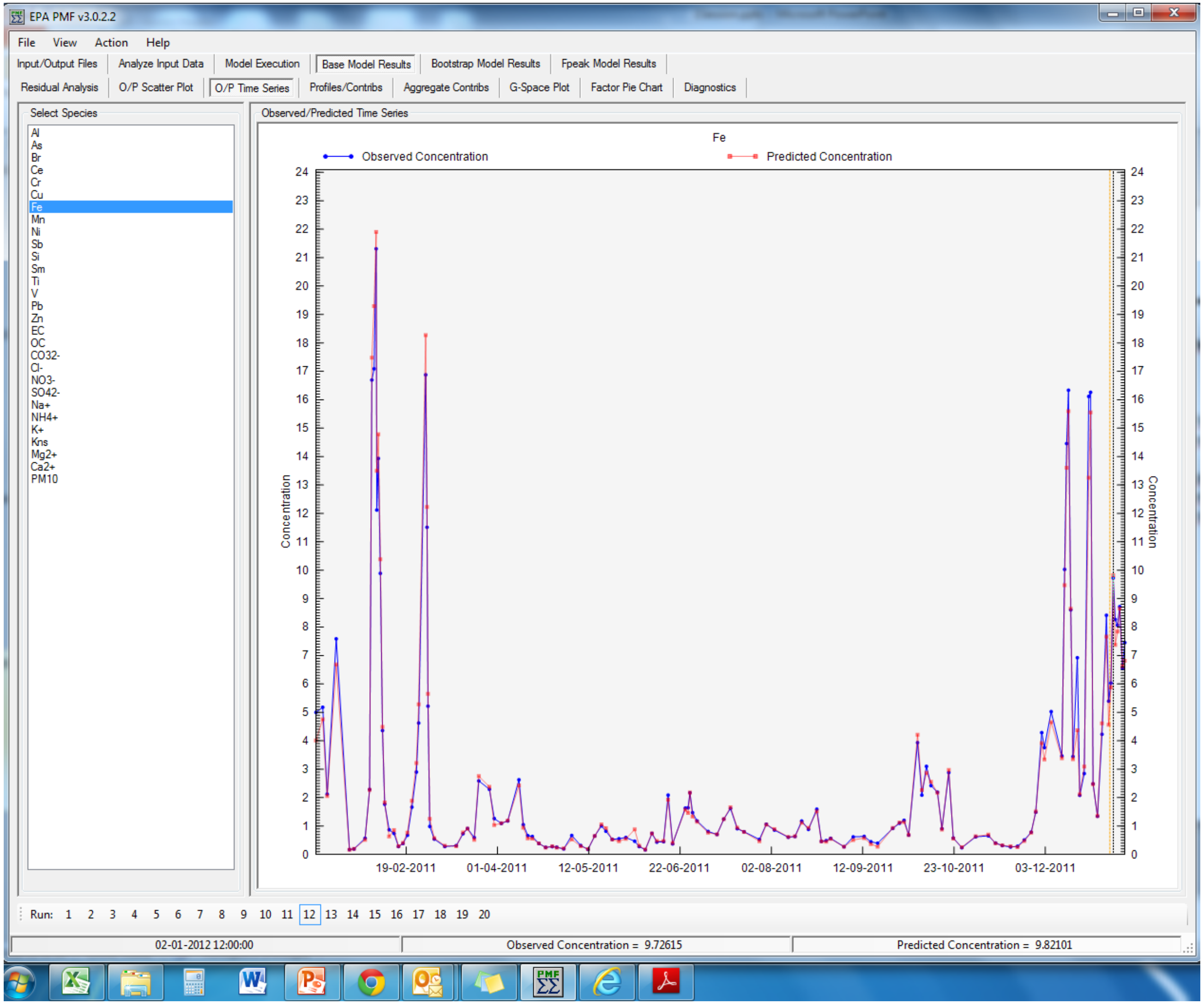
Observed/Predicted scatter plot

Objective:

- To determine how well the model fit each species

- Evaluate if a species should be down weighted or excluded from the model

Base model results



02-01-2012 12:00:00

Observed Concentration = 9.72615

Predicted Concentration = 9.82101

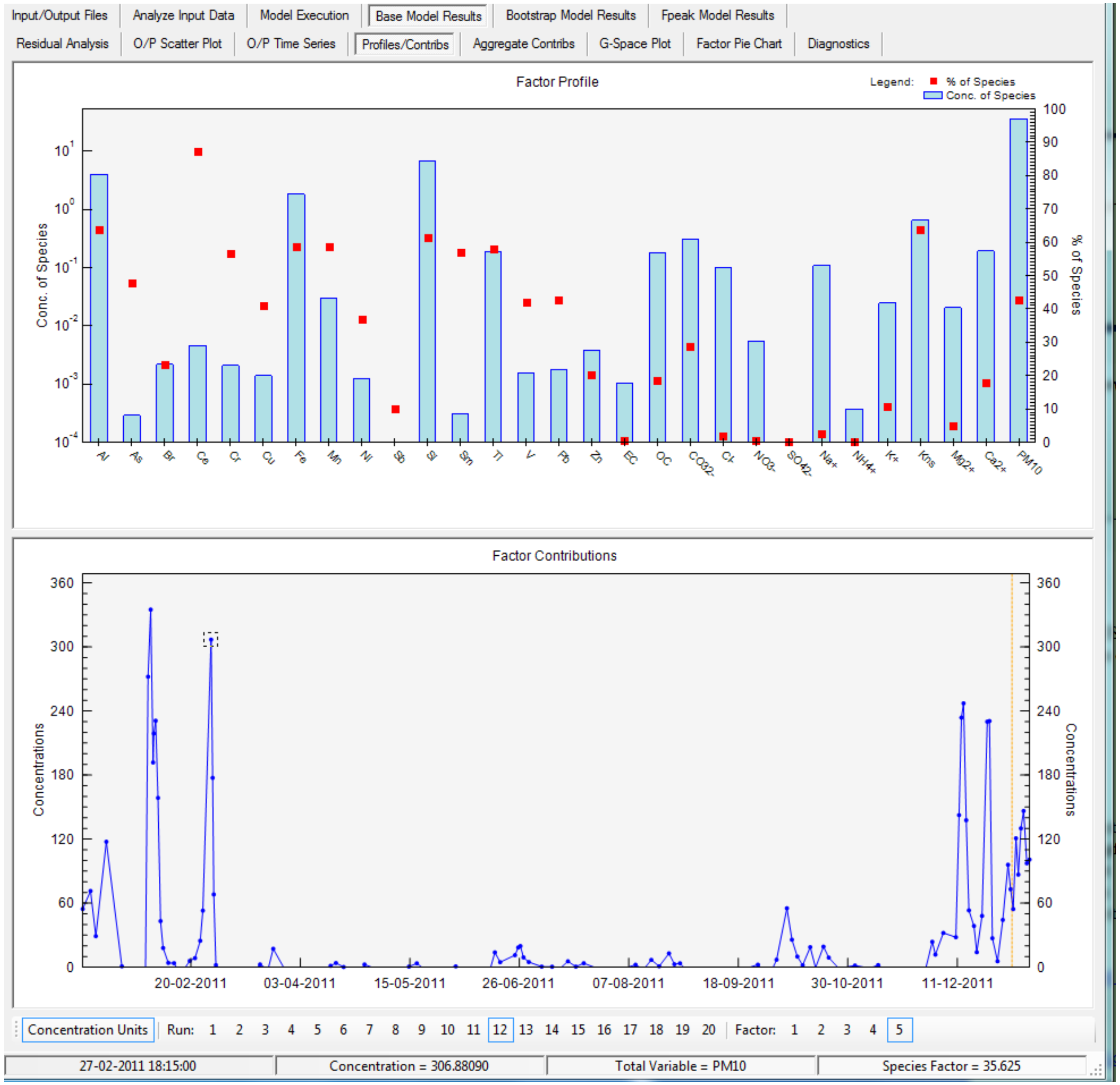
Observed/Predicted time series

Objective:

- To determine how well the model fit each species

- Evaluate if a species should be down weighted or excluded from the model

Base model results

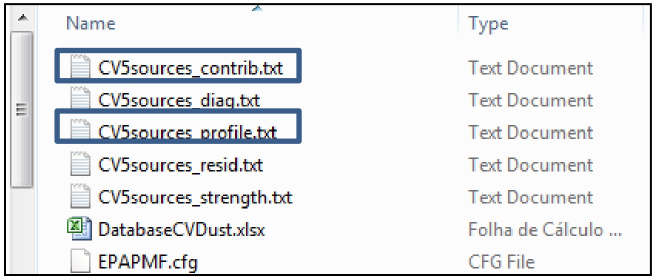


27-02-2011 18:15:00 | Concentration = 306.88090 | Total Variable = PM10 | Species Factor = 35.625

Profiles/Contribs

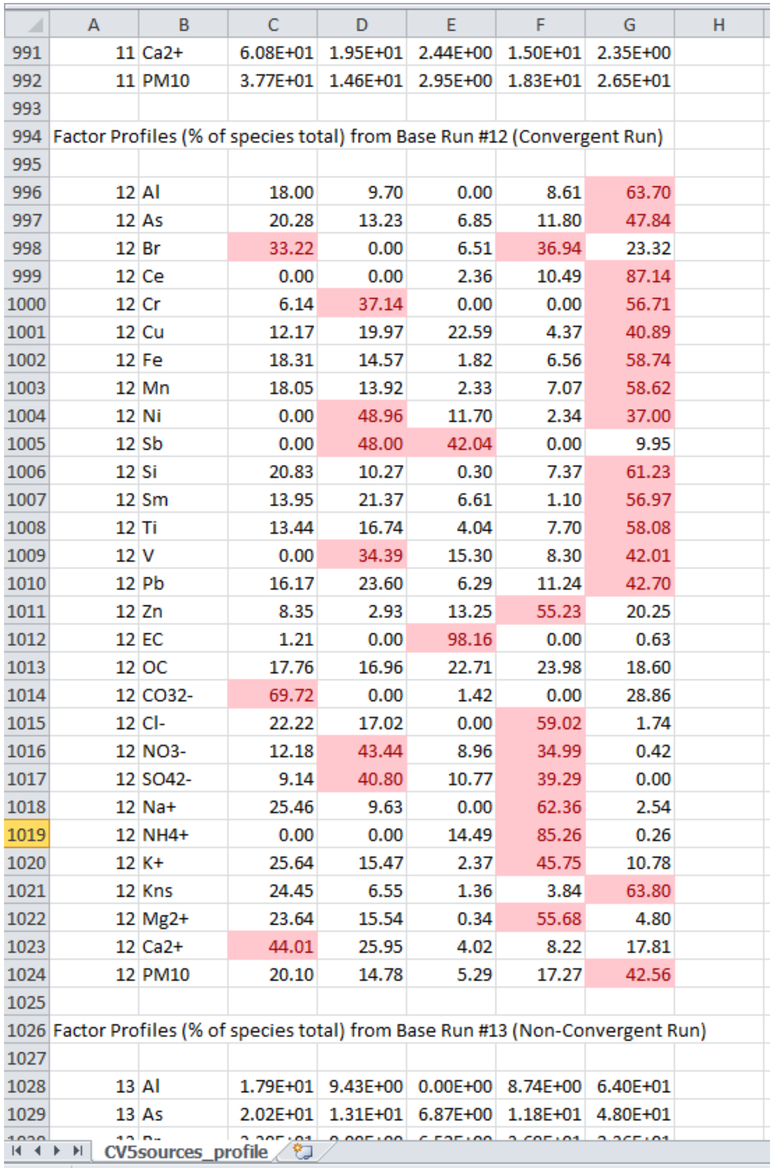
Objective:

- Display the factors resolved by PMF;
- Two graphs are displayed:
  - Profile graph
    - Mass of each species apportioned to the factor 🟦
    - Percent of each species apportionned to the factor 🟥
  - Contribution of the factor to the total mass by sample



Name	Type
CV5sources_contrib.txt	Text Document
CV5sources_diag.txt	Text Document
CV5sources_profile.txt	Text Document
CV5sources_resid.txt	Text Document
CV5sources_strength.txt	Text Document
DatabaseCVDust.xlsx	Folha de Cálculo ...
EPAPMF.cfg	CFG File

Base model results

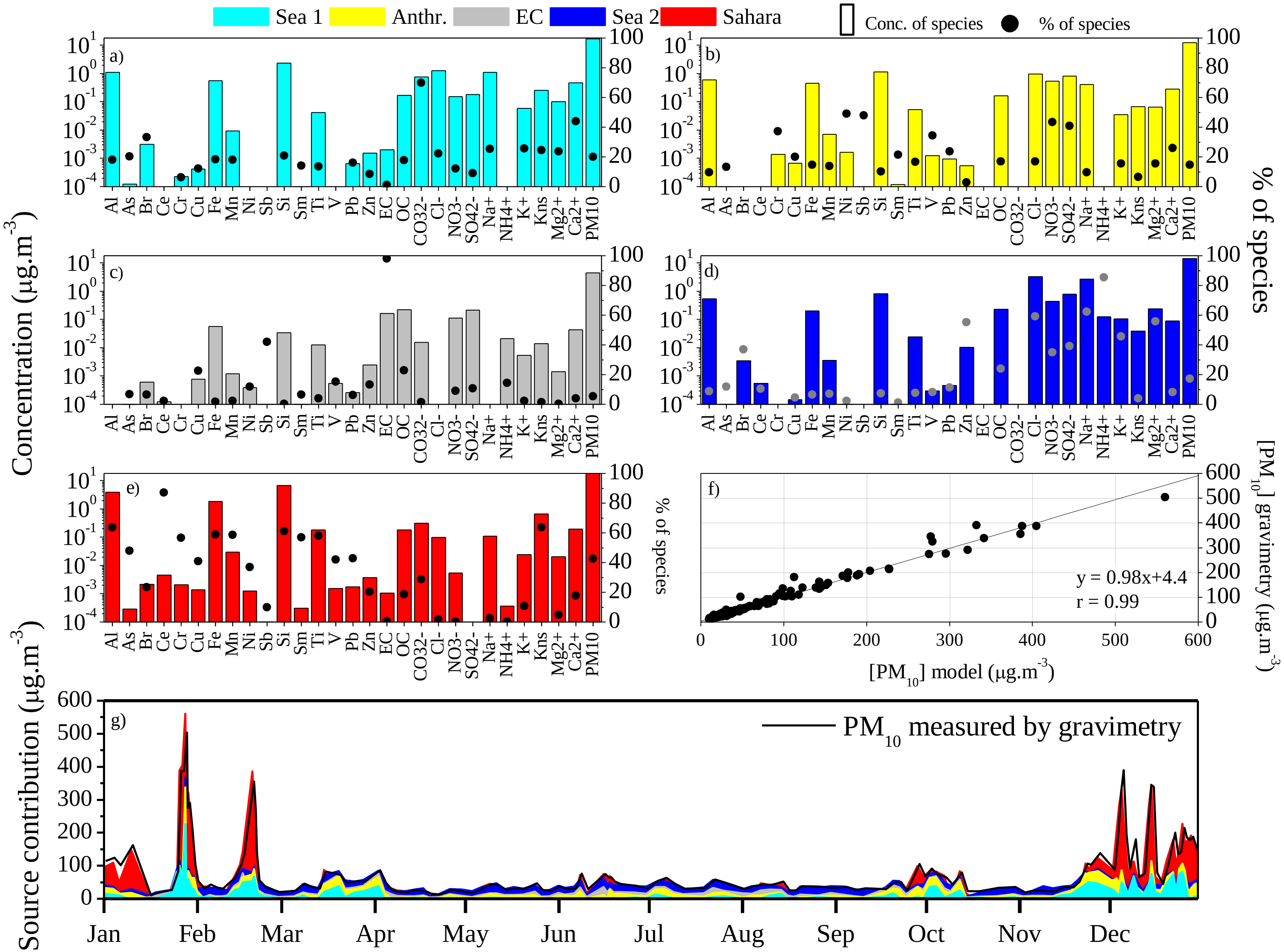


	A	B	C	D	E	F	G	H
991	11	Ca2+	6.08E+01	1.95E+01	2.44E+00	1.50E+01	2.35E+00	
992	11	PM10	3.77E+01	1.46E+01	2.95E+00	1.83E+01	2.65E+01	
993								
994	Factor Profiles (% of species total) from Base Run #12 (Convergent Run)							
995								
996	12	Al	18.00	9.70	0.00	8.61	63.70	
997	12	As	20.28	13.23	6.85	11.80	47.84	
998	12	Br	33.22	0.00	6.51	36.94	23.32	
999	12	Ce	0.00	0.00	2.36	10.49	87.14	
1000	12	Cr	6.14	37.14	0.00	0.00	56.71	
1001	12	Cu	12.17	19.97	22.59	4.37	40.89	
1002	12	Fe	18.31	14.57	1.82	6.56	58.74	
1003	12	Mn	18.05	13.92	2.33	7.07	58.62	
1004	12	Ni	0.00	48.96	11.70	2.34	37.00	
1005	12	Sb	0.00	48.00	42.04	0.00	9.95	
1006	12	Si	20.83	10.27	0.30	7.37	61.23	
1007	12	Sm	13.95	21.37	6.61	1.10	56.97	
1008	12	Ti	13.44	16.74	4.04	7.70	58.08	
1009	12	V	0.00	34.39	15.30	8.30	42.01	
1010	12	Pb	16.17	23.60	6.29	11.24	42.70	
1011	12	Zn	8.35	2.93	13.25	55.23	20.25	
1012	12	EC	1.21	0.00	98.16	0.00	0.63	
1013	12	OC	17.76	16.96	22.71	23.98	18.60	
1014	12	CO32-	69.72	0.00	1.42	0.00	28.86	
1015	12	Cl-	22.22	17.02	0.00	59.02	1.74	
1016	12	NO3-	12.18	43.44	8.96	34.99	0.42	
1017	12	SO42-	9.14	40.80	10.77	39.29	0.00	
1018	12	Na+	25.46	9.63	0.00	62.36	2.54	
1019	12	NH4+	0.00	0.00	14.49	85.26	0.26	
1020	12	K+	25.64	15.47	2.37	45.75	10.78	
1021	12	Kns	24.45	6.55	1.36	3.84	63.80	
1022	12	Mg2+	23.64	15.54	0.34	55.68	4.80	
1023	12	Ca2+	44.01	25.95	4.02	8.22	17.81	
1024	12	PM10	20.10	14.78	5.29	17.27	42.56	
1025								
1026	Factor Profiles (% of species total) from Base Run #13 (Non-Convergent Run)							
1027								
1028	13	Al	1.79E+01	9.43E+00	0.00E+00	8.74E+00	6.40E+01	
1029	13	As	2.02E+01	1.31E+01	6.87E+00	1.18E+01	4.80E+01	

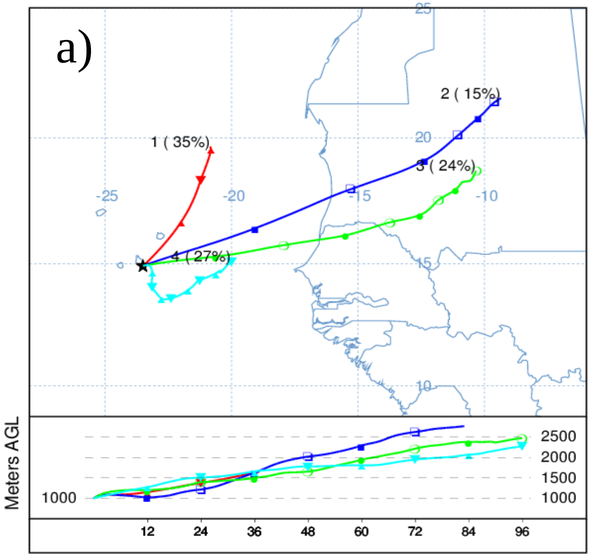
Identifications of sources

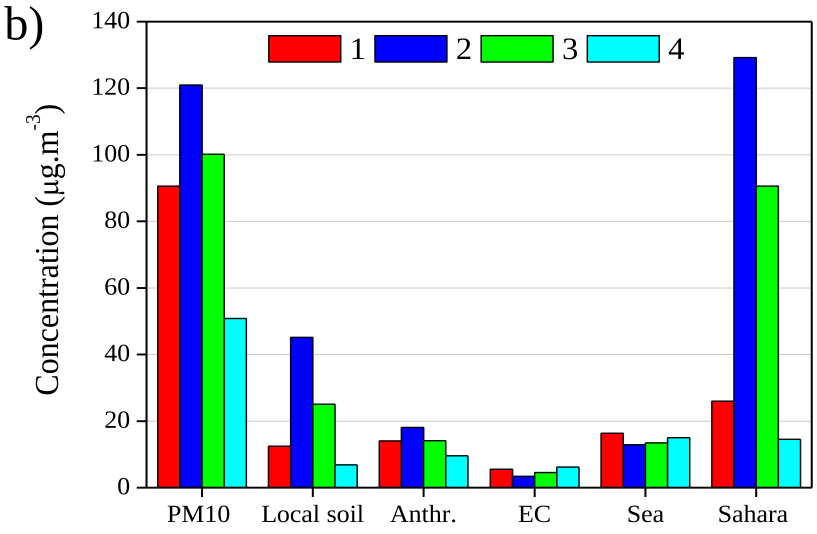
	Species	Source
F1	Br, CO_3^{2-}, Ca^{2+}	Local soil
F2	Cr, Ni, Sb, V, NO_3^-, SO_4^{2-}	Anthropogenic source
F3	Sb, EC	Anthropogenic source
F4	Br, Zn, Cl^-, NO_3^-, SO_4^{2-}, Na^+, NH_4^+, K^+, Mg^{2+}	Sea
F5	Al, As, Ce, Cr, Cu, Fe, Mn, Mi, Si, Sm, Ti, V, Pb, Kns	Sahara dust

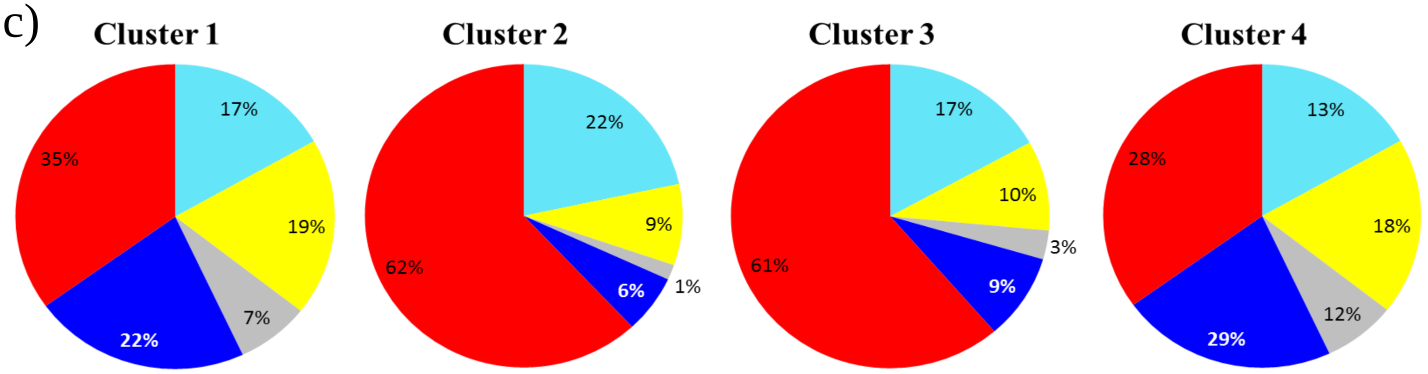
Base model results



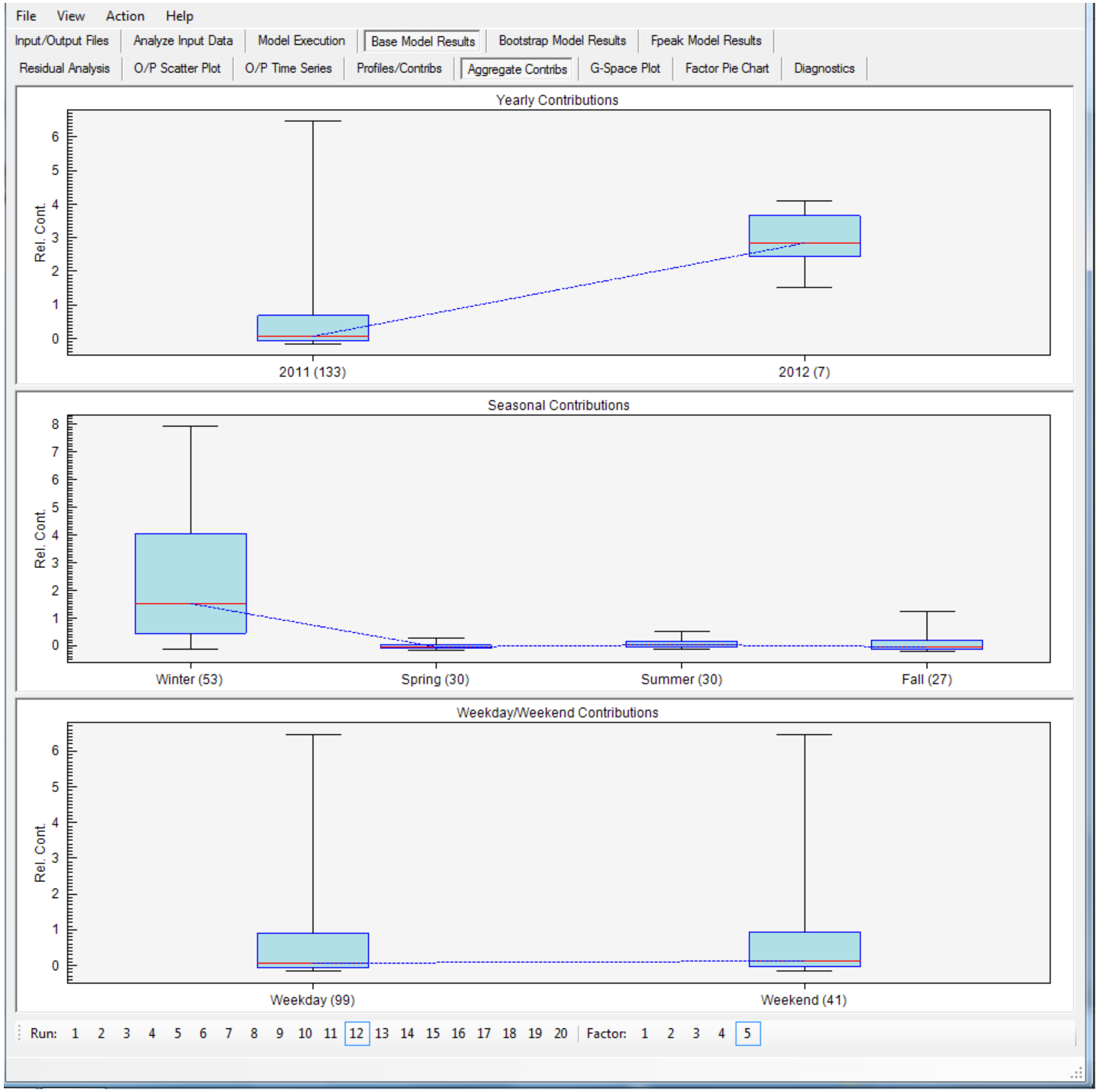
Base model results







Base model results

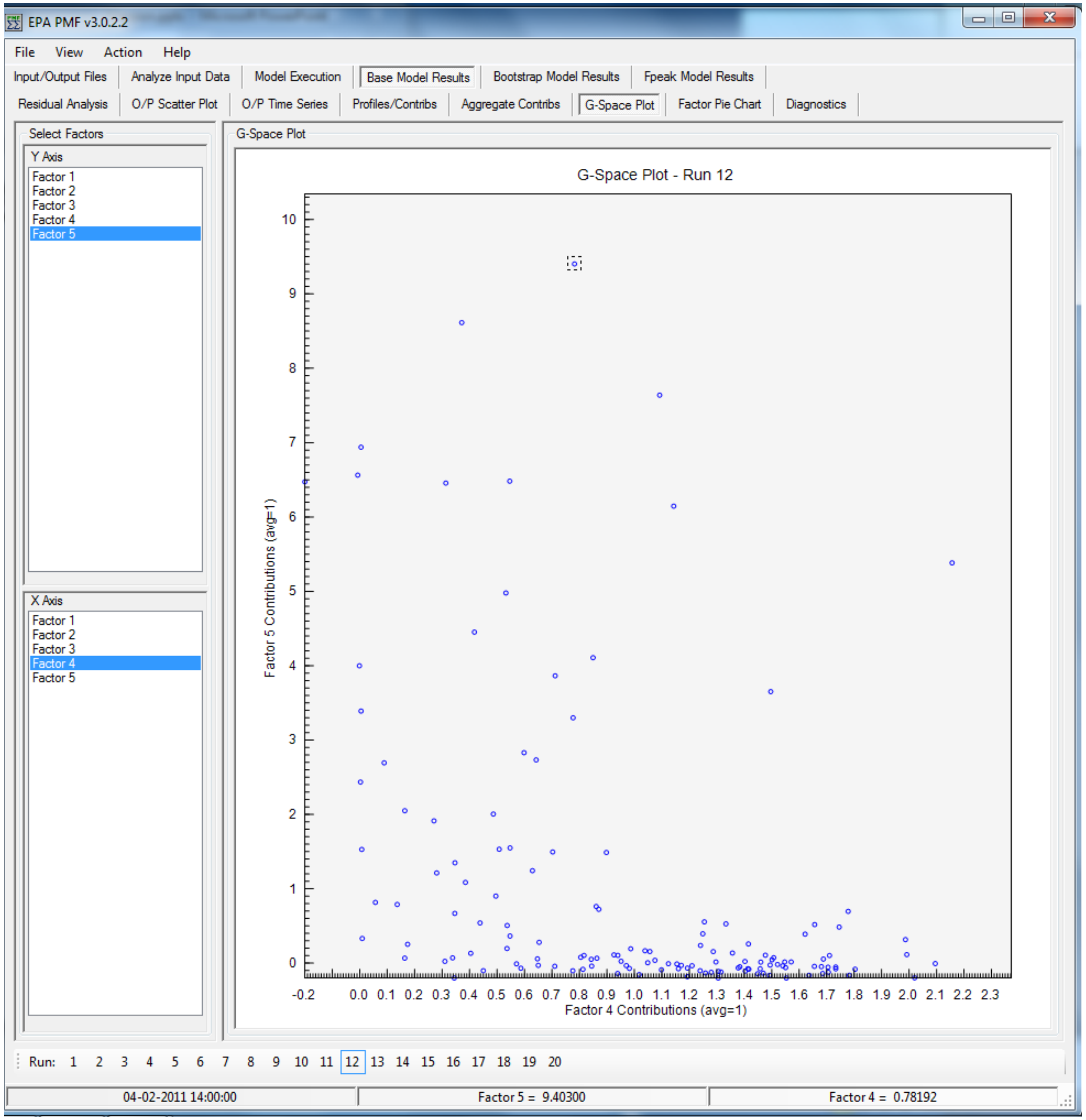


Aggregate Contribs

Objective:

- Display the factors contribution discriminated:
  - by year;
  - season;
  - weekday/wekeend

Base model results



04-02-2011 14:00:00

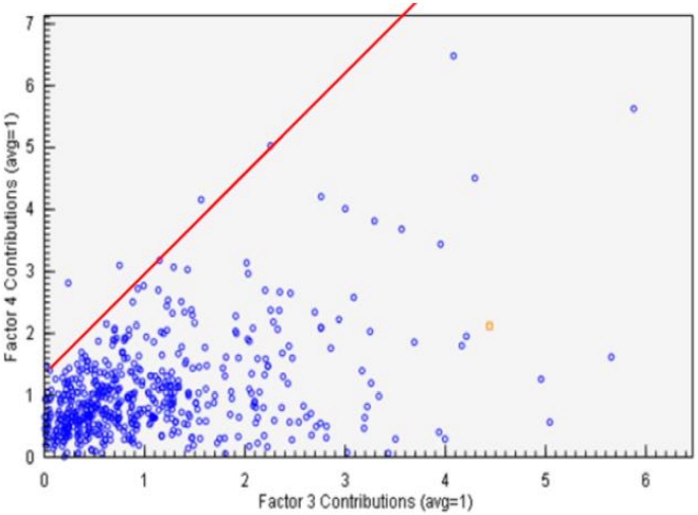
Factor 5 = 9.40300

Factor 4 = 0.78192

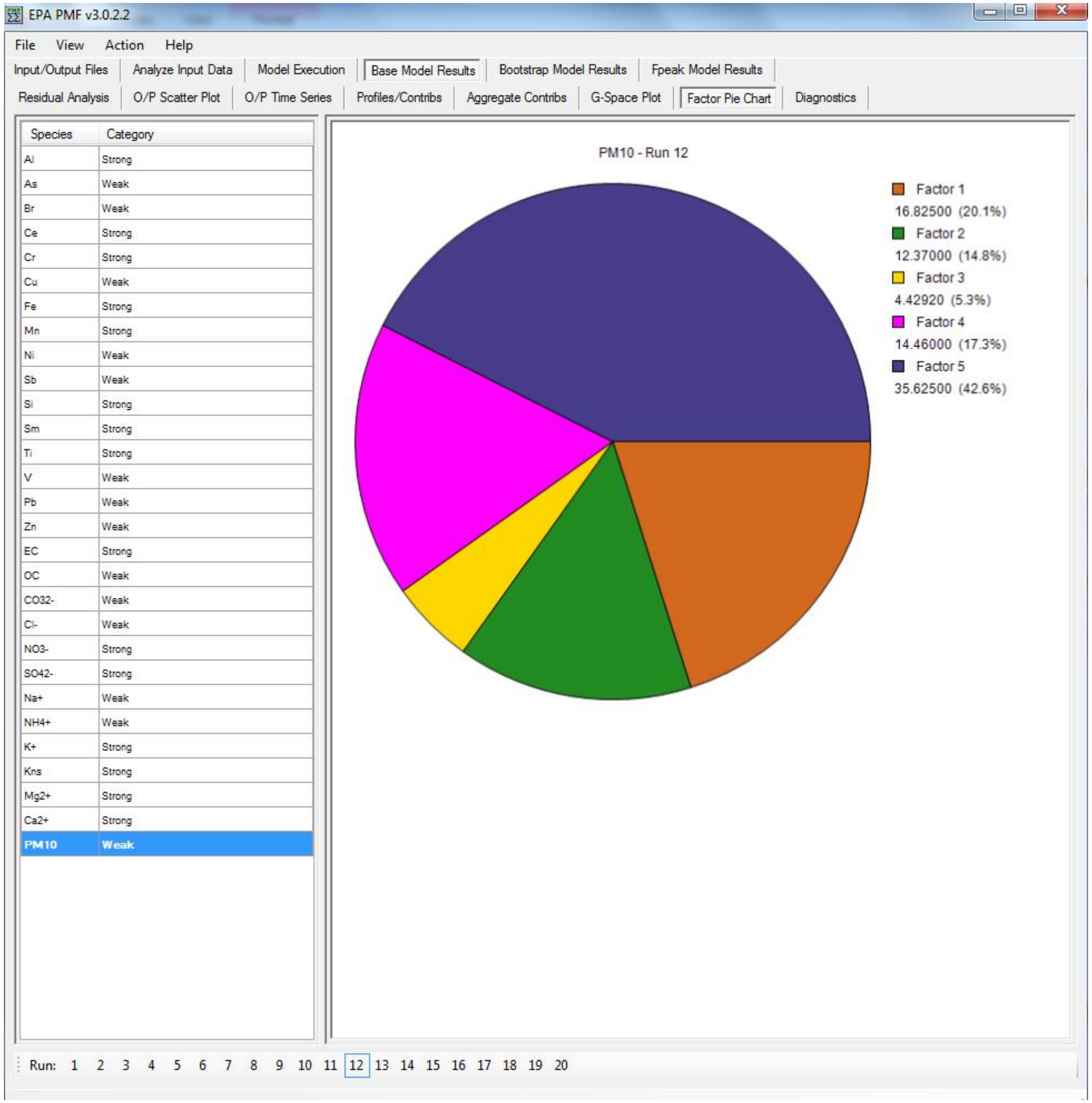
G-Space plot

Objective:

- Display 1 factor contribution versus the other;
- Determine if the solution
  - 1- has filled the solution space – edge of the scatter plot will correspond to the axis
  - 2 - has some rotational ambiguity – appears na oblique edge on a space plot



Base model results



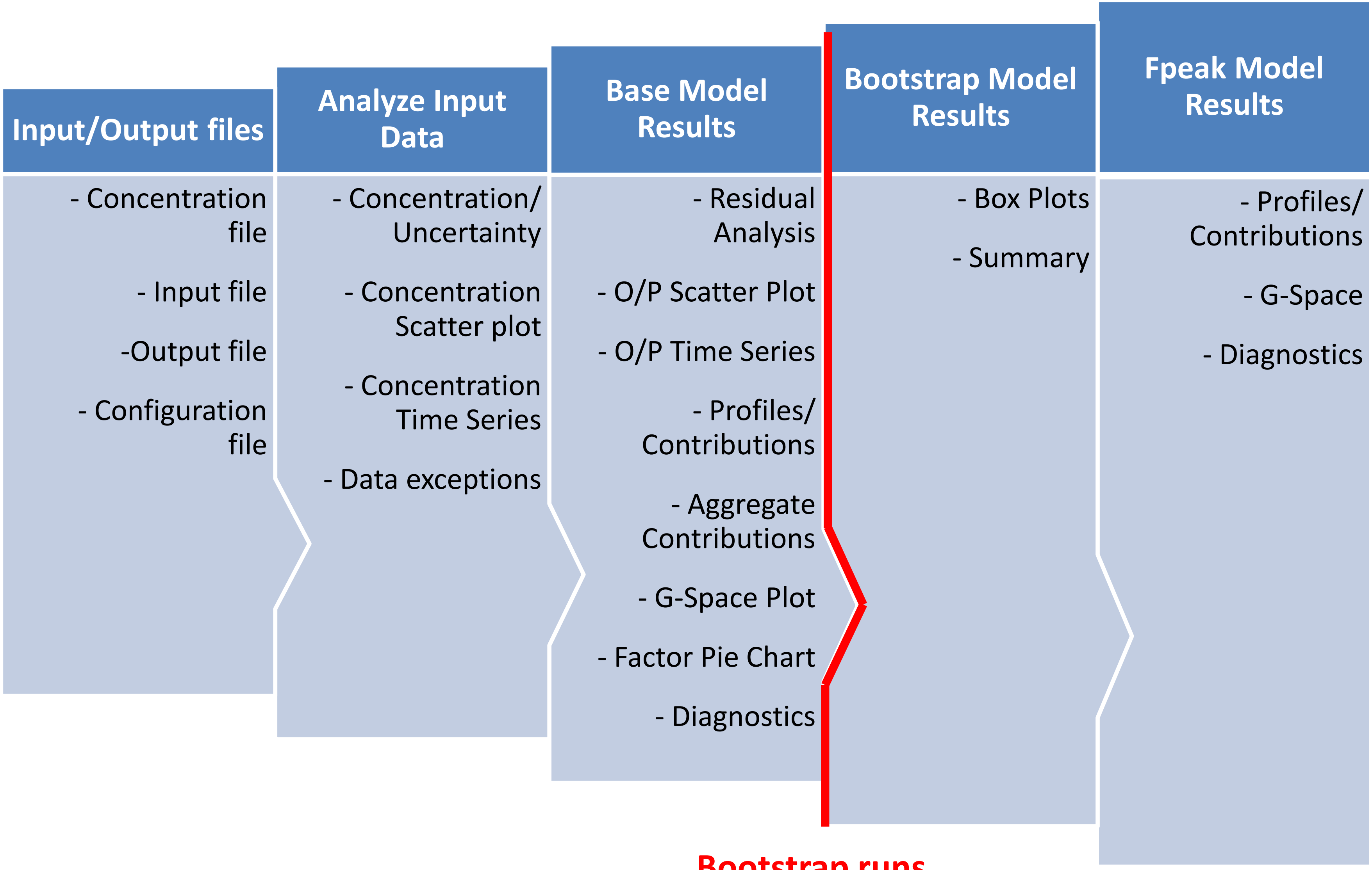
Species	Category
Al	Strong
As	Weak
Br	Weak
Ce	Strong
Cr	Strong
Cu	Weak
Fe	Strong
Mn	Strong
Ni	Weak
Sb	Weak
Si	Strong
Sm	Strong
Ti	Strong
V	Weak
Pb	Weak
Zn	Weak
EC	Strong
OC	Weak
CO32-	Weak
Cl-	Weak
NO3-	Strong
SO42-	Strong
Na+	Weak
NH4+	Weak
K+	Strong
Kns	Strong
Mg2+	Strong
Ca2+	Strong
PM10	Weak

Factor Pie Chart

Objective:

- Display the distribution of each species among the factors resolved by PMF

Order of the operations



Input/Output files

- Concentration file
- Input file
- Output file
- Configuration file

Analyze Input Data

- Concentration/ Uncertainty
- Concentration Scatter plot
- Concentration Time Series
- Data exceptions

Base Model Results

- Residual Analysis
- O/P Scatter Plot
- O/P Time Series
- Profiles/ Contributions
- Aggregate Contributions
- G-Space Plot
- Factor Pie Chart
- Diagnostics

Bootstrap Model Results

- Box Plots
- Summary

Fpeak Model Results

- Profiles/ Contributions
- G-Space
- Diagnostics

Bootstrap runs

Bootstrap runs

Objective: Estimate the stability and uncertainty of the solution

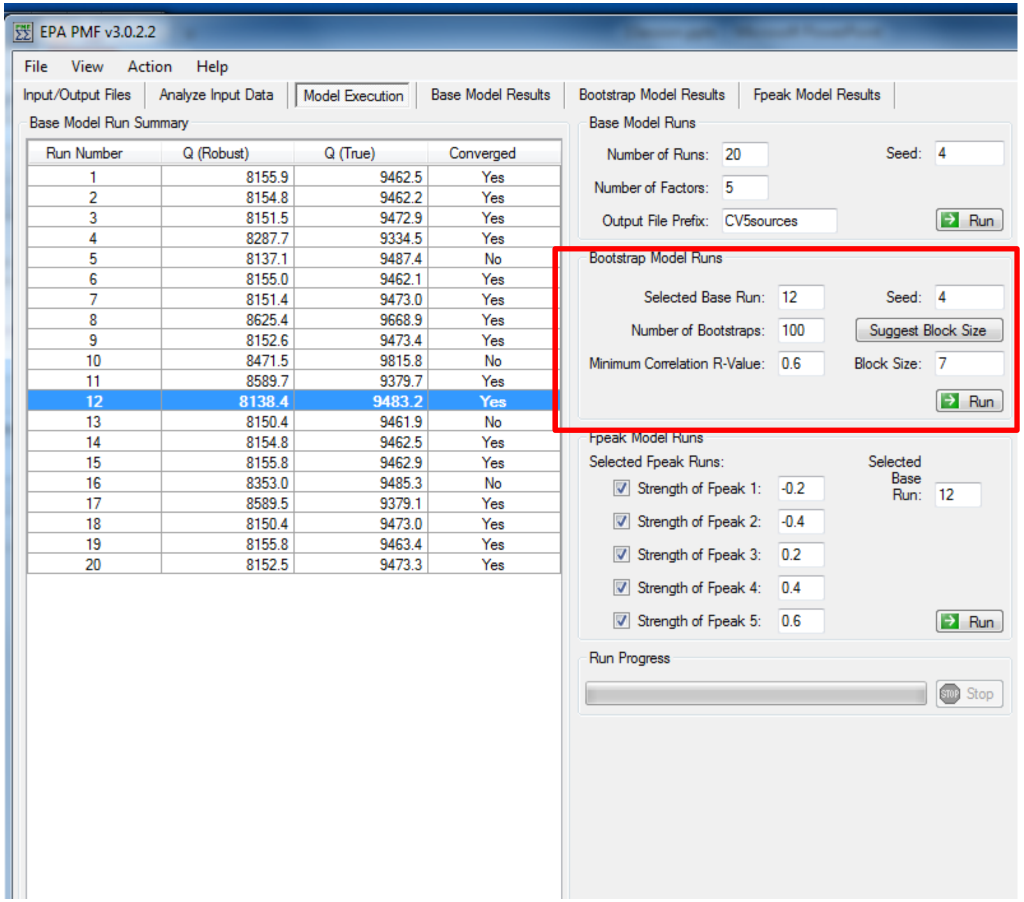
How?

1 – Randomly selects non-overlapping blocks of samples and creats a new input data file of the selected samples;

2 - PMF is run on the new data set and each bootstrap factor is mapped to a base run factor with which the bootstrap factor has the highest correlation, above a user-specific threshold;

3 – If no base factors have a correlation above the thresholdfor a given bootstrap factor , the factor is considered unmapped;

Bootstrap runs



Base Model Run Summary

Run Number	Q (Robust)	Q (True)	Converged
1	8155.9	9462.5	Yes
2	8154.8	9462.2	Yes
3	8151.5	9472.9	Yes
4	8287.7	9334.5	Yes
5	8137.1	9487.4	No
6	8155.0	9462.1	Yes
7	8151.4	9473.0	Yes
8	8625.4	9668.9	Yes
9	8152.6	9473.4	Yes
10	8471.5	9815.8	No
11	8589.7	9379.7	Yes
12	8138.4	9483.2	Yes
13	8150.4	9461.9	No
14	8154.8	9462.5	Yes
15	8155.8	9462.9	Yes
16	8353.0	9485.3	No
17	8589.5	9379.1	Yes
18	8150.4	9473.0	Yes
19	8155.8	9463.4	Yes
20	8152.5	9473.3	Yes

Base Model Runs

Number of Runs: 20
Seed: 4
Number of Factors: 5
Output File Prefix: CV5sources

Bootstrap Model Runs

Selected Base Run: 12
Seed: 4
Number of Bootstraps: 100
Minimum Correlation R-Value: 0.6
Block Size: 7

Fpeak Model Runs

Selected Fpeak Runs:
Selected Base Run: 12
☑ Strength of Fpeak 1: -0.2
☑ Strength of Fpeak 2: -0.4
☑ Strength of Fpeak 3: 0.2
☑ Strength of Fpeak 4: 0.4
☑ Strength of Fpeak 5: 0.6

Selected base run – the base run to be used to map each bootstrap run

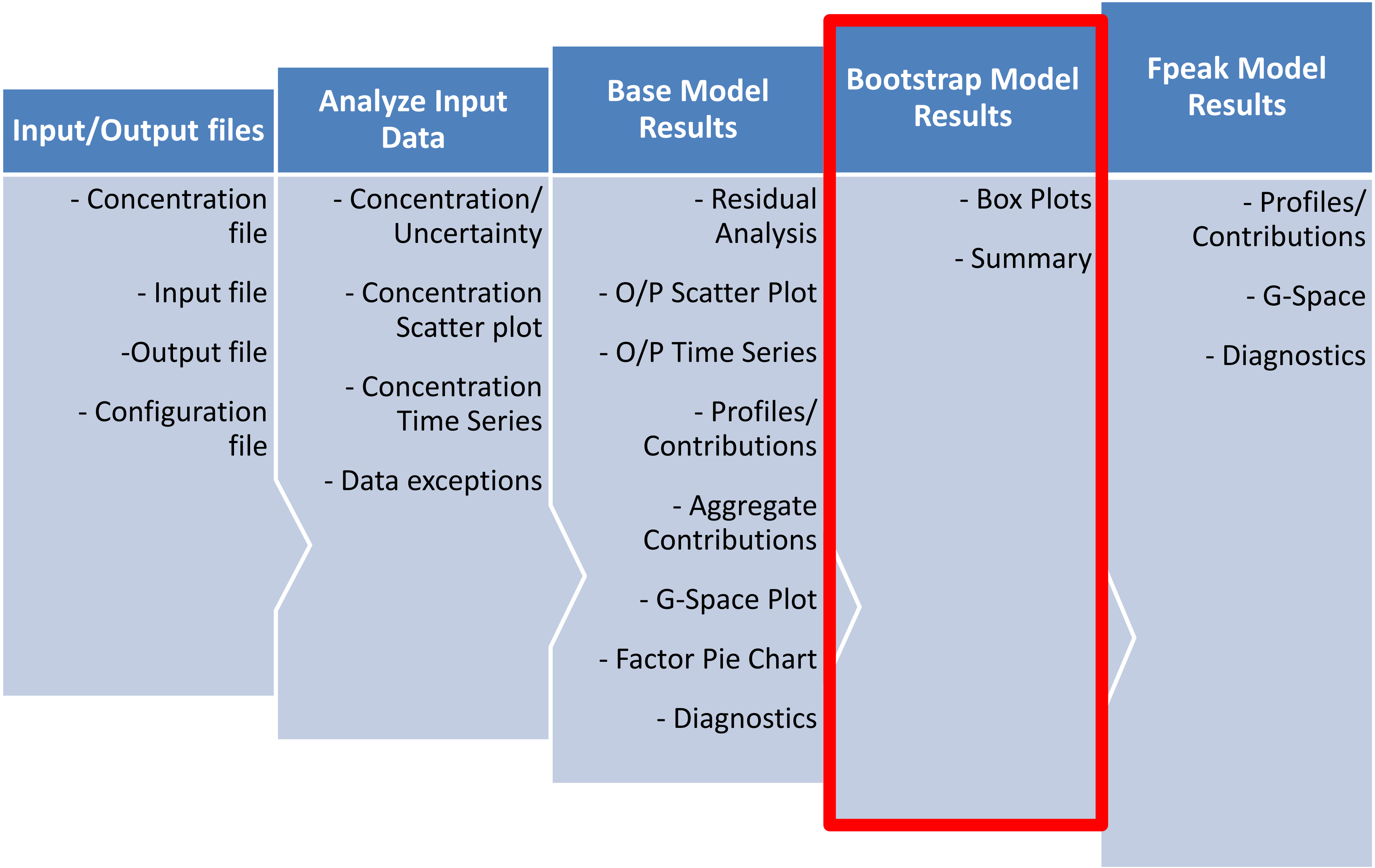
Number of bootstraps – It is recommended at least 100 bootstrap runs to ensure the robustness of the statistics

Minimum Correlation R-Values – the minimum Pearson correlation coefficient that will be used in the assignement of a bootstrap run factor to a base run factor. Default value 0.6 (can decrease if a large number of factors are unmapped)

Seed – similar to base runs, starting point for iterations

Block size – the number of samples that will be selected in each step for resampling

Order of the operations



Input/Output files

- Concentration file
- Input file
- Output file
- Configuration file

Analyze Input Data

- Concentration/ Uncertainty
- Concentration Scatter plot
- Concentration Time Series
- Data exceptions

Base Model Results

- Residual Analysis
- O/P Scatter Plot
- O/P Time Series
- Profiles/ Contributions
- Aggregate Contributions
- G-Space Plot
- Factor Pie Chart
- Diagnostics

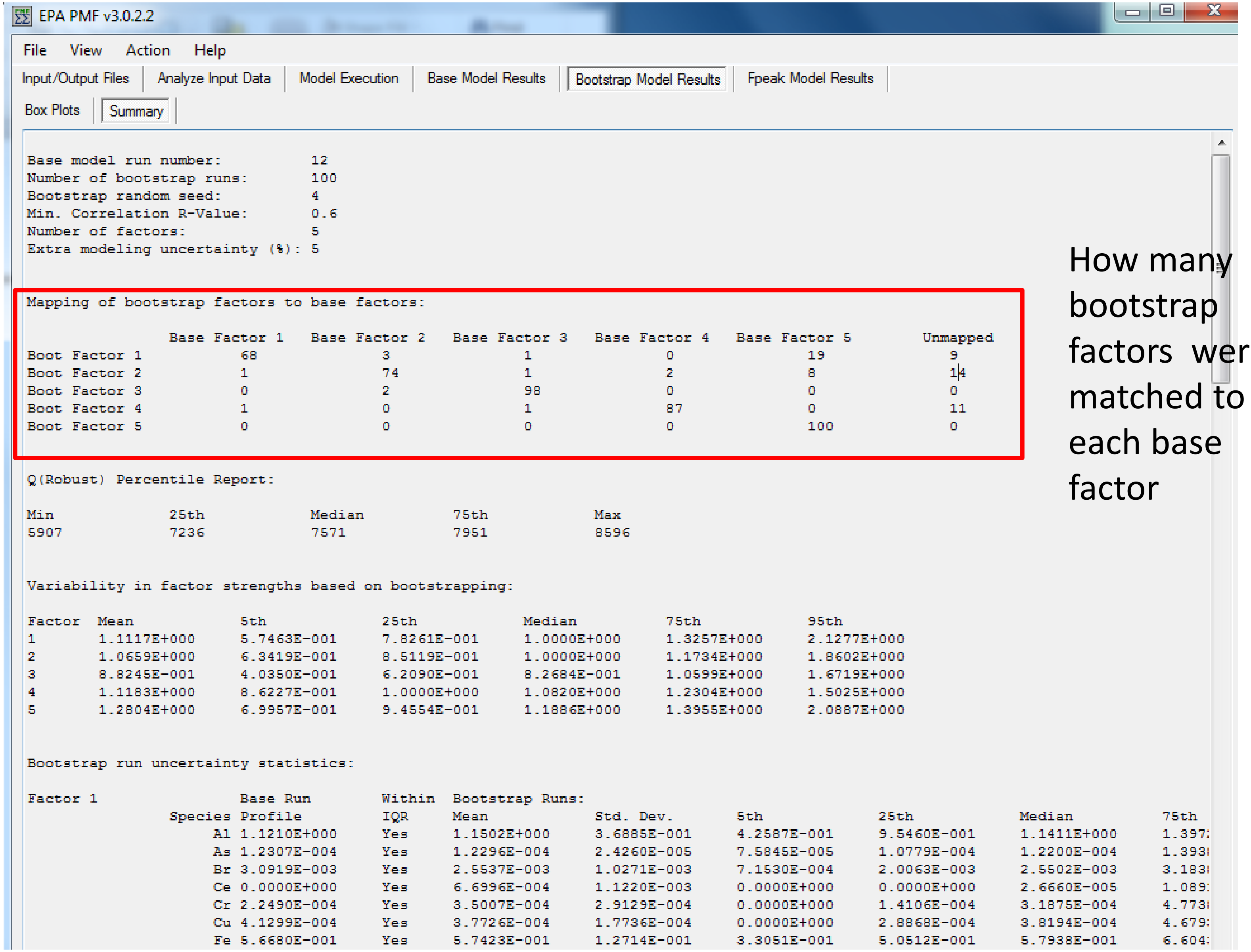
Bootstrap Model Results

- Box Plots
- Summary

Fpeak Model Results

- Profiles/ Contributions
- G-Space
- Diagnostics

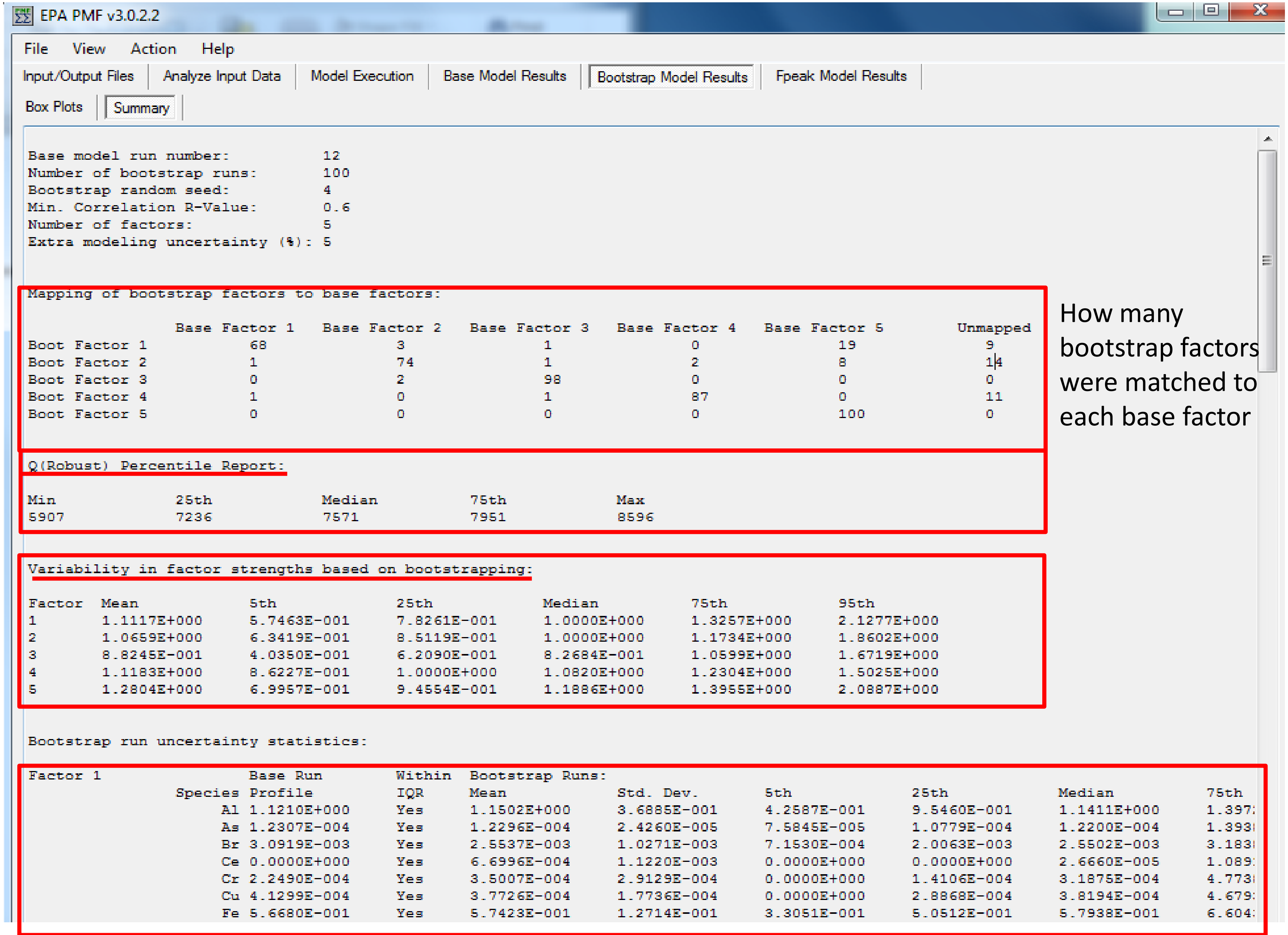
Bootstrap Model Results



```
Base model run number:         12
Number of bootstrap runs:      100
Bootstrap random seed:         4
Min. Correlation R-Value:      0.6
Number of factors:             5
Extra modeling uncertainty (%): 5


Mapping of bootstrap factors to base factors:

                Base Factor 1   Base Factor 2   Base Factor 3   Base Factor 4   Base Factor 5        Unmapped
Boot Factor 1            68               3               1               0              19               9
Boot Factor 2             1              74               1               2               8              14
Boot Factor 3             0               2              98               0               0               0
Boot Factor 4             1               0               1              87               0              11
Boot Factor 5             0               0               0               0             100               0


Q(Robust) Percentile Report:

Min             25th            Median          75th            Max
5907            7236            7571            7951            8596


Variability in factor strengths based on bootstrapping:

Factor  Mean            5th             25th            Median          75th            95th
1       1.1117E+000     5.7463E-001     7.8261E-001     1.0000E+000     1.3257E+000     2.1277E+000
2       1.0659E+000     6.3419E-001     8.5119E-001     1.0000E+000     1.1734E+000     1.8602E+000
3       8.8245E-001     4.0350E-001     6.2090E-001     8.2684E-001     1.0599E+000     1.6719E+000
4       1.1183E+000     8.6227E-001     1.0000E+000     1.0820E+000     1.2304E+000     1.5025E+000
5       1.2804E+000     6.9957E-001     9.4554E-001     1.1886E+000     1.3955E+000     2.0887E+000


Bootstrap run uncertainty statistics:

Factor 1                Base Run        Within  Bootstrap Runs:
                Species Profile         IQR     Mean            Std. Dev.       5th             25th            Median          75th
                     Al 1.1210E+000     Yes     1.1502E+000     3.6885E-001     4.2587E-001     9.5460E-001     1.1411E+000     1.397
                     As 1.2307E-004     Yes     1.2296E-004     2.4260E-005     7.5845E-005     1.0779E-004     1.2200E-004     1.393
                     Br 3.0919E-003     Yes     2.5537E-003     1.0271E-003     7.1530E-004     2.0063E-003     2.5502E-003     3.183
                     Ce 0.0000E+000     Yes     6.6996E-004     1.1220E-003     0.0000E+000     0.0000E+000     2.6660E-005     1.089
                     Cr 2.2490E-004     Yes     3.5007E-004     2.9129E-004     0.0000E+000     1.4106E-004     3.1875E-004     4.773
                     Cu 4.1299E-004     Yes     3.7726E-004     1.7736E-004     0.0000E+000     2.8868E-004     3.8194E-004     4.679
                     Fe 5.6680E-001     Yes     5.7423E-001     1.2714E-001     3.3051E-001     5.0512E-001     5.7938E-001     6.604
```

How many bootstrap factors were matched to each base factor

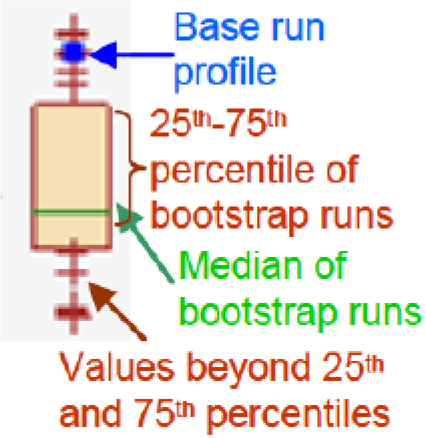
Bootstrap Model Results



```
Base model run number:          12
Number of bootstrap runs:       100
Bootstrap random seed:          4
Min. Correlation R-Value:       0.6
Number of factors:              5
Extra modeling uncertainty (%): 5


Mapping of bootstrap factors to base factors:

                Base Factor 1   Base Factor 2   Base Factor 3   Base Factor 4   Base Factor 5       Unmapped
Boot Factor 1            68               3               1               0              19              9
Boot Factor 2             1              74               1               2               8             14
Boot Factor 3             0               2              98               0               0              0
Boot Factor 4             1               0               1              87               0             11
Boot Factor 5             0               0               0               0             100              0

Q(Robust) Percentile Report:

Min             25th            Median          75th            Max
5907            7236            7571            7951            8596


Variability in factor strengths based on bootstrapping:

Factor  Mean            5th             25th            Median          75th            95th
1       1.1117E+000     5.7463E-001     7.8261E-001     1.0000E+000     1.3257E+000     2.1277E+000
2       1.0659E+000     6.3419E-001     8.5119E-001     1.0000E+000     1.1734E+000     1.8602E+000
3       8.8245E-001     4.0350E-001     6.2090E-001     8.2684E-001     1.0599E+000     1.6719E+000
4       1.1183E+000     8.6227E-001     1.0000E+000     1.0820E+000     1.2304E+000     1.5025E+000
5       1.2804E+000     6.9957E-001     9.4554E-001     1.1886E+000     1.3955E+000     2.0887E+000


Bootstrap run uncertainty statistics:

Factor 1                Base Run        Within  Bootstrap Runs:
                Species Profile         IQR     Mean            Std. Dev.       5th             25th            Median          75th
                     Al 1.1210E+000     Yes     1.1502E+000     3.6885E-001     4.2587E-001     9.5460E-001     1.1411E+000     1.397
                     As 1.2307E-004     Yes     1.2296E-004     2.4260E-005     7.5845E-005     1.0779E-004     1.2200E-004     1.393
                     Br 3.0919E-003     Yes     2.5537E-003     1.0271E-003     7.1530E-004     2.0063E-003     2.5502E-003     3.183
                     Ce 0.0000E+000     Yes     6.6996E-004     1.1220E-003     0.0000E+000     0.0000E+000     2.6660E-005     1.089
                     Cr 2.2490E-004     Yes     3.5007E-004     2.9129E-004     0.0000E+000     1.4106E-004     3.1875E-004     4.773
                     Cu 4.1299E-004     Yes     3.7726E-004     1.7736E-004     0.0000E+000     2.8868E-004     3.8194E-004     4.679
                     Fe 5.6680E-001     Yes     5.7423E-001     1.2714E-001     3.3051E-001     5.0512E-001     5.7938E-001     6.604
```

How many bootstrap factors were matched to each base factor

Bootstrap Model Results



Base run profile

25th-75th percentile of bootstrap runs

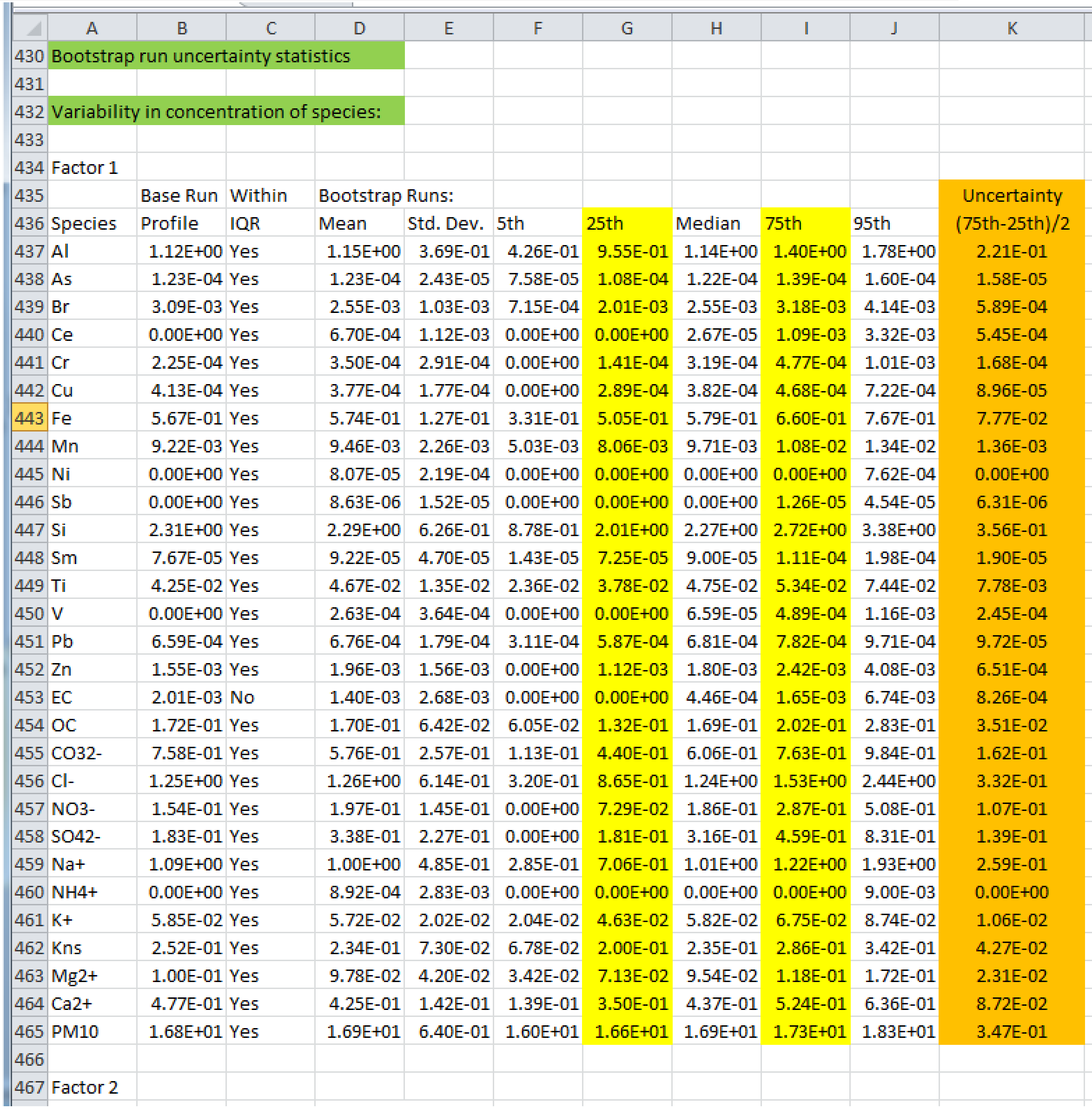
Median of bootstrap runs

Values beyond 25th and 75th percentiles

Bootstrap Model Results

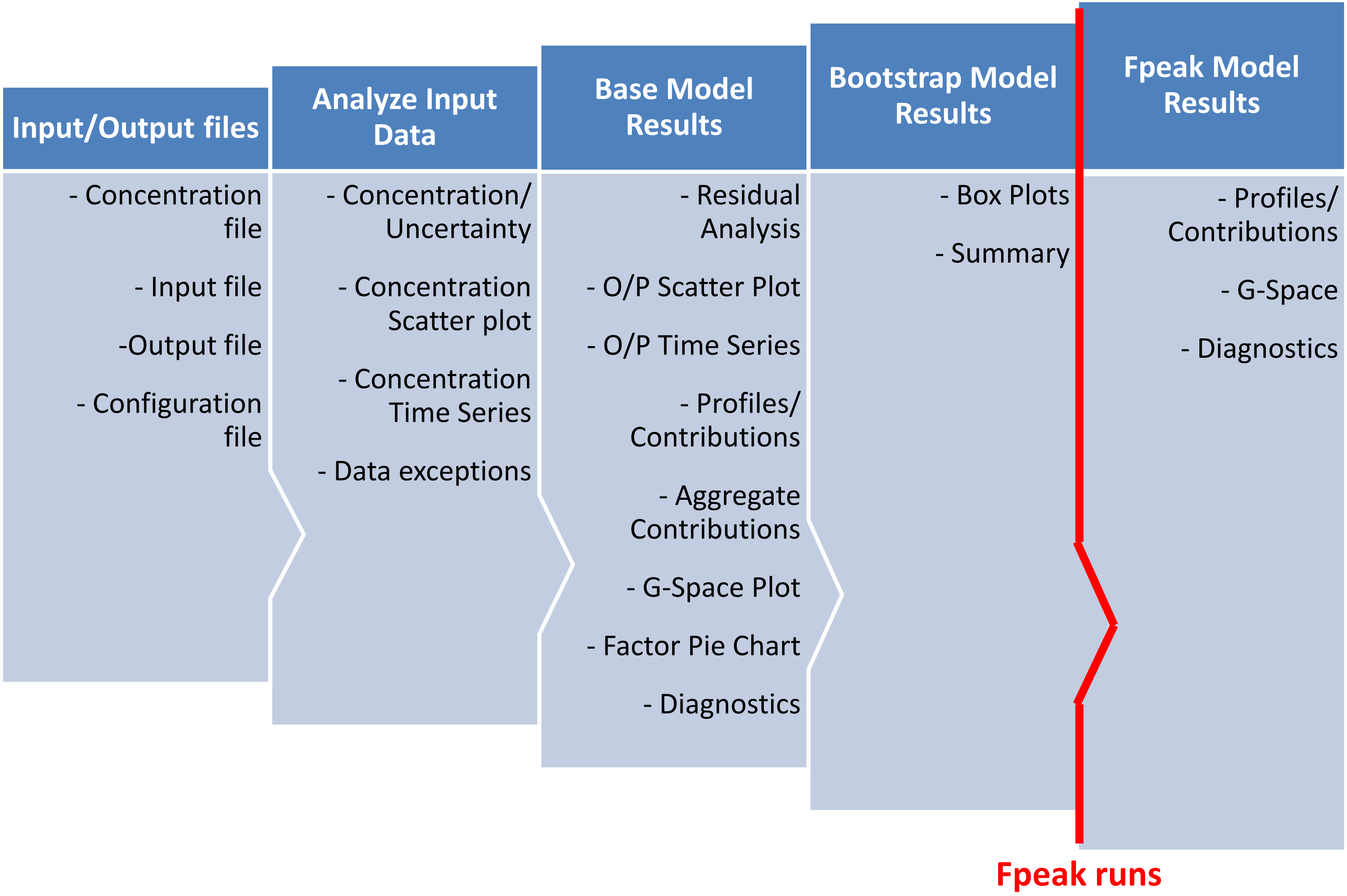
Uncertainty of the solution

- Go to output file “*_diag.txt”.
- In “Bootstrap run uncertainty statistics - Variability in concentration of species” we find output of the bootstrap factor by factor.
- The uncertainty is the absolute difference between the 75th and 25th percentiles divided by two.



	A	B	C	D	E	F	G	H	I	J	K
430	Bootstrap run uncertainty statistics										
431											
432	Variability in concentration of species:										
433											
434	Factor 1										
435		Base Run	Within	Bootstrap Runs:							Uncertainty
436	Species	Profile	IQR	Mean	Std. Dev.	5th	25th	Median	75th	95th	(75th-25th)/2
437	Al	1.12E+00	Yes	1.15E+00	3.69E-01	4.26E-01	9.55E-01	1.14E+00	1.40E+00	1.78E+00	2.21E-01
438	As	1.23E-04	Yes	1.23E-04	2.43E-05	7.58E-05	1.08E-04	1.22E-04	1.39E-04	1.60E-04	1.58E-05
439	Br	3.09E-03	Yes	2.55E-03	1.03E-03	7.15E-04	2.01E-03	2.55E-03	3.18E-03	4.14E-03	5.89E-04
440	Ce	0.00E+00	Yes	6.70E-04	1.12E-03	0.00E+00	0.00E+00	2.67E-05	1.09E-03	3.32E-03	5.45E-04
441	Cr	2.25E-04	Yes	3.50E-04	2.91E-04	0.00E+00	1.41E-04	3.19E-04	4.77E-04	1.01E-03	1.68E-04
442	Cu	4.13E-04	Yes	3.77E-04	1.77E-04	0.00E+00	2.89E-04	3.82E-04	4.68E-04	7.22E-04	8.96E-05
443	Fe	5.67E-01	Yes	5.74E-01	1.27E-01	3.31E-01	5.05E-01	5.79E-01	6.60E-01	7.67E-01	7.77E-02
444	Mn	9.22E-03	Yes	9.46E-03	2.26E-03	5.03E-03	8.06E-03	9.71E-03	1.08E-02	1.34E-02	1.36E-03
445	Ni	0.00E+00	Yes	8.07E-05	2.19E-04	0.00E+00	0.00E+00	0.00E+00	0.00E+00	7.62E-04	0.00E+00
446	Sb	0.00E+00	Yes	8.63E-06	1.52E-05	0.00E+00	0.00E+00	0.00E+00	1.26E-05	4.54E-05	6.31E-06
447	Si	2.31E+00	Yes	2.29E+00	6.26E-01	8.78E-01	2.01E+00	2.27E+00	2.72E+00	3.38E+00	3.56E-01
448	Sm	7.67E-05	Yes	9.22E-05	4.70E-05	1.43E-05	7.25E-05	9.00E-05	1.11E-04	1.98E-04	1.90E-05
449	Ti	4.25E-02	Yes	4.67E-02	1.35E-02	2.36E-02	3.78E-02	4.75E-02	5.34E-02	7.44E-02	7.78E-03
450	V	0.00E+00	Yes	2.63E-04	3.64E-04	0.00E+00	0.00E+00	6.59E-05	4.89E-04	1.16E-03	2.45E-04
451	Pb	6.59E-04	Yes	6.76E-04	1.79E-04	3.11E-04	5.87E-04	6.81E-04	7.82E-04	9.71E-04	9.72E-05
452	Zn	1.55E-03	Yes	1.96E-03	1.56E-03	0.00E+00	1.12E-03	1.80E-03	2.42E-03	4.08E-03	6.51E-04
453	EC	2.01E-03	No	1.40E-03	2.68E-03	0.00E+00	0.00E+00	4.46E-04	1.65E-03	6.74E-03	8.26E-04
454	OC	1.72E-01	Yes	1.70E-01	6.42E-02	6.05E-02	1.32E-01	1.69E-01	2.02E-01	2.83E-01	3.51E-02
455	CO32-	7.58E-01	Yes	5.76E-01	2.57E-01	1.13E-01	4.40E-01	6.06E-01	7.63E-01	9.84E-01	1.62E-01
456	Cl-	1.25E+00	Yes	1.26E+00	6.14E-01	3.20E-01	8.65E-01	1.24E+00	1.53E+00	2.44E+00	3.32E-01
457	NO3-	1.54E-01	Yes	1.97E-01	1.45E-01	0.00E+00	7.29E-02	1.86E-01	2.87E-01	5.08E-01	1.07E-01
458	SO42-	1.83E-01	Yes	3.38E-01	2.27E-01	0.00E+00	1.81E-01	3.16E-01	4.59E-01	8.31E-01	1.39E-01
459	Na+	1.09E+00	Yes	1.00E+00	4.85E-01	2.85E-01	7.06E-01	1.01E+00	1.22E+00	1.93E+00	2.59E-01
460	NH4+	0.00E+00	Yes	8.92E-04	2.83E-03	0.00E+00	0.00E+00	0.00E+00	0.00E+00	9.00E-03	0.00E+00
461	K+	5.85E-02	Yes	5.72E-02	2.02E-02	2.04E-02	4.63E-02	5.82E-02	6.75E-02	8.74E-02	1.06E-02
462	Kns	2.52E-01	Yes	2.34E-01	7.30E-02	6.78E-02	2.00E-01	2.35E-01	2.86E-01	3.42E-01	4.27E-02
463	Mg2+	1.00E-01	Yes	9.78E-02	4.20E-02	3.42E-02	7.13E-02	9.54E-02	1.18E-01	1.72E-01	2.31E-02
464	Ca2+	4.77E-01	Yes	4.25E-01	1.42E-01	1.39E-01	3.50E-01	4.37E-01	5.24E-01	6.36E-01	8.72E-02
465	PM10	1.68E+01	Yes	1.69E+01	6.40E-01	1.60E+01	1.66E+01	1.69E+01	1.73E+01	1.83E+01	3.47E-01
466											
467	Factor 2										

Order of the operations



Input/Output files

- Concentration file
- Input file
-Output file
- Configuration file

Analyze Input Data

- Concentration/ Uncertainty
- Concentration Scatter plot
- Concentration Time Series
- Data exceptions

Base Model Results

- Residual Analysis
- O/P Scatter Plot
- O/P Time Series
- Profiles/ Contributions
- Aggregate Contributions
- G-Space Plot
- Factor Pie Chart
- Diagnostics

Bootstrap Model Results

- Box Plots
- Summary

Fpeak Model Results

- Profiles/ Contributions
- G-Space
- Diagnostics

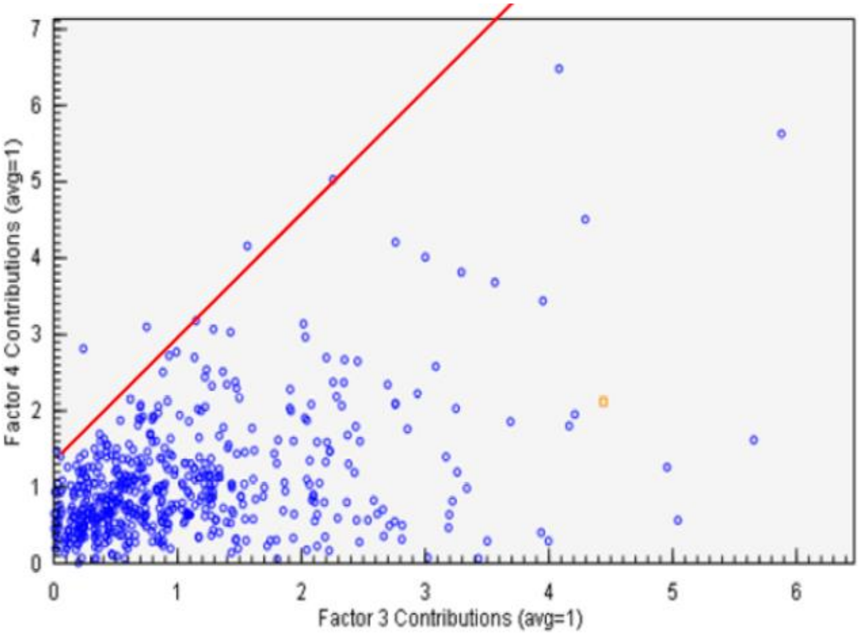
Fpeak runs

Fpeak runs

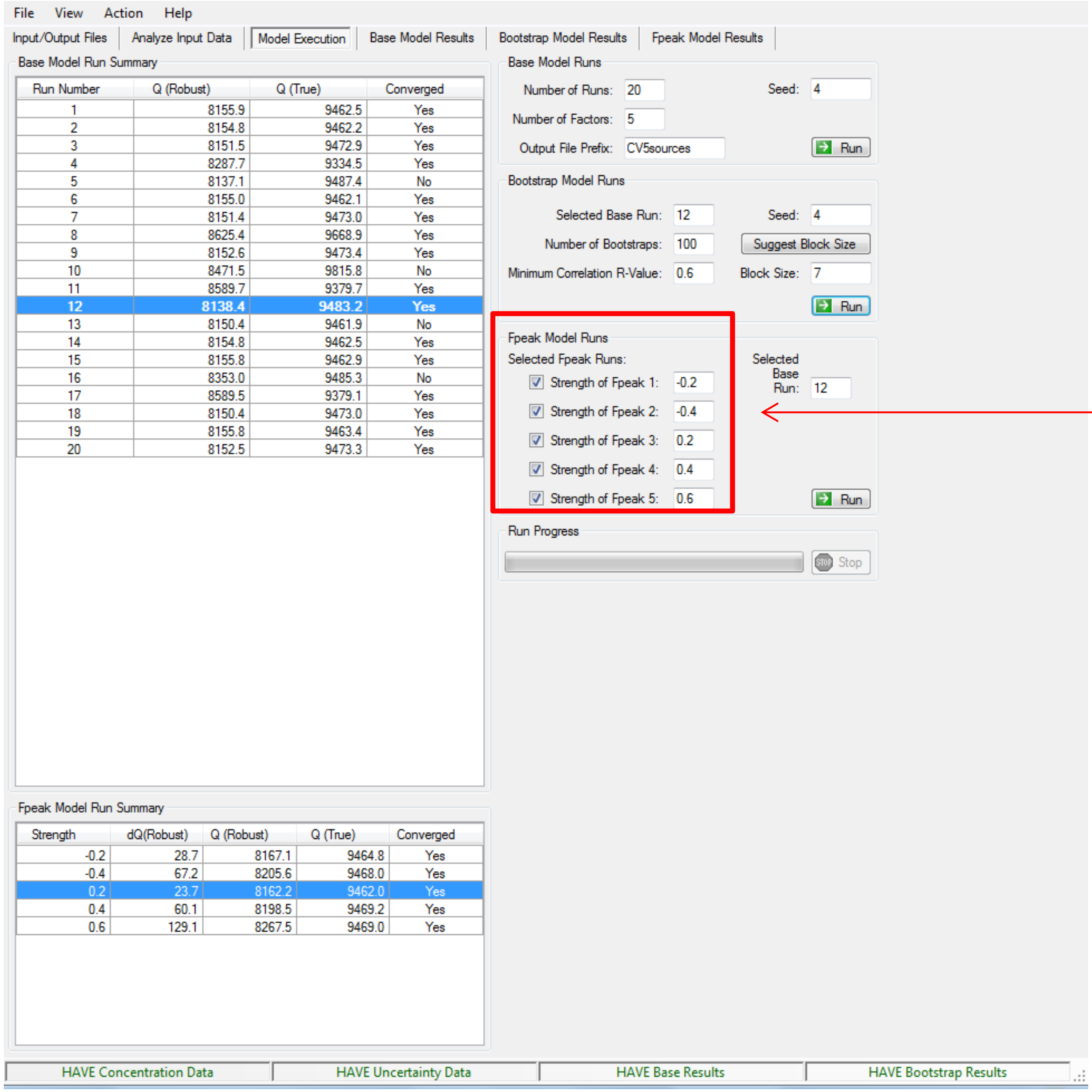
Rotation ambiguity

- The user should inspect the G-space plots for each pair of factors in the original solution to determine if an incorrect rotation has occured .

- In this case the solution must to be rotated to the real solution using **Fpeak**



Fpeak runs



Base Model Run Summary

Run Number	Q (Robust)	Q (True)	Converged
1	8155.9	9462.5	Yes
2	8154.8	9462.2	Yes
3	8151.5	9472.9	Yes
4	8287.7	9334.5	Yes
5	8137.1	9487.4	No
6	8155.0	9462.1	Yes
7	8151.4	9473.0	Yes
8	8625.4	9668.9	Yes
9	8152.6	9473.4	Yes
10	8471.5	9815.8	No
11	8589.7	9379.7	Yes
12	8138.4	9483.2	Yes
13	8150.4	9461.9	No
14	8154.8	9462.5	Yes
15	8155.8	9462.9	Yes
16	8353.0	9485.3	No
17	8589.5	9379.1	Yes
18	8150.4	9473.0	Yes
19	8155.8	9463.4	Yes
20	8152.5	9473.3	Yes

Base Model Runs

Number of Runs: 20

Seed: 4

Number of Factors: 5

Output File Prefix: CV5sources

Bootstrap Model Runs

Selected Base Run: 12

Seed: 4

Number of Bootstraps: 100

Minimum Correlation R-Value: 0.6

Block Size: 7

Fpeak Model Runs

Selected Fpeak Runs:

☑ Strength of Fpeak 1: -0.2

☑ Strength of Fpeak 2: -0.4

☑ Strength of Fpeak 3: 0.2

☑ Strength of Fpeak 4: 0.4

☑ Strength of Fpeak 5: 0.6

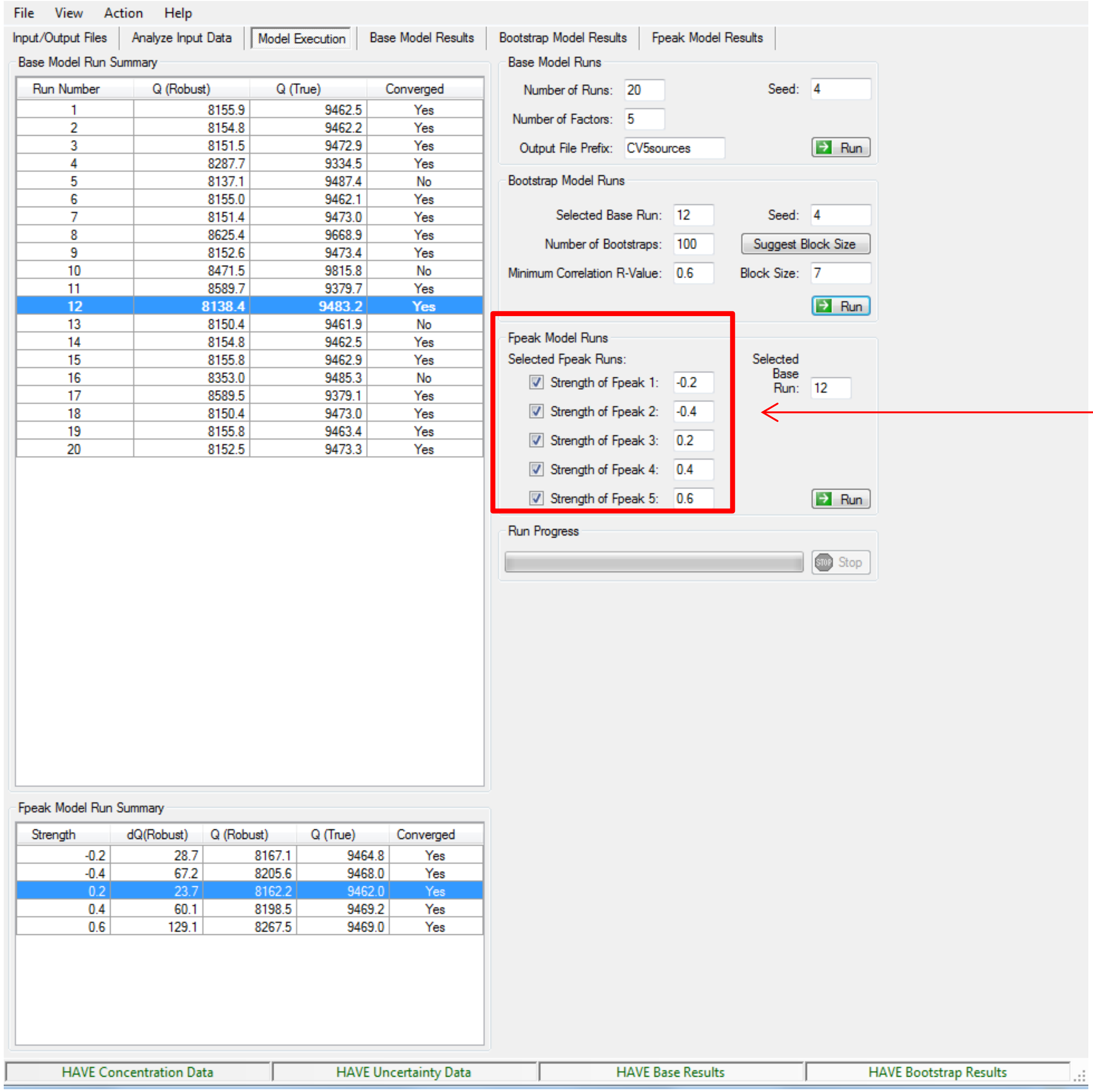
Selected Base Run: 12

Fpeak Model Run Summary

Strength	dQ(Robust)	Q (Robust)	Q (True)	Converged
-0.2	28.7	8167.1	9464.8	Yes
-0.4	67.2	8205.6	9468.0	Yes
0.2	23.7	8162.2	9462.0	Yes
0.4	60.1	8198.5	9469.2	Yes
0.6	129.1	8267.5	9469.0	Yes

Values should be between -5 and 5 (Fpeak ≠0)

Fpeak runs



Base Model Run Summary

Run Number	Q (Robust)	Q (True)	Converged
1	8155.9	9462.5	Yes
2	8154.8	9462.2	Yes
3	8151.5	9472.9	Yes
4	8287.7	9334.5	Yes
5	8137.1	9487.4	No
6	8155.0	9462.1	Yes
7	8151.4	9473.0	Yes
8	8625.4	9668.9	Yes
9	8152.6	9473.4	Yes
10	8471.5	9815.8	No
11	8589.7	9379.7	Yes
12	8138.4	9483.2	Yes
13	8150.4	9461.9	No
14	8154.8	9462.5	Yes
15	8155.8	9462.9	Yes
16	8353.0	9485.3	No
17	8589.5	9379.1	Yes
18	8150.4	9473.0	Yes
19	8155.8	9463.4	Yes
20	8152.5	9473.3	Yes

Base Model Runs

Number of Runs: 20
Seed: 4
Number of Factors: 5
Output File Prefix: CV5sources

Bootstrap Model Runs

Selected Base Run: 12
Seed: 4
Number of Bootstraps: 100
Minimum Correlation R-Value: 0.6
Block Size: 7

Fpeak Model Runs

Selected Fpeak Runs:

- ☑ Strength of Fpeak 1: -0.2
- ☑ Strength of Fpeak 2: -0.4
- ☑ Strength of Fpeak 3: 0.2
- ☑ Strength of Fpeak 4: 0.4
- ☑ Strength of Fpeak 5: 0.6

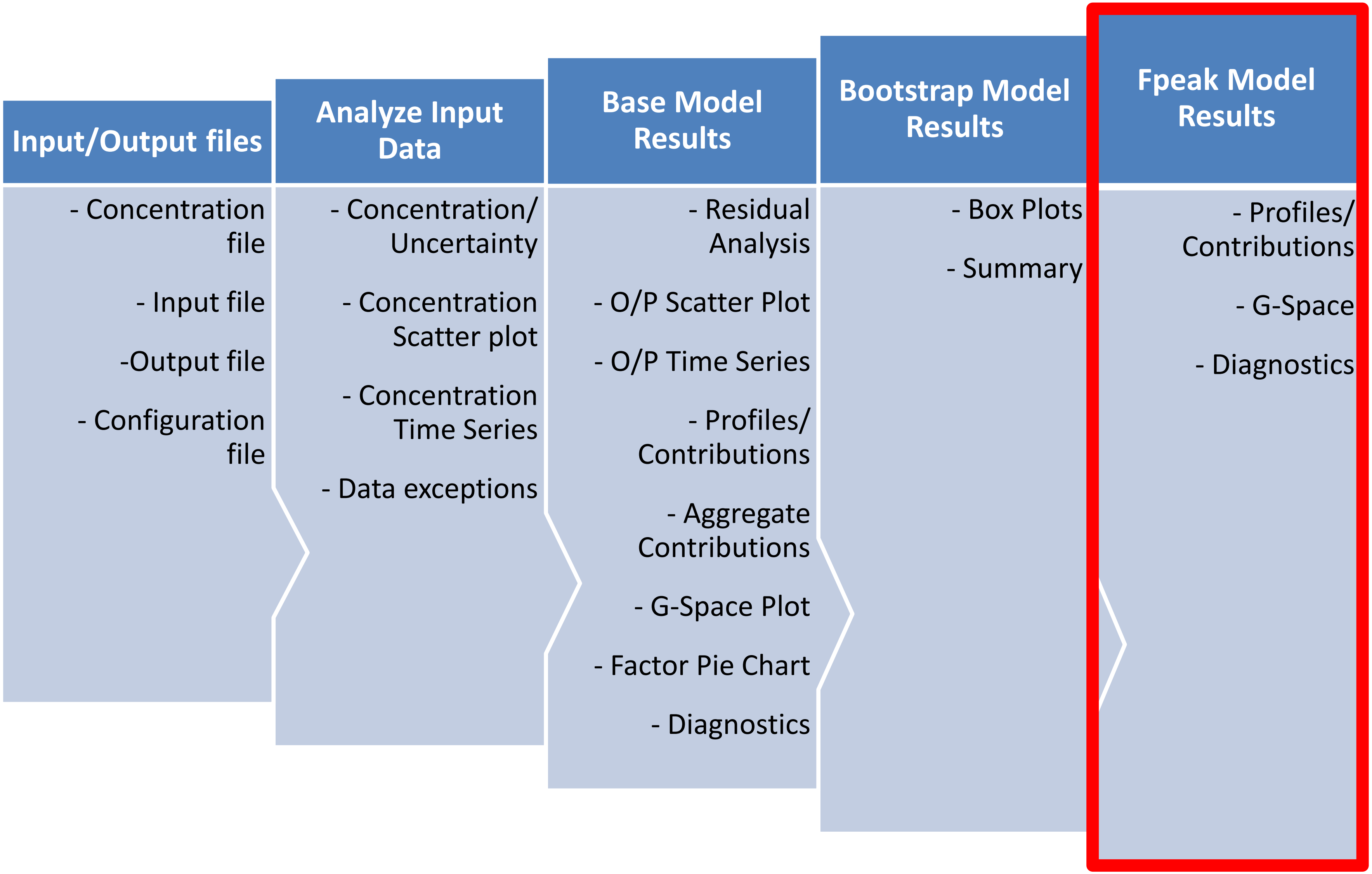
Selected Base Run: 12

Fpeak Model Run Summary

Strength	dQ(Robust)	Q (Robust)	Q (True)	Converged
-0.2	28.7	8167.1	9464.8	Yes
-0.4	67.2	8205.6	9468.0	Yes
0.2	23.7	8162.2	9462.0	Yes
0.4	60.1	8198.5	9469.2	Yes
0.6	129.1	8267.5	9469.0	Yes

Values should be between -5 and 5 (Fpeak ≠0)

Order of the operations



Input/Output files

- Concentration file
- Input file
-Output file
- Configuration file

Analyze Input Data

- Concentration/ Uncertainty
- Concentration Scatter plot
- Concentration Time Series
- Data exceptions

Base Model Results

- Residual Analysis
- O/P Scatter Plot
- O/P Time Series
- Profiles/ Contributions
- Aggregate Contributions
- G-Space Plot
- Factor Pie Chart
- Diagnostics

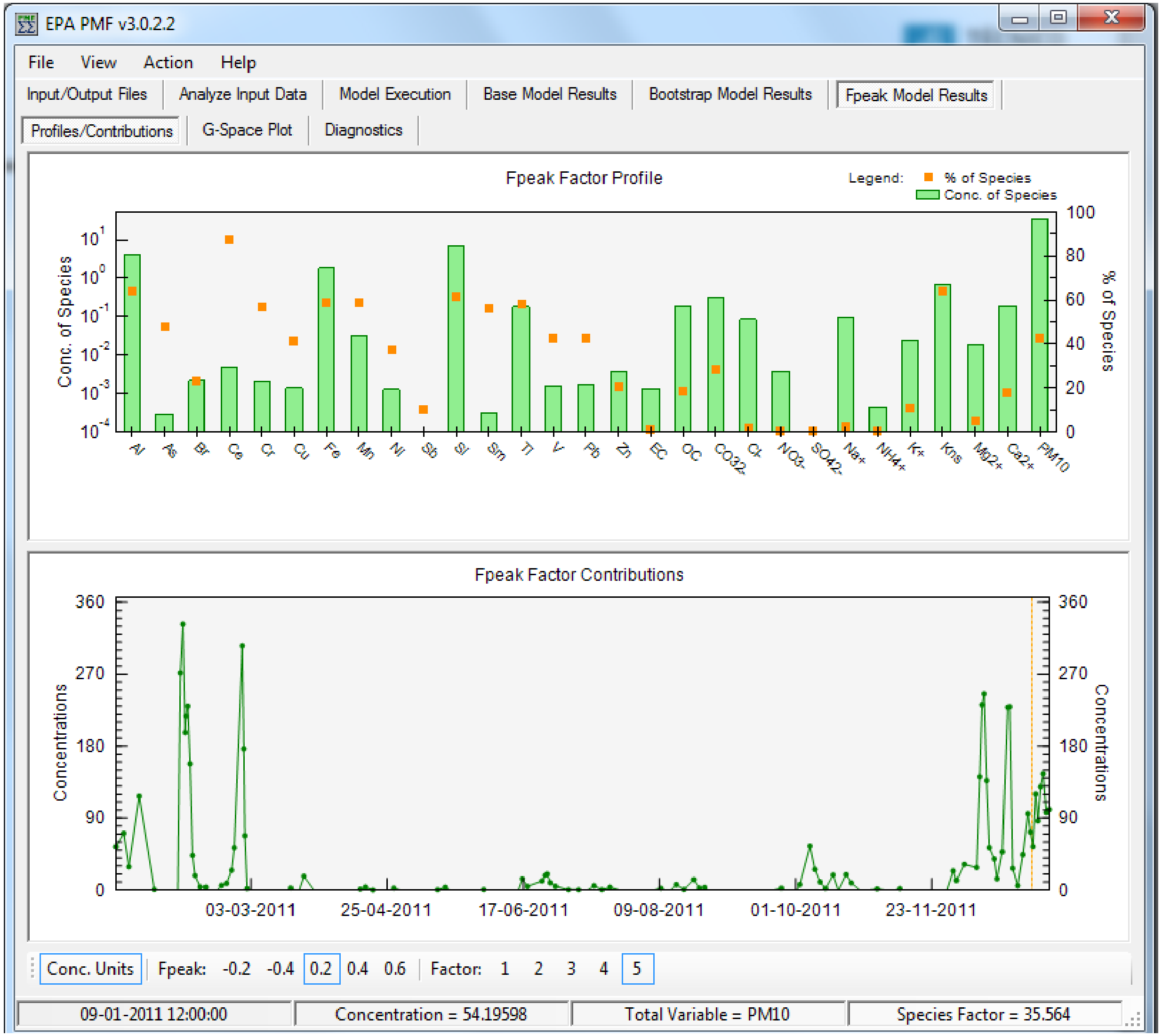
Bootstrap Model Results

- Box Plots
- Summary

Fpeak Model Results

- Profiles/ Contributions
- G-Space
- Diagnostics

Fpeak model results



Fpeak: 0.2

Factor: 5

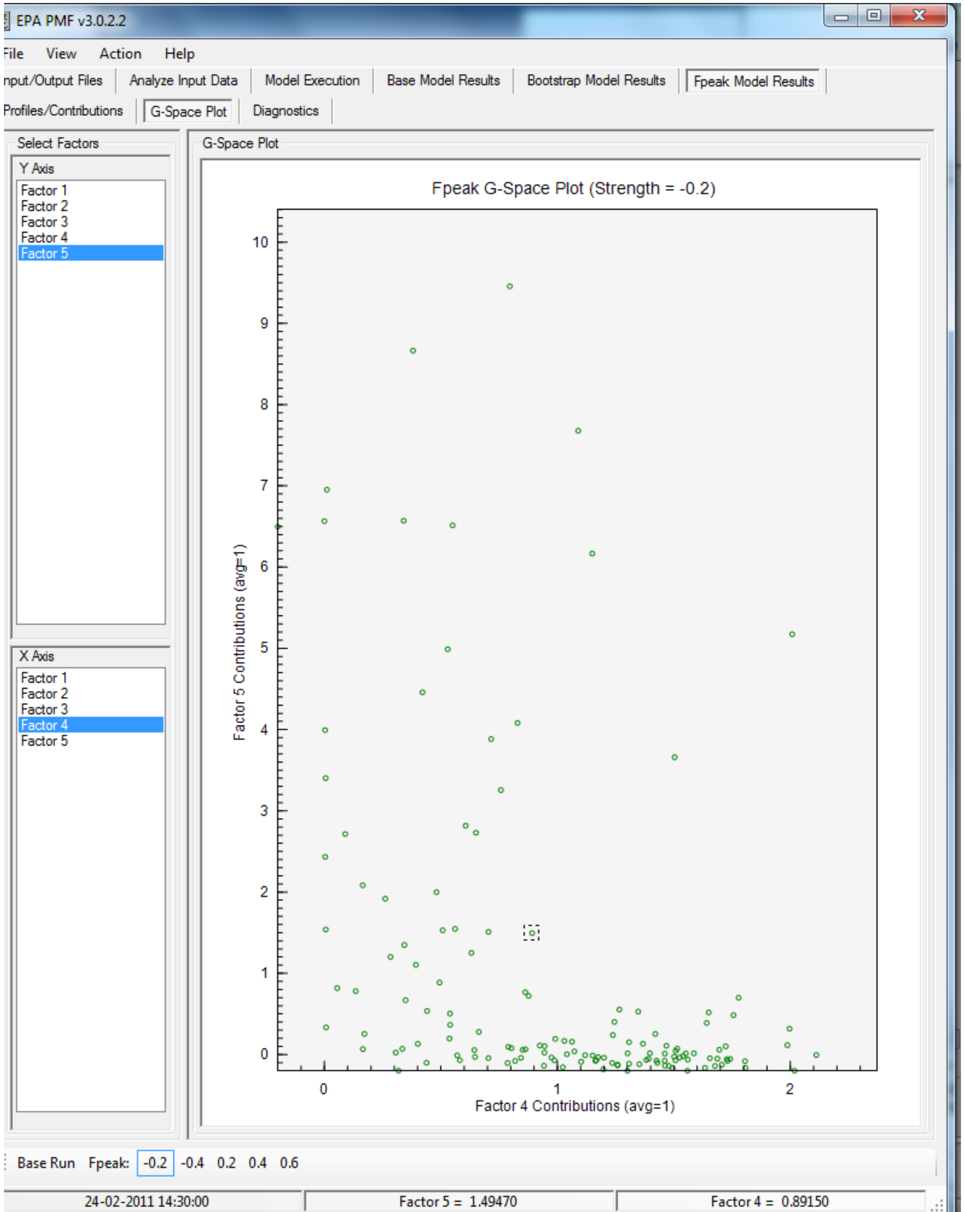
09-01-2011 12:00:00

Concentration = 54.19598

Total Variable = PM10

Species Factor = 35.564

Fpeak model results



Y Axis: Factor 5

X Axis: Factor 4

Fpeak: -0.2

24-02-2011 14:30:00

Factor 5 = 1.49470

Factor 4 = 0.89150

Thank you for your attention