Paul Scherrer Institute

The Source Finder (SoFi)

F. Canonaco, A. Tobler, C. Bozzetti, K. Dällenbach, A.S.H. Prévôt and many more

General – Program and logistics

Tuesday	
Time	*Activity*
	+++ LUNCH +++
13:30 – 15:00	• General information • Presentations of the participants
	+++ COFFEE BREAK +++
15:30 – 17:00	• Presentations of the participants • Software installation and download of sample data if not done so far

General – Program and logistics

Wednesday	
Time	*Activity*
9:00 – 10:30	• Theory input on PMF, ME-2, Q-space, robust mode, rotational tools (a-value)
	+++ COFFEE BREAK +++
10:50 – 12:30	• Guide through SoFi, main features (import raw data, treat data for PMF run, call ME-2)
	+++ LUNCH +++
13:30 – 15:00	• Explore PMF results in SoFi
	+++ COFFEE BREAK +++
15:30 – 17:00	• Explore PMF results in SoFi

General – Program and logistics

Thursday	
Time	*Activity*
9:00 – 10:30	• Offline
	+++ COFFEE BREAK +++
10:50 – 12:30	• Offline
	+++ LUNCH +++
13:30 – 15:00	• Individual work: participants work on their own data set (support provided)
	+++ COFFEE BREAK +++
15:30 – 17:00	• Individual work: participants work on their own data set (support provided)

General – Program and logistics

Friday	
Time	*Activity*
9:00 – 10:30	• Individual work: participants work on their own data set (support provided)
	+++ COFFEE BREAK +++
10:50 – 12:30	• Individual work: participants work on their own data set (support provided)
	+++ LUNCH +++
13:30 – 15:00	• SoFi Pro
	+++ COFFEE BREAK +++
15:30 – 17:00	• SoFi Pro

General information

- Current policy using SoFi standard only
  - collaboration for at least two peer-reviewed manuscripts per scientific group (F. Canonaco, A. Prevot, A. Tobler and PSI staff supporting your analysis during the workshop and beyond)
  - cite the SoFi paper in AMT (Canonaco et al. 2013)

- SoFi Pro
  - license-based
  - release this summer
  - No collaborative obligations

Pricing list

The following table summarizes the costs for a SoFi Pro license. Prices are given **per PC and year** and are reported without VAT. 25 % discount is applied for the purchase of multi-user or multi-year licenses.

	1 PC	multi-user (3 PCs)	multi-year 1 PC (3 years)	multi-user & -year (3 PCs and 3 years)
costs per PC & year in CHF	1000	750	750	500

IGOR-based software

- currently working with IGOR 6.37
- will be made compatible with IGOR 8 this winter

Main concept in SoFi

- Explorative analysis
  - exploration of a base case using a priori information, constrain what is known to find out what is unknown (iterative way)
- Assess uncertainties of base case
  - rotational uncertainties using random a values (soon with pulling equations DISP-like) and statistical uncertainties using bootstrap

Canonaco et al., Atmos Meas. Tech., 6, 3649–3661, 2013
Canonaco et al., Atmos Meas Tech., in prep.
A. Vlachou *et al.*, Atmos. Chem. Phys. Discuss., in review, 2018.
P. Rai *et al.*, Atmos. Chem. Phys. Discuss., in review, 2019.

General – Download area

ftp://datalystica.com

- Access using:
  username: sofi_workshop
  password: sofi_password

- /Data contains the PMF input (SoFi standard for Wednesday)

- /Results contains the result folder test_II with a simple rolling PMF result (SoFi Pro for Friday)
  – Save it under …\ME2_engine\ME2_Result

PMF – Useful ressources

- Francesco Canonaco @ PSI and datalystica: francesco.canonaco@psi.ch or francesco.canonaco@datalystica.com
- Carlo Bozzetti @ datalystica: carlo.bozzetti@datalystica.com
- Anna Tobler @ PSI: anna.tobler@psi.ch
- Kaspar Dällenbach from Helsinki: kaspar.dallenbach@helsinki.fi

- **Website** : https://www.datalystica.com
- Subscribe to our **mailing list**: info@datalystica.com (check on www.datalystica.com)

PMF – Useful ressources

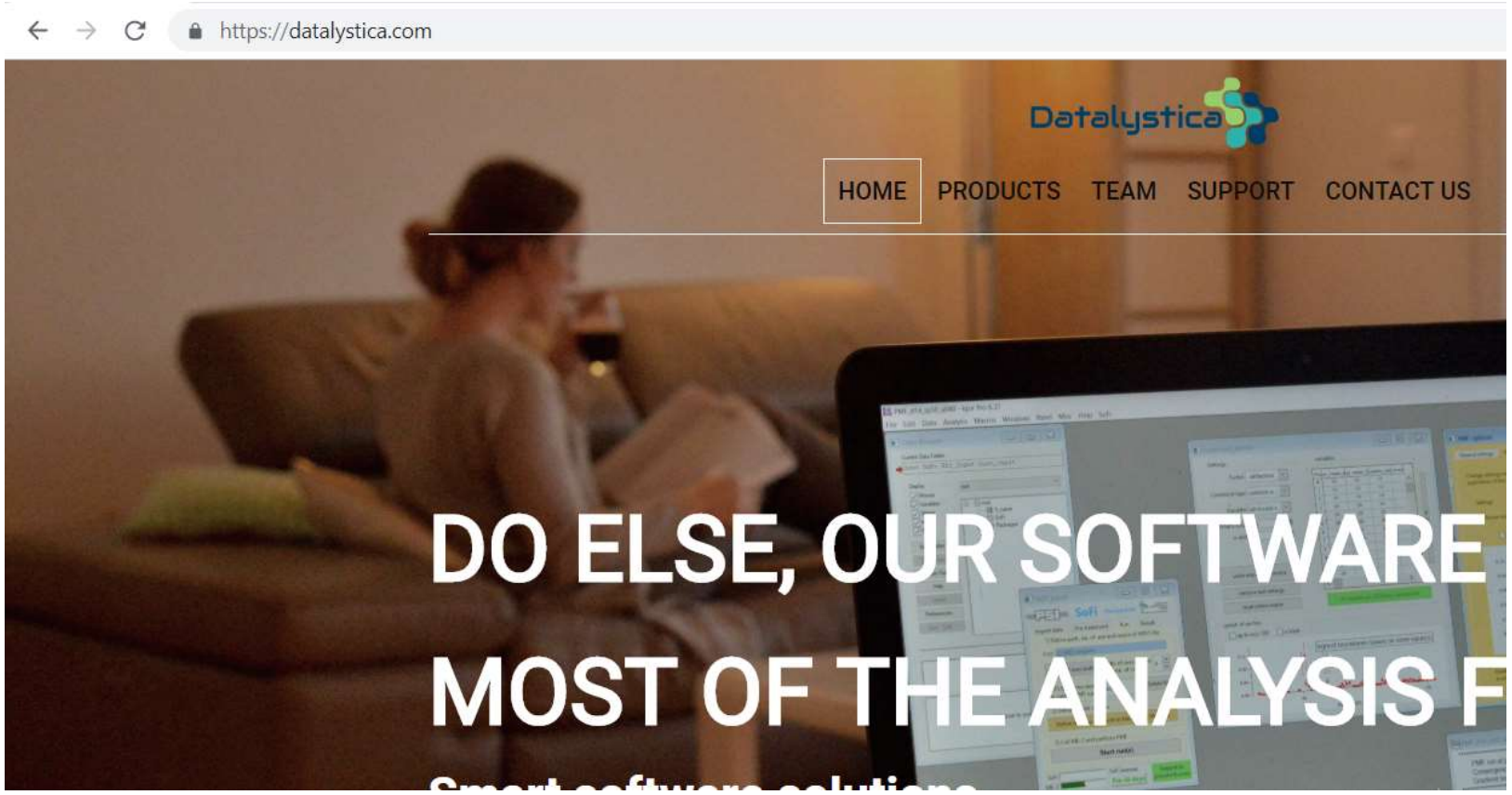
- Source Apportionment Guide :
  European guide on air pollution source apportionment with receptor models (2013)

Useful Ressources

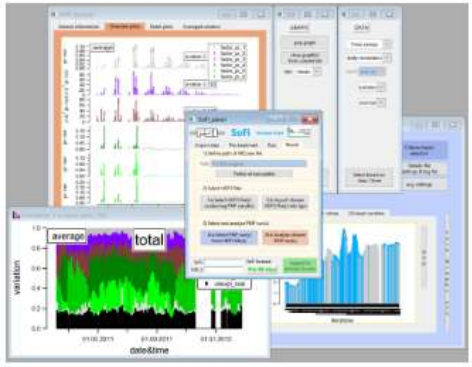
- Paatero's papers : (1994, 1999, 2009) on Positive Matrix Factorization and Multi-Linear Engine
- Canonaco's paper : (2013), on SoFi's toolkit http://www.atmos-meas-tech.net/6/3649/2013/
- Zhang's paper : (2011), Review on PMF https://link.springer.com/article/10.1007/s00216-011-5355-y
- Wiki page by Jimenez Group (for AMS data set) :
  - http://cires1.colorado.edu/jimenez-group/wiki/index.php/PMF-AMS_Analysis_Guide
  - AMS Spectral Database (HR) : http://cires1.colorado.edu/jimenez-group/HRAMSsd/
  - AMS Spectral Database (UMR) : http://cires1.colorado.edu/jimenez-group/AMSsd/

Datalystica

- responsible for the technical support, maintenance and further development of SoFi



General – Datalystica



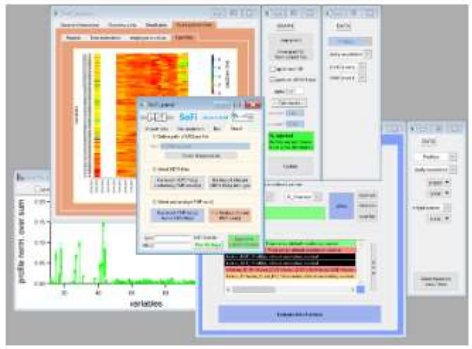
SoFi standard

The Source Finder (SoFi) software allows to efficiently analyze your multivariate data with factor analytic tools.

Read more information on SoFi and how to acquire the SoFi standard package.

freeware

Sign up to the mailing list: info@datalystica.com for the credentials



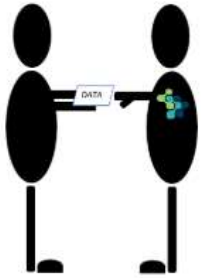
SoFi Pro

The SoFi Pro package offers a myriad of relevant features that highly enhance and support you along your data analysis journey.

Learn more about the features of SoFi Pro and how to acquire a SoFi Pro license.

license-based

Licenses can be purchased starting from this July in the datalystica kiosk



Data analysis service

Are you struggeling with your data? Not sufficient time for the data anylsis or missing the expertise in your own company?

Read more information on our data analysis service.

General – Presentation of the participants

Name	Last Name	Affiliation
Aliki	Christodoulou	The Cyprus Institute
Anja	Tremper	Aerosol Science Team, King's College London
Anna	Font	Aerosol Science Team, King's College London
Axel	Eriksson	Lund University
Bojan	Radvić	Mining and metallurgy Institute Bor
Cristina Antonia	Marin	National Institute of Research and Development for Optoelectronics, INOE2000
Ettore	Petralia	ENEA
Hannes	Keernik	Estonian Environmental Research Centre
Harald	Flentje	Deutscher Wetterdienst
Julija	Pauraite-Dudek	SRI Center for Physical Sciences and Technology
Katarzyna	Styszko	AGH University of Science and Technology, Faculty of Energy and Fuels
Katja	Dzepina	University of Rijeka, Rijeka, Croatia
Liene	Sustere	University of Latvia, Faculty of Geography and Earth Sciences
Marco	Paglione	CNR-ISAC
Marta	Via Gonzalez	IDAEA - CSIC
Maurizio	Gualtieri	ENEA-SSPT-MET-INAT
Otakar	Makeš	Institute of Chemical Process Fundamentals of the CAS
Petra	Pokorna	Department of Aerosol Chemistry and Physics, Institute of Chemical Process
Renata	Kovačević	Mining and metallurgy Institute Bor
Rosa Maria	Flores Rangel	Marmara University
Stephen	Platt	Norwegian institute for air research-NILU

PAUL SCHERRER INSTITUT

PSI

PSI, 01.10.2018

Positive Matrix Factorization (PMF) – general

Keywords:
(multi-) linear regression, PMF, CMB, PMF2, ME-2, Q-space, robust mode, seed runs, local/global minima, rotational ambiguity / uncertainty, model uncertainty

Model - Positive Matrix Factorization (PMF)

Simple linear regression

relationship between the dependent variable y and the regressor x is linear.
This relationship is modeled through an error variable ε (random variable that adds "noise" to the linear relationship):

$$y_i = \alpha + \beta x_i + \varepsilon_i .$$

Dependent variable (vector)

Intercept, if present

Slope of x (scalar value)

Independent variable or regressor (vector)

Random error to be minimized in object function Q

https://en.wikipedia.org/wiki/Simple_linear_regression (last visited on the 03,06,2018)

Model - Positive Matrix Factorization (PMF)

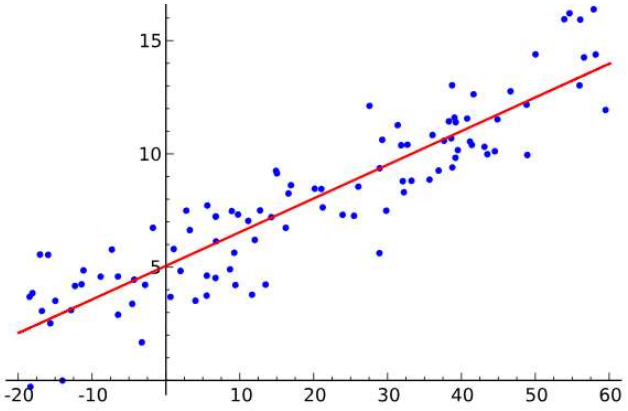
Simple linear regression

Ordinary least squares model (OLS) minimizes the residuals ε (on the y-axis only, assuming known regressor **X**) for the following object function and finding the slope β:

$$\text{Find } \min_{\alpha,\,\beta} Q(\alpha,\beta), \quad \text{for } Q(\alpha,\beta) = \sum_{i=1}^{n} \hat{\varepsilon}_i^{\,2} = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \,.$$

With the analytical solution for α and β:

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta}\,\bar{x}, \\ \hat{\beta} &= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\ &= \frac{\operatorname{Cov}(x,y)}{\operatorname{Var}(x)} \end{aligned}$$



https://en.wikipedia.org/wiki/Simple_linear_regression (last visited on the 03,06,2018)

Model - Positive Matrix Factorization (PMF)

Multivariate linear regression

relationship between the dependent variable **Y** and the regressor **X** is linear.
This relationship is modeled by minimizing the residual ε (random variable that adds "noise" to the linear relationship):

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{i1} + \beta_{2j}X_{i2} + \ldots + \beta_{pj}X_{ip} + \epsilon_{ij}$$

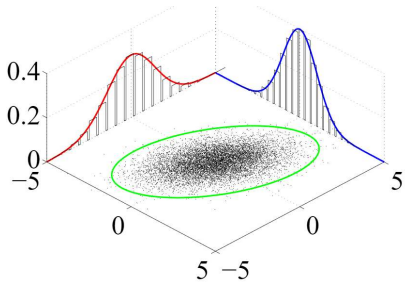
Dependent variable (matrix)

Intercept, if present

Slope of x_{ij}, (matrix)

Independent variable or regressor (matrix)

Error to be minimized in object function Q to become Gaussian-like



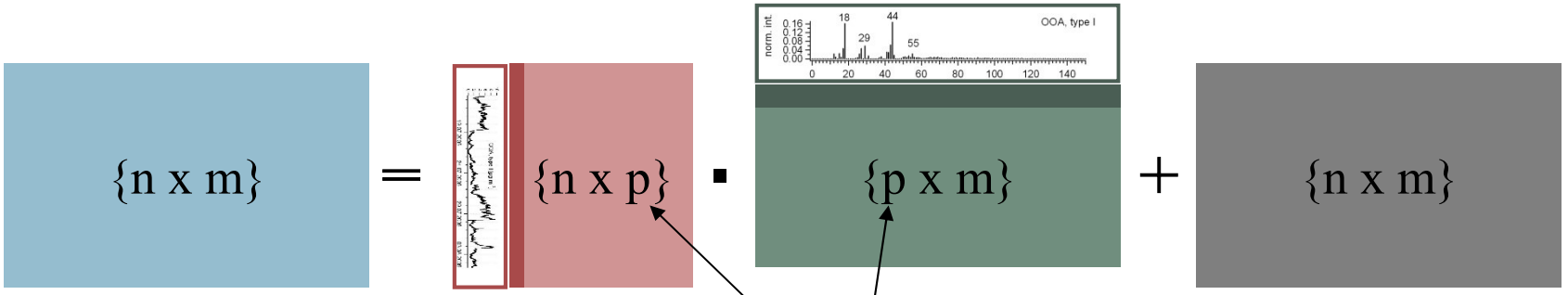
https://en.wikipedia.org/wiki/linear_regression (last visited on the 03,06,2018)

Model - Positive Matrix Factorization (PMF)

Bilinear factor analytic algorithm

$$X_{measured} = G \cdot F + E$$

mass spectral matrix — factor contribution — factor profile — residual matrix



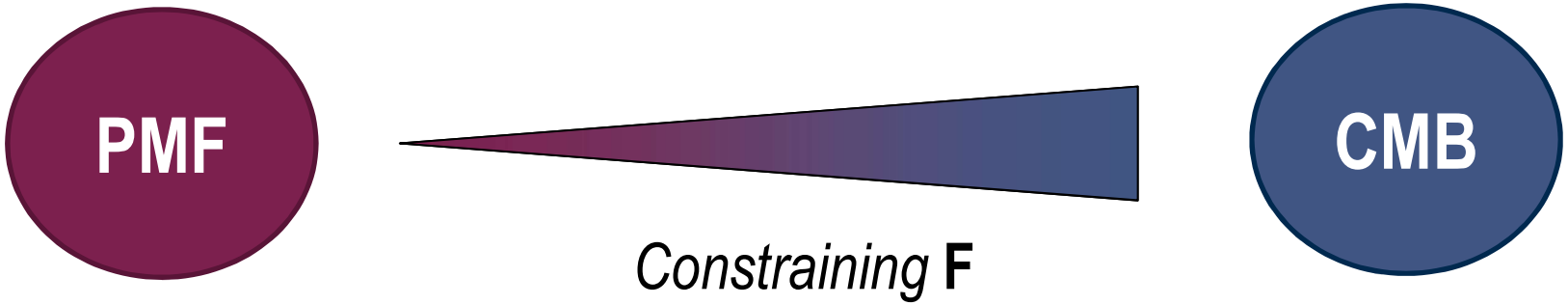
Values in G & F are non-negative
Factors represent sources / processes

Goal
Factor solution must be environmentally reasonable

Paatero, P., Tapper, U., 1994. Environmetrics 5, 111-126.

Model – PMF/CMB/solvers

- PMF / CMB (chemical mass balance) approach



- Solvers

Solver	Unconstrained	Constrained	Communication
PMF2 / PMF3	X	only to zero	Limited
ME-2	X	X	All quantities easily accessible

Paatero, P., 1999. J Comput Graph Stat 8, 854-888.

Model – Positive Matrix Factorization (PMF)

Least-squares algorithm

$$Q = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\frac{e_{ij}}{\sigma_{ij}}\right)^2$$

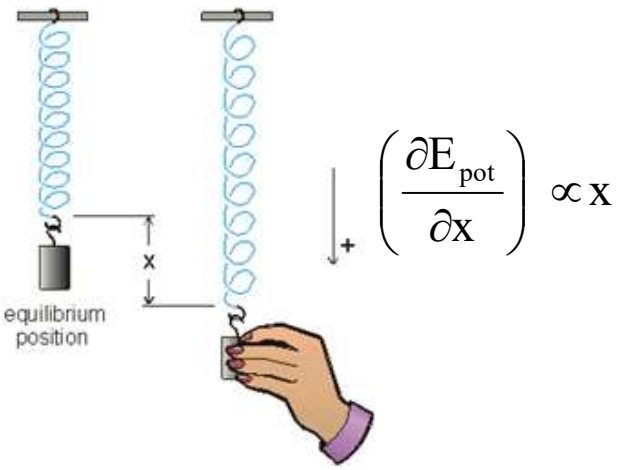
e_{ij}: difference (measured – model)

σ_{ij}: uncertainty (statistical error)

- Q will be minimized with respect to all model variables
  - ME-2 starts the conjugate gradient algorithm for solving this task (multivariate linear regression)

- Quantity minimized in ME-2 scales with the residual

$$\left(\frac{\partial Q_{ij}}{\partial e_{ij}}\right) \propto e_{ij}$$



Paatero, P., 1997. Chemometr Intell Lab 37, 23-35.

PMF run (non-robust mode)

- Minimized quantity is proportional to the residual (in theory ideal)
- Outliers, e.g. transient sources, wrong nb. of factors, electronic recording issues, etc. violate this relation and PMF could spend more time, reducing "wrong" entries in **Q**

PMF run (robust mode)

- Allow for this dependency only in a certain range and damp afterwards (robust mode, default value = 4)

$$\text{if} \left| \frac{e_{ij}}{\sigma_{ij}} \right| \leq 4 \Rightarrow \left(\frac{\partial Q_{ij}}{\partial e_{ij}} \right) \propto e_{ij} \qquad \text{else} \left| \frac{e_{ij}}{\sigma_{ij}} \right| > 4 \Rightarrow \left(\frac{\partial Q_{ij}}{\partial e_{ij}} \right) \propto 4$$

PMF – Q/Q_{exp}

Weight Q by Q_{exp}, the remaining degrees of freedom

$$Q_{exp} = n \cdot m - p \cdot (n + m) \sim n \cdot m$$

- If all residuals were similar as their σ's, Q / Q_{exp} ~1

- Monitor Q / Q_{exp} values (with and without the robust mode) → Too high values might indicate systematic problems of the PMF result
- Monitor the changes of Q/Q_{exp} over various model runs with different settings, e.g. nb. of factors, specific constraints

PMF - problems

Disadvantages

- Assess number of factors
- Constant factor profiles (mass spectra)
- Uncertainties are not fully defined, minimal Q-value is not necessarily the best solution
- Bilinear factor analytic models suffer from rotational ambiguity

$$\mathbf{X_{model}} = \mathbf{G} \cdot \mathbf{F} = \mathbf{G} \cdot \mathbf{T} \cdot \mathbf{T^{-1}} \cdot \mathbf{F} = \mathbf{G'} \cdot \mathbf{F'}$$

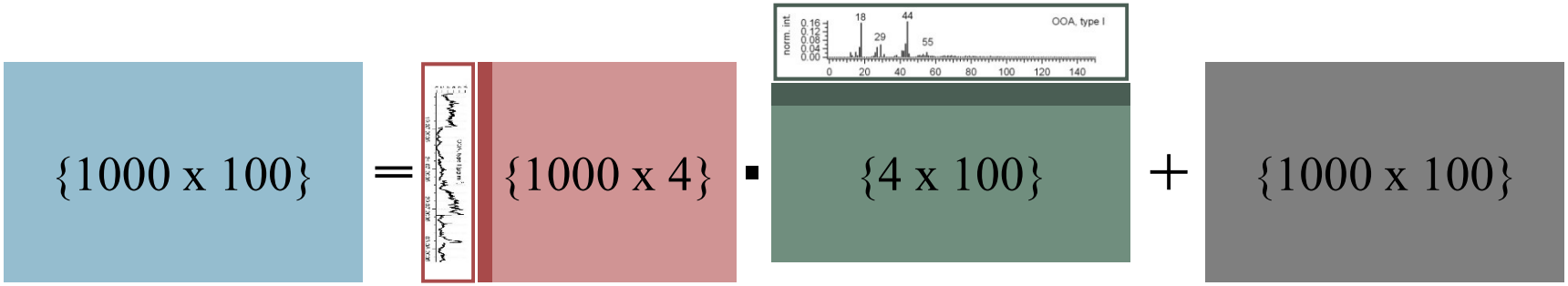
→ explore the PMF results

– vary number of factors

– vary the entries in G and F randomly (seed), controlled (e.g. a value, if good reasons) to find the global minimum/minima

PMF – Q-space

- Real case
  - ACSM data with 100 variables for 1000 scans, four factors, unconstrained



→ 4400 model variables

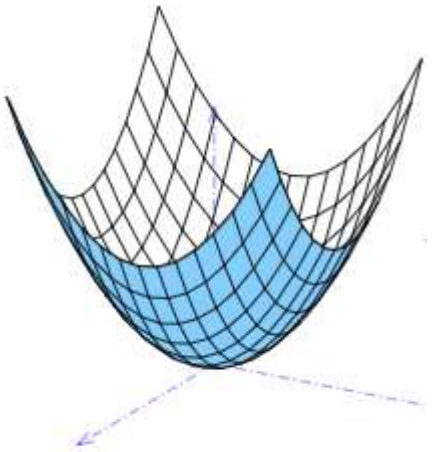
→ Q(4400 model variables), multidimensional Q-space

- Simplified case
  - Simply the real case with two model variables

→ Q(2 model variables), three dimensional Q-space

- Global minimum with two variables
- Paraboloidal function:

$$z = x^2 + y^2$$



- ME-2 finds the global minimum with one run, since every starting point on the plane rolls down following the steepest descent to the global minimum

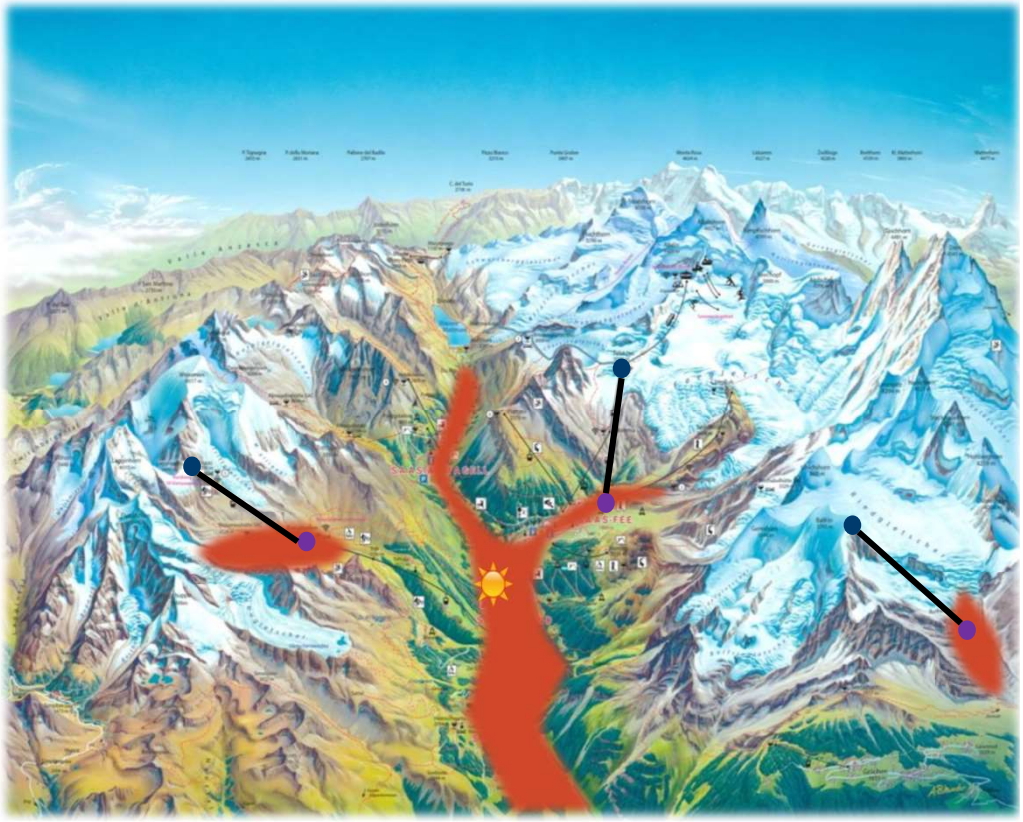
https://en.wikipedia.org/wiki/Paraboloid (last visited on the 03,06,2018)

PMF – rotational ambiguity

- Q(2 model variables) similar to the height h(x,y) in the map

- PMF is performed through the conjugate gradient algorithm minimizing Q based on the starting conditions (from blue to the purple)

- Goal is to find the smallest possible Q-value (global minimum) (red area) together with interpretable PMF runs (☀)

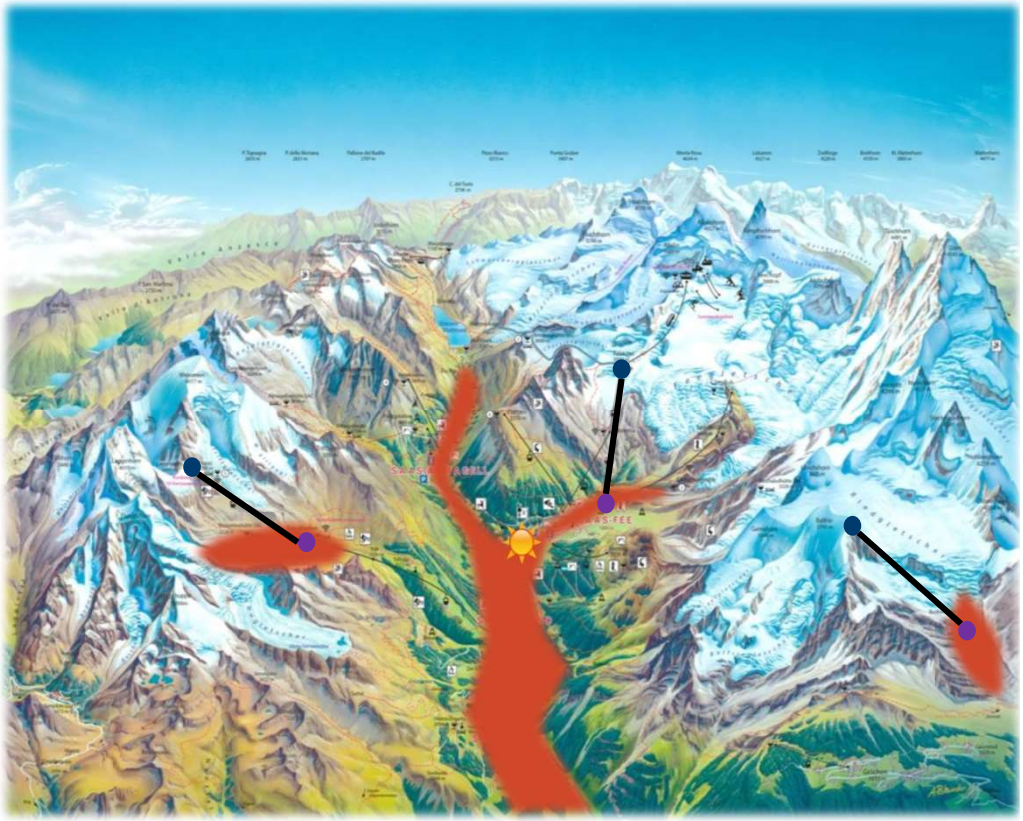


PMF – rotational ambiguity

- There are many points on the map, for which h(x,y) is equal → rotational ambiguity

- the task is tedious using only pure random entries in **G** and **F** → could take lots of PMF runs

- Explore the rotational ambiguity with proper techniques (global fpeak, individual fpeak, **a-value**, pulling equations)

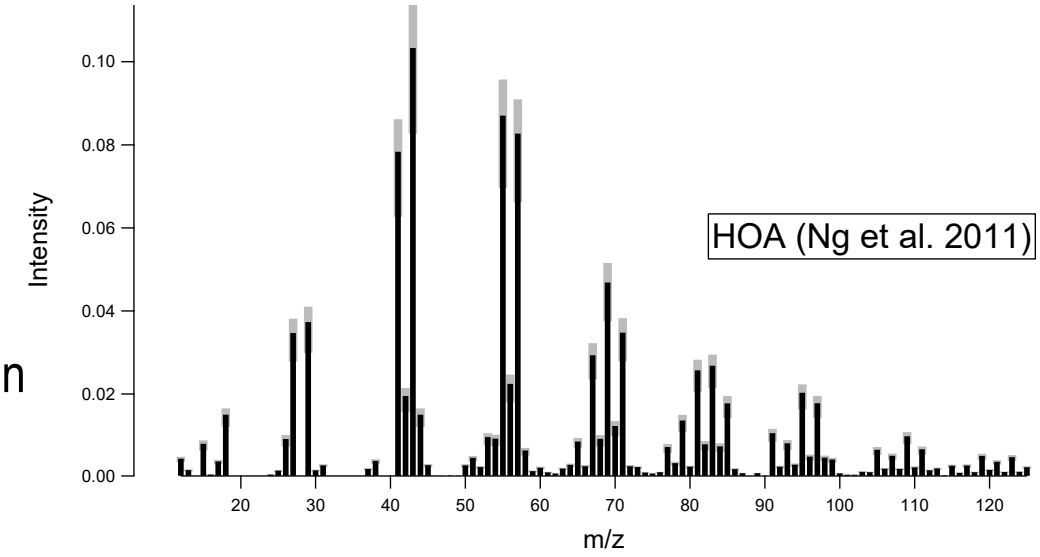


PMF – a-value technique

- full Q-space can potentially be investigated
- advantage: easy to perform and computationally inexpensive
- if **a priori information** available, wiser to confine solution space by constraining entries in **G** and / or **F**
- disadvantage: sensitivity analysis on the constrained model variables

PMF – a-value technique

- full Q-space can potentially be investigated
- advantage: easy to perform and computationally inexpensive
- if **a priori information** available, wiser to confine solution space by constraining entries in **G** and / or **F**
- disadvantage: sensitivity analysis on the constrained model variables

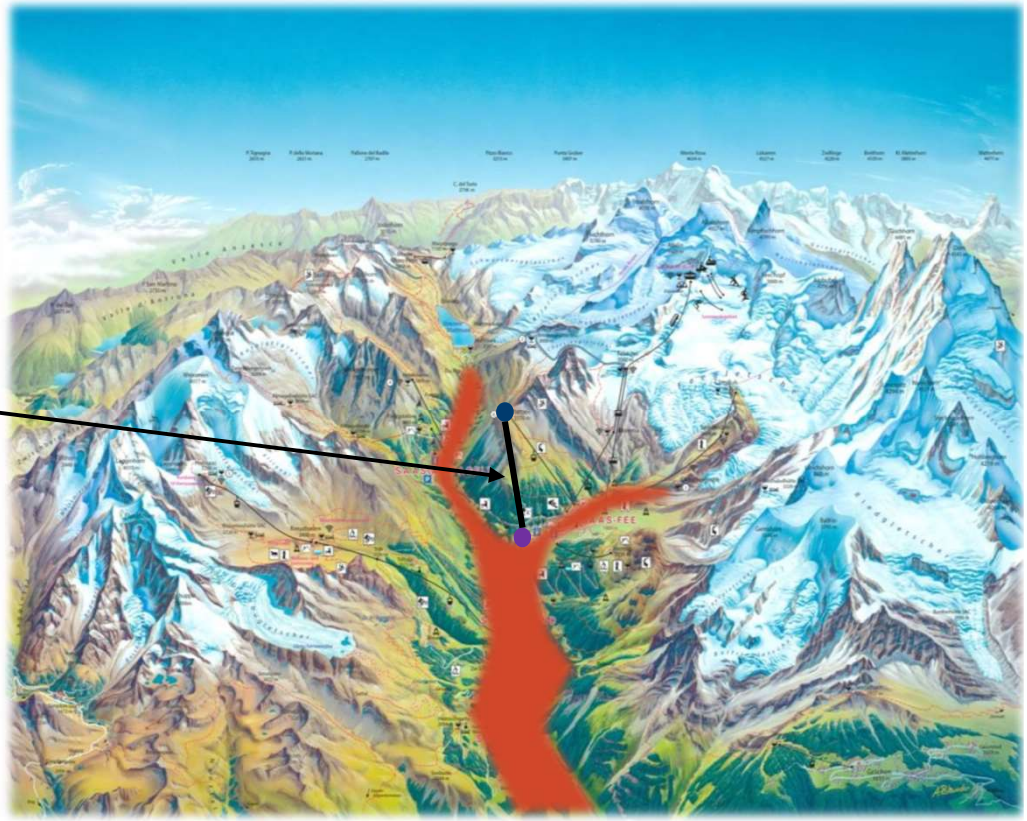


$$f_{p,j,solution} = f_{p,j} \pm a \cdot f_{p,j}$$

PMF – a-value / pulling technique

- a value approach allows to explore specific regions that wouldn't be visible in a pure unconstrained / see run

- Investigate the solution along the descent (a value)

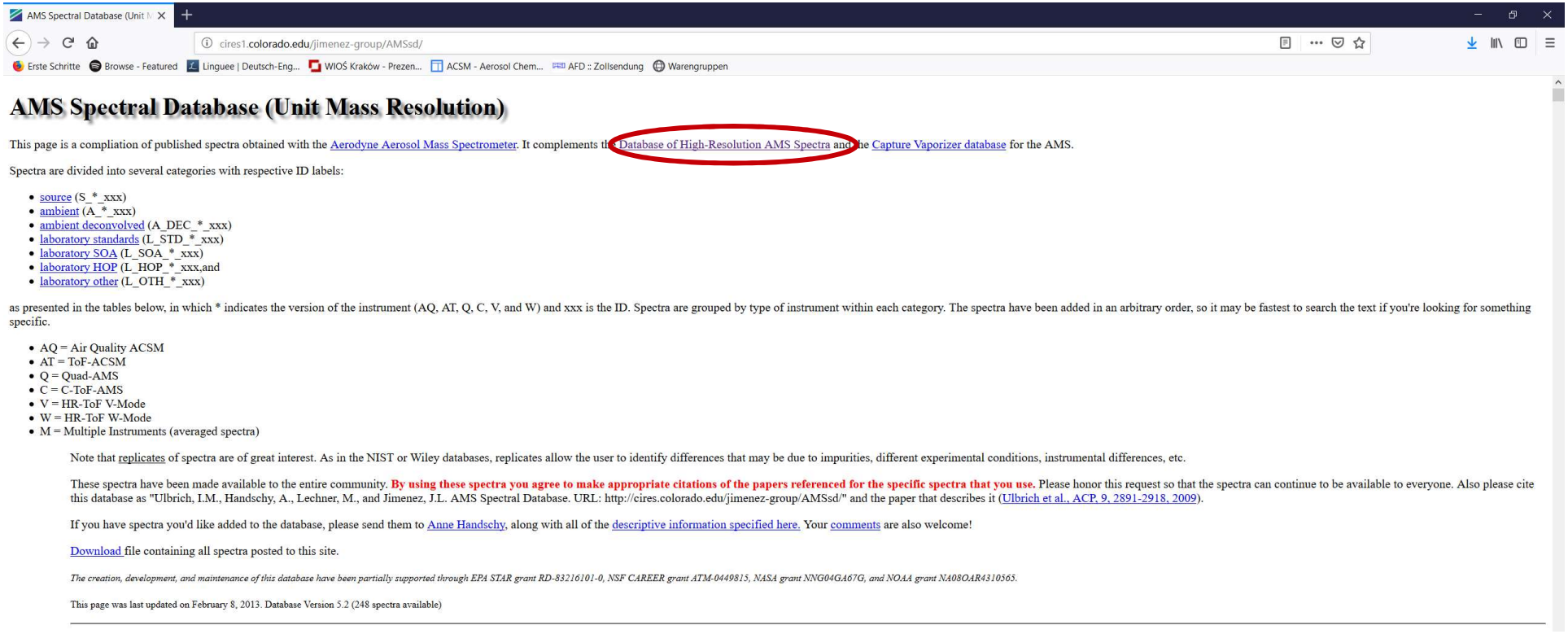


PMF – a-value technique

- sensitivity analysis performed on the constrained anchor does not find the good solution
- change position of initial value (dashed line), i.e., change factor profile

PMF – a priori information

- ~ a decade of know-how on AMS fingerprints of various POA & SOA, exploit this information in PMF



AMS Spectral Database (Unit Mass Resolution)

This page is a compliation of published spectra obtained with the Aerodyne Aerosol Mass Spectrometer. It complements the Database of High-Resolution AMS Spectra and the Capture Vaporizer database for the AMS.

Spectra are divided into several categories with respective ID labels:

- source (S_*_xxx)
- ambient (A_*_xxx)
- ambient deconvolved (A_DEC_*_xxx)
- laboratory standards (L_STD_*_xxx)
- laboratory SOA (L_SOA_*_xxx)
- laboratory HOP (L_HOP_*_xxx,and
- laboratory other (L_OTH_*_xxx)

as presented in the tables below, in which * indicates the version of the instrument (AQ, AT, Q, C, V, and W) and xxx is the ID. Spectra are grouped by type of instrument within each category. The spectra have been added in an arbitrary order, so it may be fastest to search the text if you're looking for something specific.

- AQ = Air Quality ACSM
- AT = ToF-ACSM
- Q = Quad-AMS
- C = C-ToF-AMS
- V = HR-ToF V-Mode
- W = HR-ToF W-Mode
- M = Multiple Instruments (averaged spectra)

Note that replicates of spectra are of great interest. As in the NIST or Wiley databases, replicates allow the user to identify differences that may be due to impurities, different experimental conditions, instrumental differences, etc.

These spectra have been made available to the entire community. **By using these spectra you agree to make appropriate citations of the papers referenced for the specific spectra that you use.** Please honor this request so that the spectra can continue to be available to everyone. Also please cite this database as "Ulbrich, I.M., Handschy, A., Lechner, M., and Jimenez, J.L. AMS Spectral Database. URL: http://cires.colorado.edu/jimenez-group/AMSsd/" and the paper that describes it (Ulbrich et al., ACP, 9, 2891-2918, 2009).

If you have spectra you'd like added to the database, please send them to Anne Handschy, along with all of the descriptive information specified here. Your comments are also welcome!

Download file containing all spectra posted to this site.

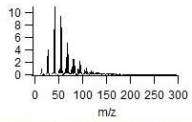
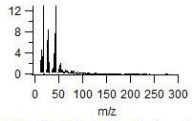
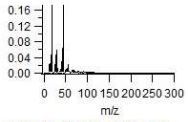
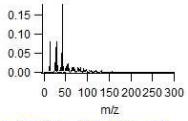
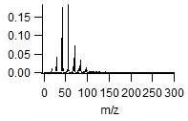
The creation, development, and maintenance of this database have been partially supported through EPA STAR grant RD-83216101-0, NSF CAREER grant ATM-0449815, NASA grant NNG04GA67G, and NOAA grant NA08OAR4310565.

This page was last updated on February 8, 2013. Database Version 5.2 (248 spectra available)

PMF – a priori information

Ambient Deconvolved Spectra

Freutel., F., Masters Thesis, Univ. Mainz, 2009 Freutel., F., Masters Thesis, Univ. Mainz, 2009 Freutel., F., Masters Thesis, Univ. Mainz, 2009 Freutel., F., Masters Thesis, Univ. Mainz, 2009

Spectra ID	Location and Date	Researcher/Group	AMS Instrument	EI Energy	Vaporizer Temp (°C)	Citation	Fig #	Comments	Data
A_DEC_Q_001	Pittsburgh, 09/2002	Qi Zhang Jimenez - Colorado	Q-AMS 014	70 eV	600	Zhang, Q., et al. **ES&T** 2005, *39*, 4938-4952.	Fig 9a1, pg. 4947	Zhang 2-component deconvolution	 A_DEC_Q_001_HOA_Pittsburgh.itx
A_DEC_Q_002	Pittsburgh, 09/2002	Qi Zhang Jimenez - Colorado	Q-AMS 014	70 eV	600	Zhang, Q., et al. **ES&T** 2005, *39*, 4938-4952.	Fig 9b1, pg. 4947	Zhang 2-component deconvolution	 A_DEC_Q_002_OOA_Pittsburgh.itx
A_DEC_Q_003	OOA, type I	V. Lanz PSI/Empa	Q-AMS	70 eV	600	Lanz, V. A., et al.,**Atmos. Chem. Phys.** 2007, *7*, 1503-1522.	Fig 3, pg. 1510	Zurich, summer 2005	 A_DEC_Q_003_OOA_I.itx
A_DEC_Q_004	OOA, type II	V. Lanz PSI/Empa	Q-AMS	70 eV	600	Lanz, V. A., et al.,**Atmos. Chem. Phys.** 2007, *7*, 1503-1522.	Fig 3, pg. 1510	Zurich, summer 2005	 A_DEC_Q_004_OOA_II.itx
A_DEC_Q_005	HOA	V. Lanz PSI/Empa	Q-AMS	70 eV	600	Lanz, V. A., et al.,**Atmos. Chem. Phys.** 2007, *7*, 1503-1522.	Fig 3, pg. 1510	Zurich, summer 2005	 A_DEC_Q_005_HOA.itx
A_DEC_Q_006	wood burning	V. Lanz	Q-AMS	70	600	Lanz, V. A., et	Fig 3, pg.	Zurich, summer 2005	

Global fpeak (ϕ) technique

- all rotations are performed at the same time
- advantage: easy to perform
- disadvantage: rotations cannot always be fully predicted

- example: three factors, rotation matrix **T** mixes all factor contributions and profiles together

$$\overline{G} = GT \text{ and } \overline{F} = T^{-1}F \qquad T_{\text{fpeak, p=3}} = \begin{bmatrix} 1 & \phi & \phi \\ \phi & 1 & \phi \\ \phi & \phi & 1 \end{bmatrix}$$

Paatero 2008

Individual fpeak (ϕ) technique

- all rotations are performed at the same time
- advantage: easy to perform
- disadvantage: rotations cannot always be fully predicted, lower estimate of the rotational uncertainty

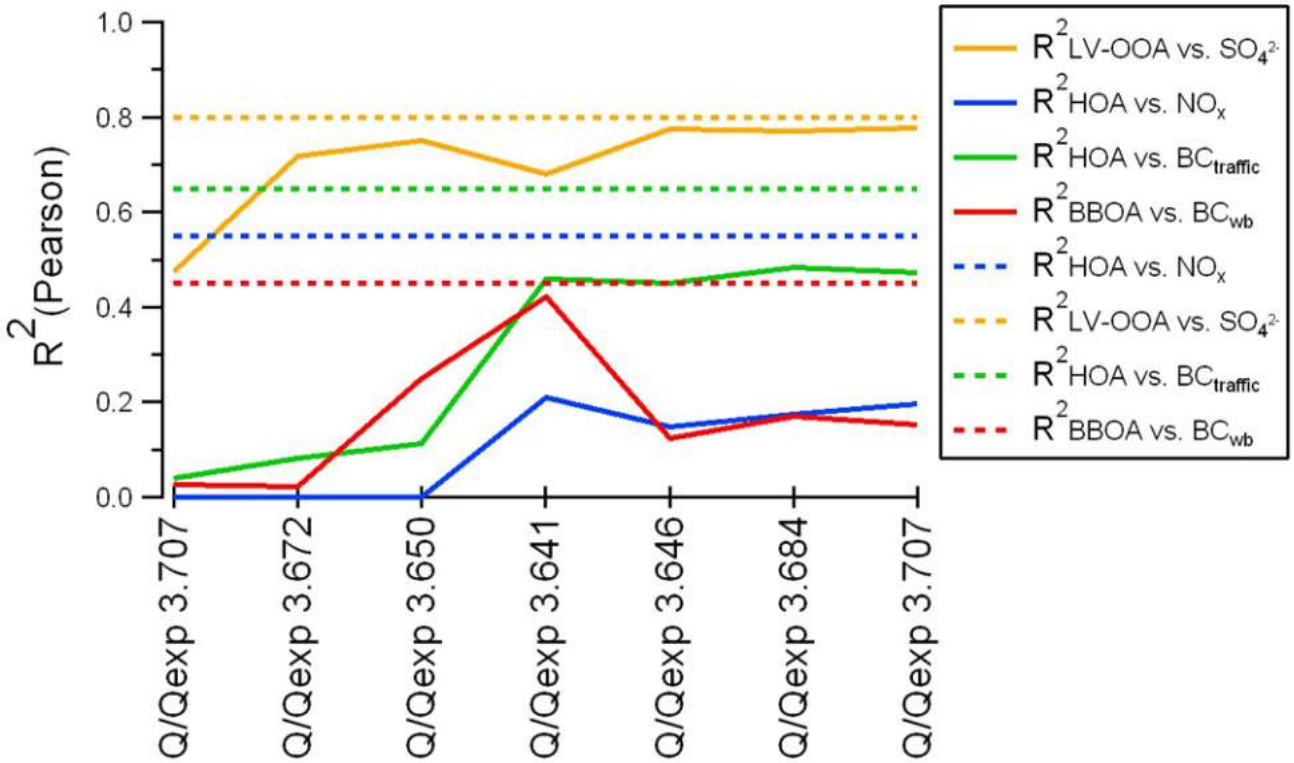
- example: three factors, rotation matrix **T** mixes only factor 3 with factor 1 *(adding contribution of factor 3 to that of factor 1 and subtracting profile of factor 3 from 1)*

$$\bar{G} = GT \text{ and } \bar{F} = T^{-1}F \qquad T_{p=3} = \begin{bmatrix} 1 & 0 & \phi \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Paatero 2008

f-peak vs. a-value approach

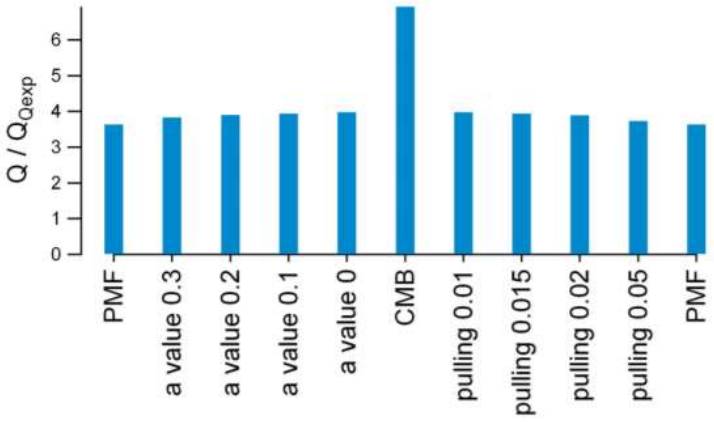
- P. Paatero and PSI **do recommend to use the a-value approach** (or in combination with the fpeak but not the global fpeak alone)

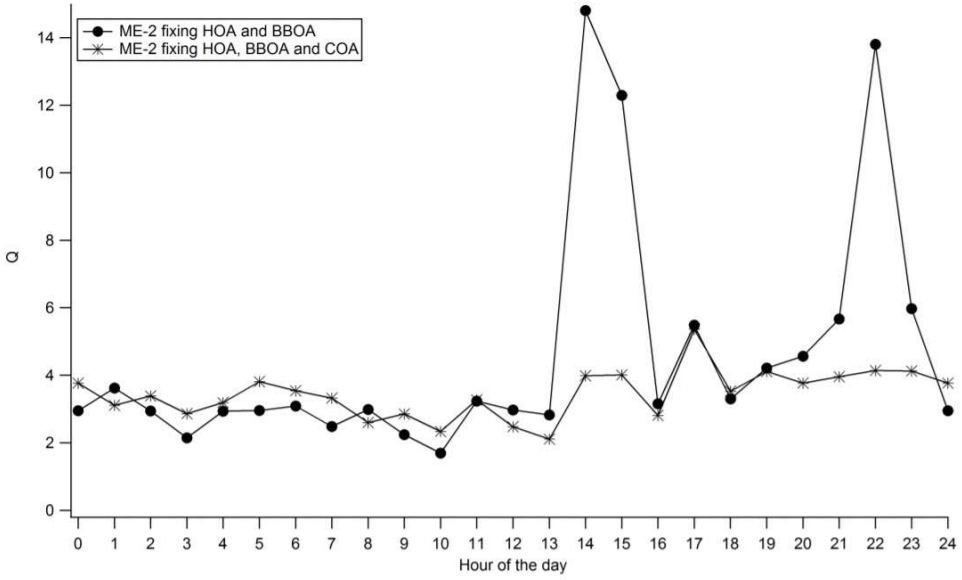


Canonaco et al., 2013

PMF results – goals

- **PMF results must be**
  - mathematically acceptable (similar Q, unstructured scaled residuals over **time** (ts, diurnal cycle, weekly cycle, etc.) and over **profile** (variables))



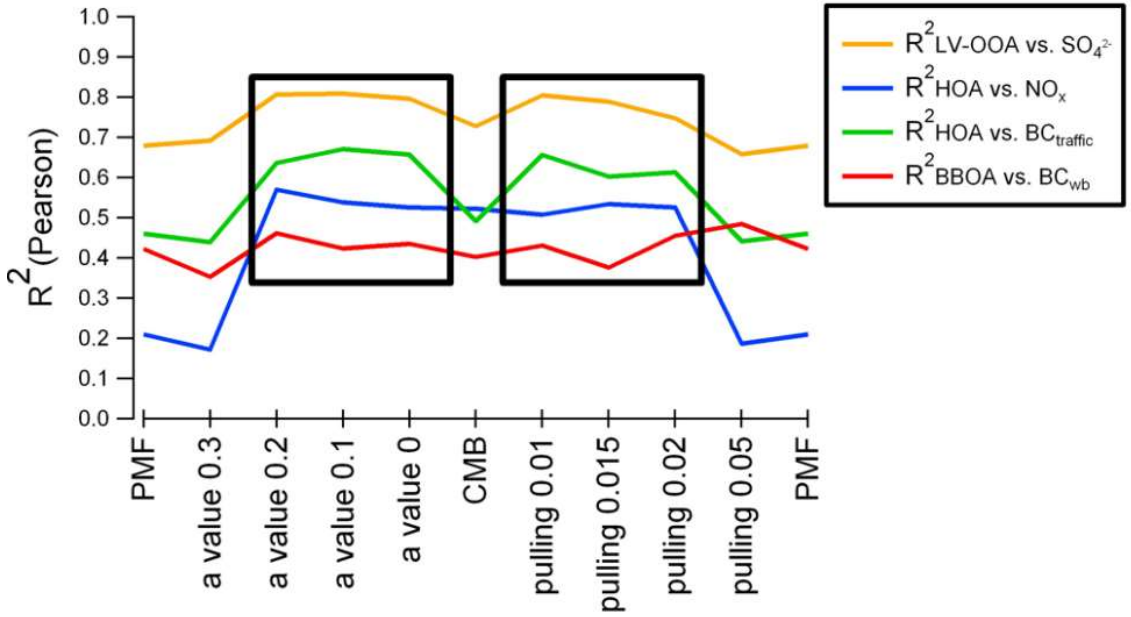


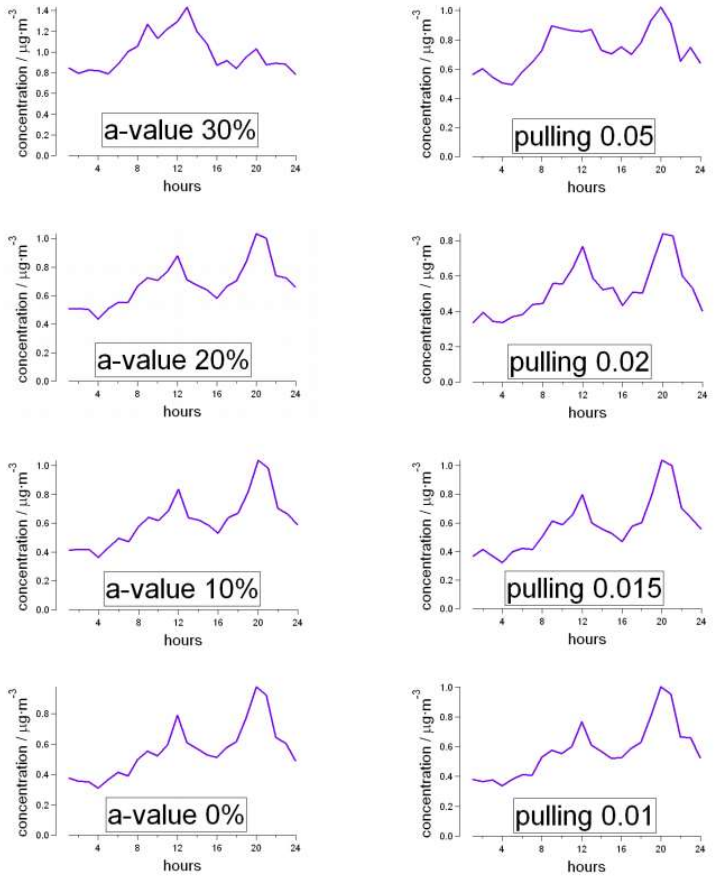
Canonaco et al. 2013

Crippa et al. 2014

PMF results – goals

- **PMF results must contain**
  - factors that are environmentally reasonable





Canonaco et al. 2013

PAUL SCHERRER INSTITUT

PSI, 01.10.2018

Tutorial – SoFi

Learning goal
- prepare the data for a PMF run in SoFi
- import and look at various PMF results

PAUL SCHERRER INSTITUT

PSI, 01.10.2018

SoFi Pro

Key words:
Automated criteria-based selection, rolling approach, rotational & statistical uncertainty, C-value approach, additional variable / time separation

Pro package of SoFi

Main Features of a SoFi Pro license

- (**support**) our full technical and reasonable scientific support with regular **updates** and **upgrades** to future IGOR versions
- (**resampling strategy**) bootstrap application on PMF input for the assessment of the statistical error and subsequent analysis using the dynamic criteria-based feature
- (**criteria-based feature**) Inspection and selection of PMF runs based on user-defined proxies/tracers
- (**statistics on average**) average over several PMF runs and visual inspection of the solution
- (**rolling technique**) user-based PMF sub-window moves over the entire PMF input allowing to model changing factor profiles. Especially relevant for long-term SA studies with profiles known to vary over time
- (**relative error scaling**) manual and automated application of the C-value, when combining data from two and more instruments and subsequent graphical support when exploring these solutions
- (**additional averaging**) hourly, daily, weekly, monthly and yearly average in SoFi for externals, PMF input and solution
- (**classes**) variables and/or time points can be classified for further analysis, e.g. PMF with data from various stations at the same time or various size-fractions
- (**saving&loading utilities**) saving and loading user-specific PMF input and constraints or user-defined criteria, graphical support for the quantification of the PMF error, statistics on the a value constraints.
- and many more...Consult the SoFi manual for more details.

Pro package of SoFi

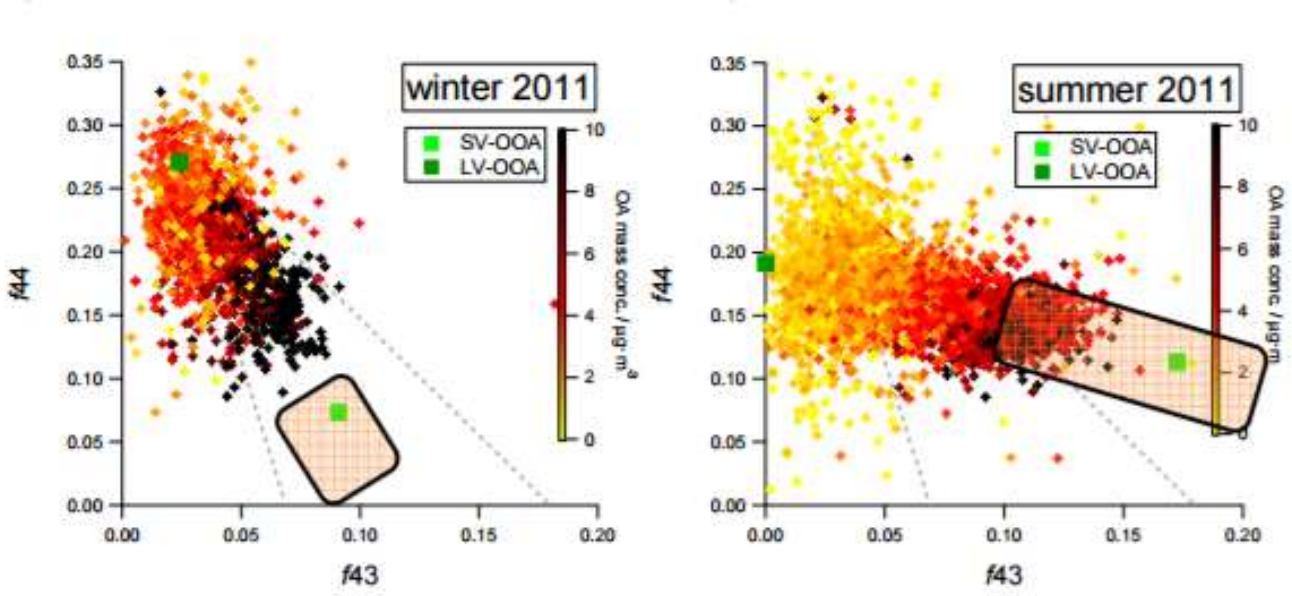
Pricing list

The following table summarizes the costs for a SoFi Pro license. Prices are given **per PC and year** and are reported without VAT. 25 % discount is applied for the purchase of multi-user or multi-year licenses.

	1 PC	multi-user (3 PCs)	multi-year 1 PC (3 years)	multi-user & -year (3 PCs and 3 years)
costs per PC & year in CHF	1000	750	750	500

Variability of factor profiles

- limitation of PMF: factor profile is constant over the PMF run
- Example: Zurich ACSM data 2011/2012

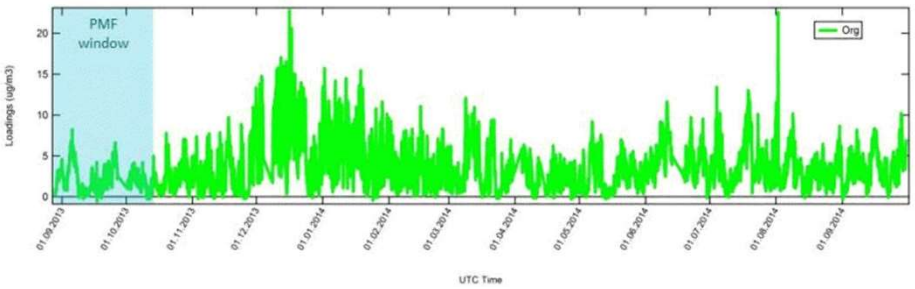


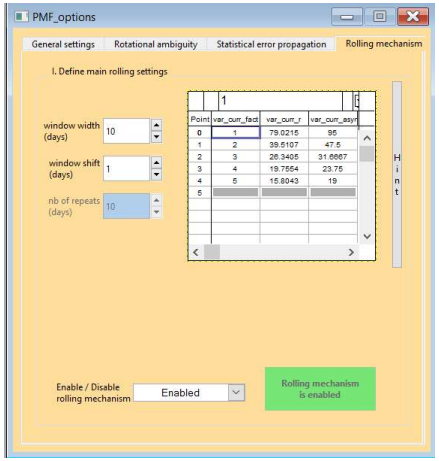
Canonaco et al. 2015

Automatic rolling PMF window

- Rolling PMF algorithm to account for seasonal and/or meteorological variations in OA sources

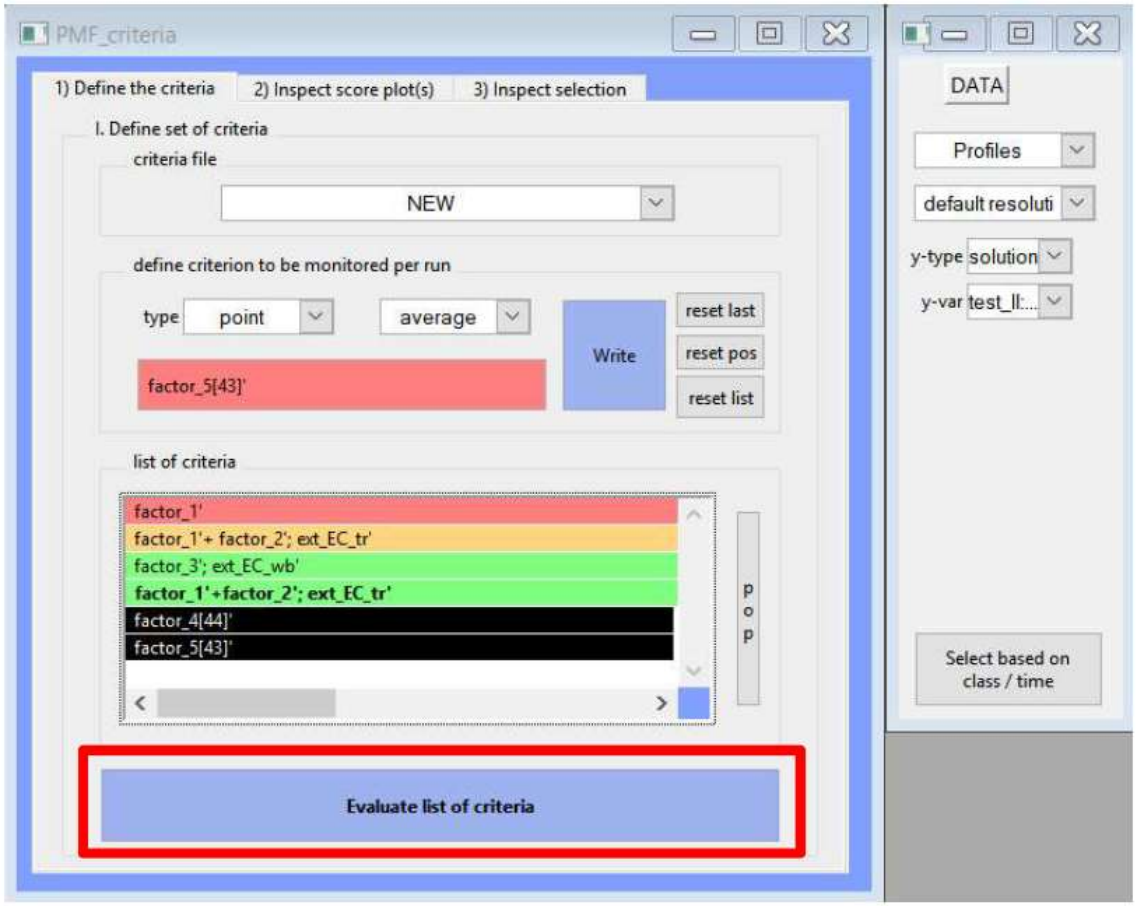
- PMF algorithm is run repeatedly on a short subset of data for a defined period
  - Assumption of constant aerosol sources during that time
  - after every shift the PMF runs are reinitialized (seed, *a*-value, fpeak, bootstrap, etc.)
- PMF window is subsequently shifted by defined period
- Thousands of runs that are sorted using the automated criteria-based selection
  - Goodness of PMF solution is estimated by applying user-defined criteria





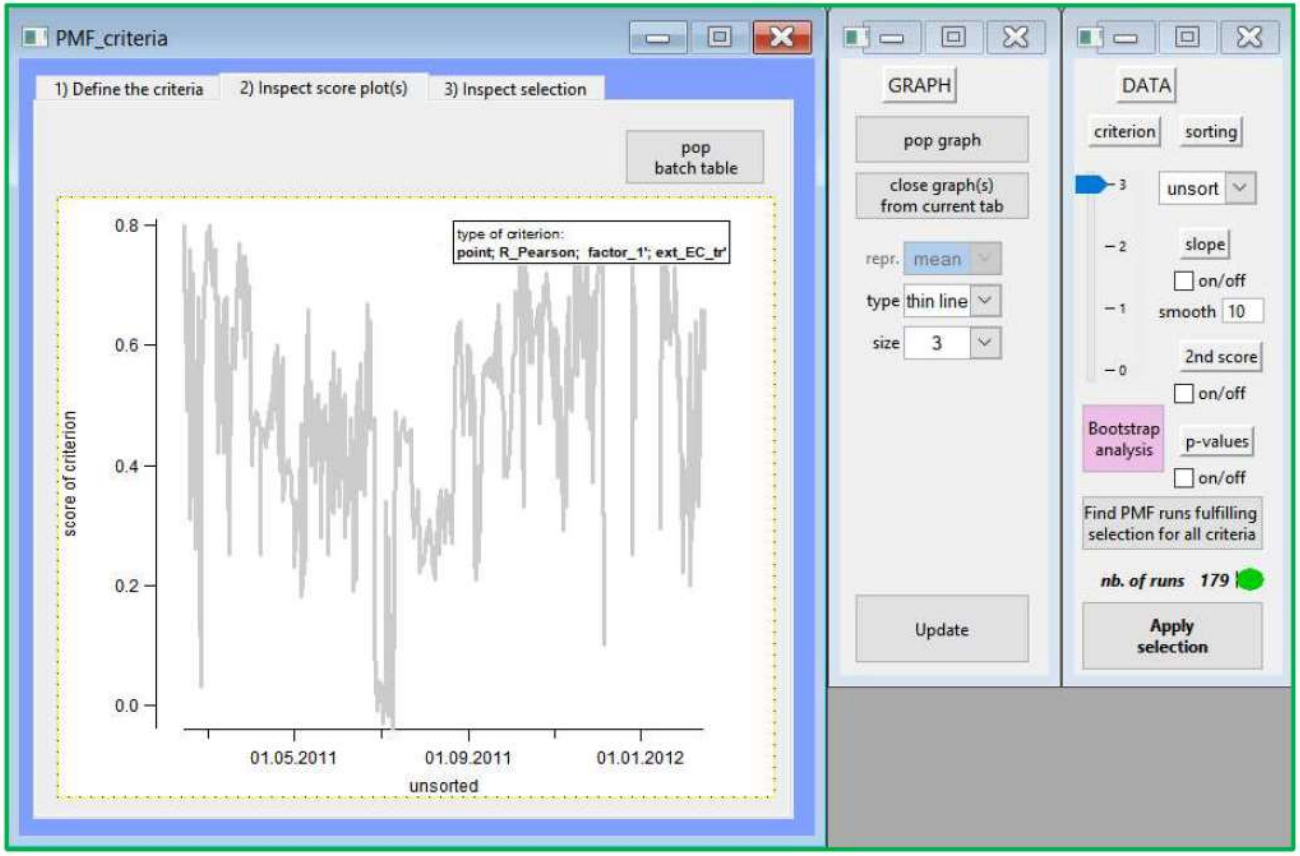
Automated criteria-based selection of PMF-runs

- User defines criteria for the factors, correlation coefficients, contributions, fractions, etc. to be monitored over the PMF runs



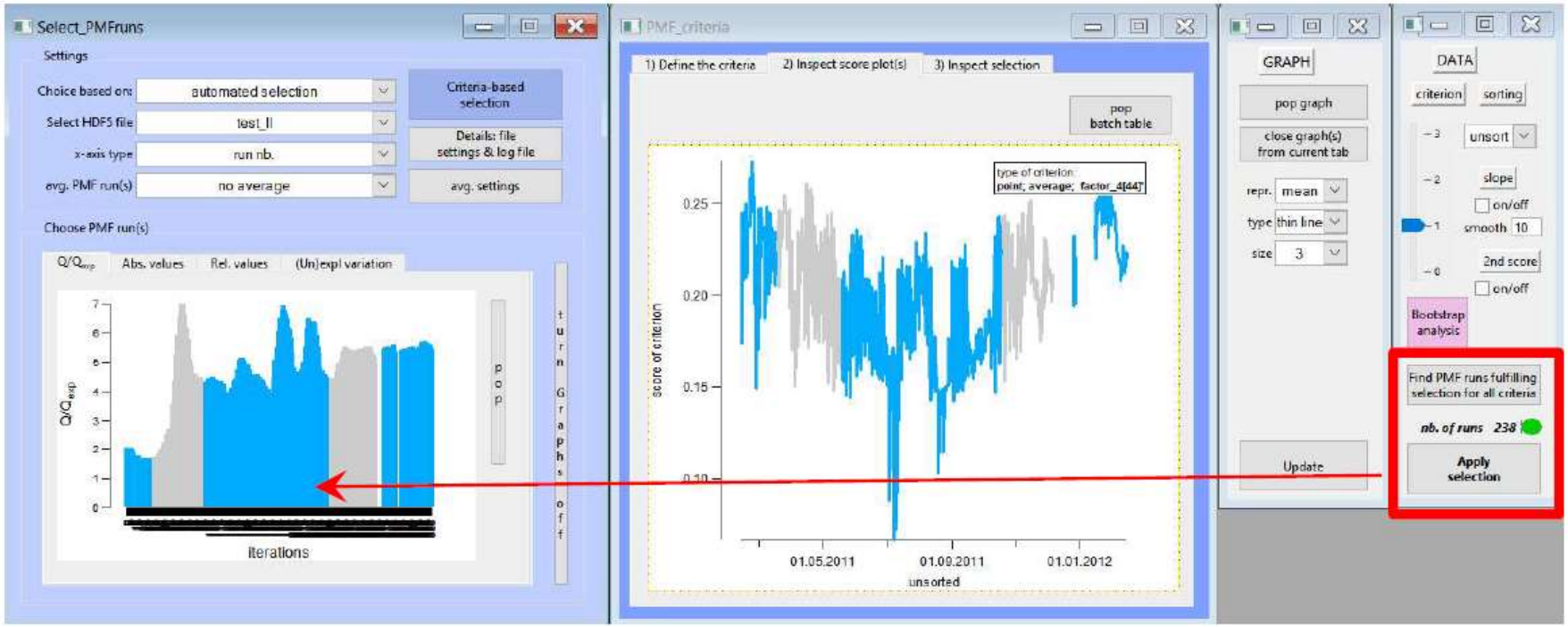
Automated criteria-based selection of PMF-runs

- all PMF runs are temporarily imported in IGOR and the scores of the criteria are evaluated for the PMF runs.



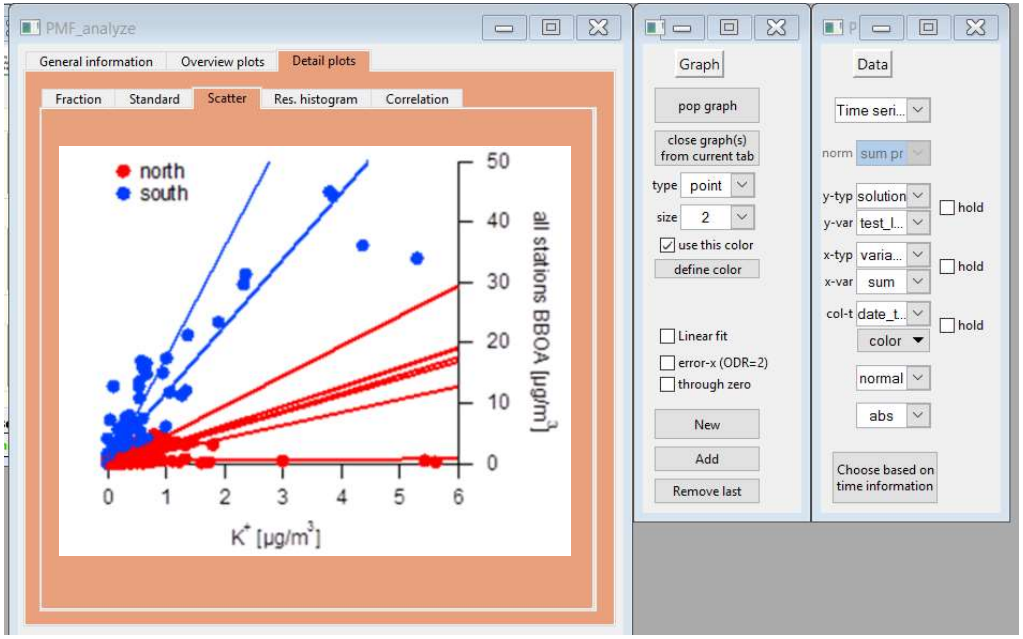
Automated criteria-based selection of PMF-runs

- PMF runs are selected based on the scores for every criterion
- Overlapping PMF runs are selected and can be further investigated as single runs or on average.



Include additional information

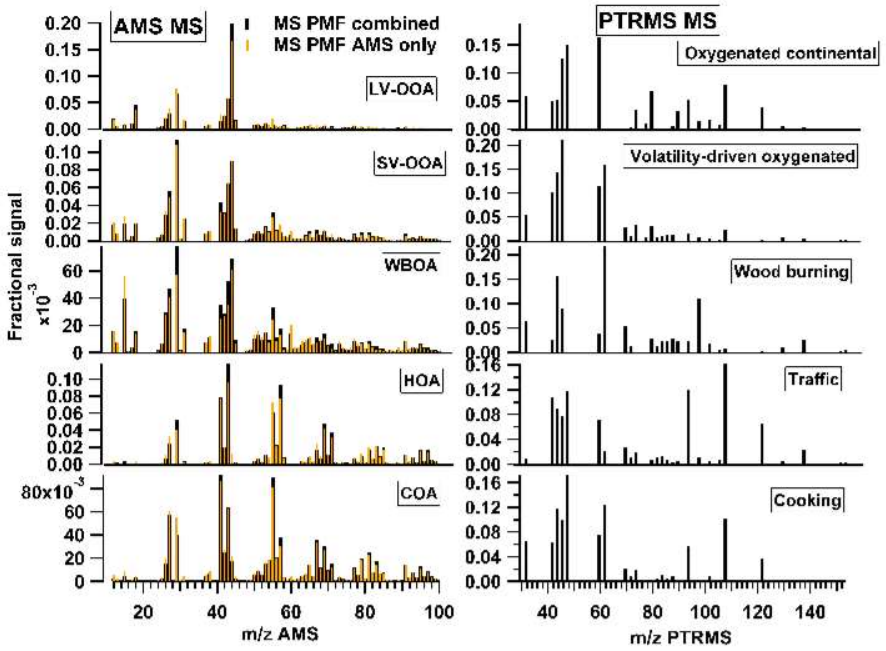
- Inspect PMF result based on additional information over variables / time
  - e.g. PMF run over data from two groups of data (north, south of the alps)

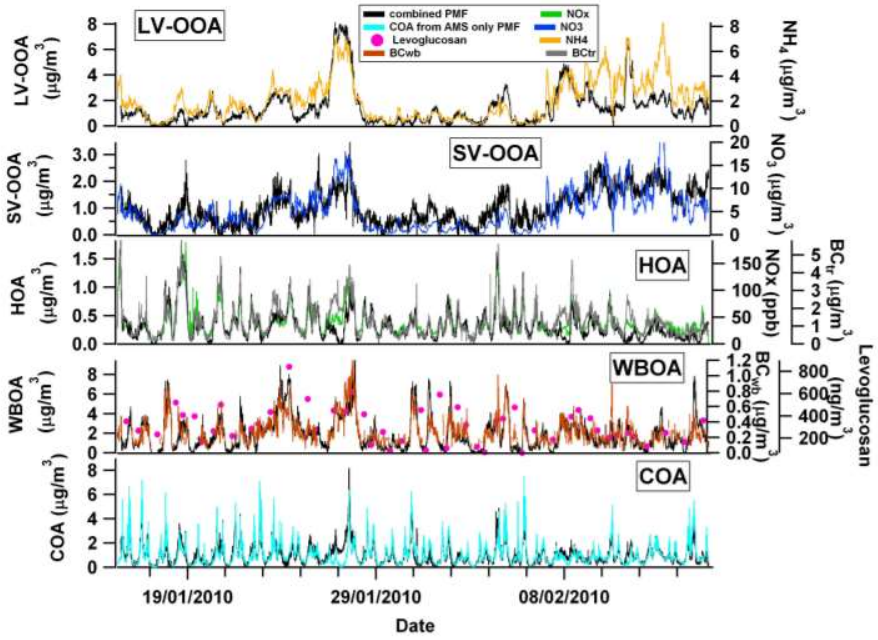


Daellenbach et al., Atmos. Chem. Phys., 17, 13265–13282, 2017

C-value approach

- automated weight of errors, e.g. when combining AMS with PTR-MS data





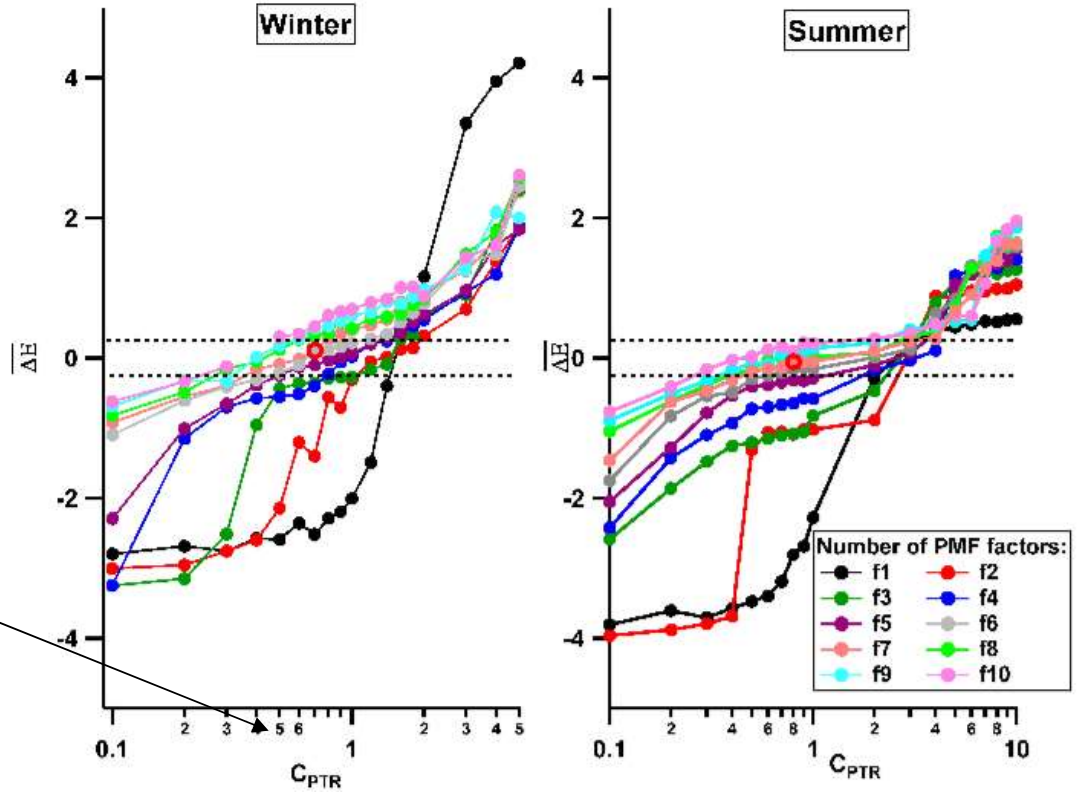
Crippa et al. 2013

C-value approach

- automated weight of errors, e.g. when combining AMS with PTR-MS data

$$\overline{\Delta E} = \overline{\left(\frac{|e_{ij}|}{s_{ij}}\right)}_{\text{AMS}} - \overline{\left(\frac{|e_{ij}|}{s_{ij}}\right)}_{\text{PTRMS}}$$

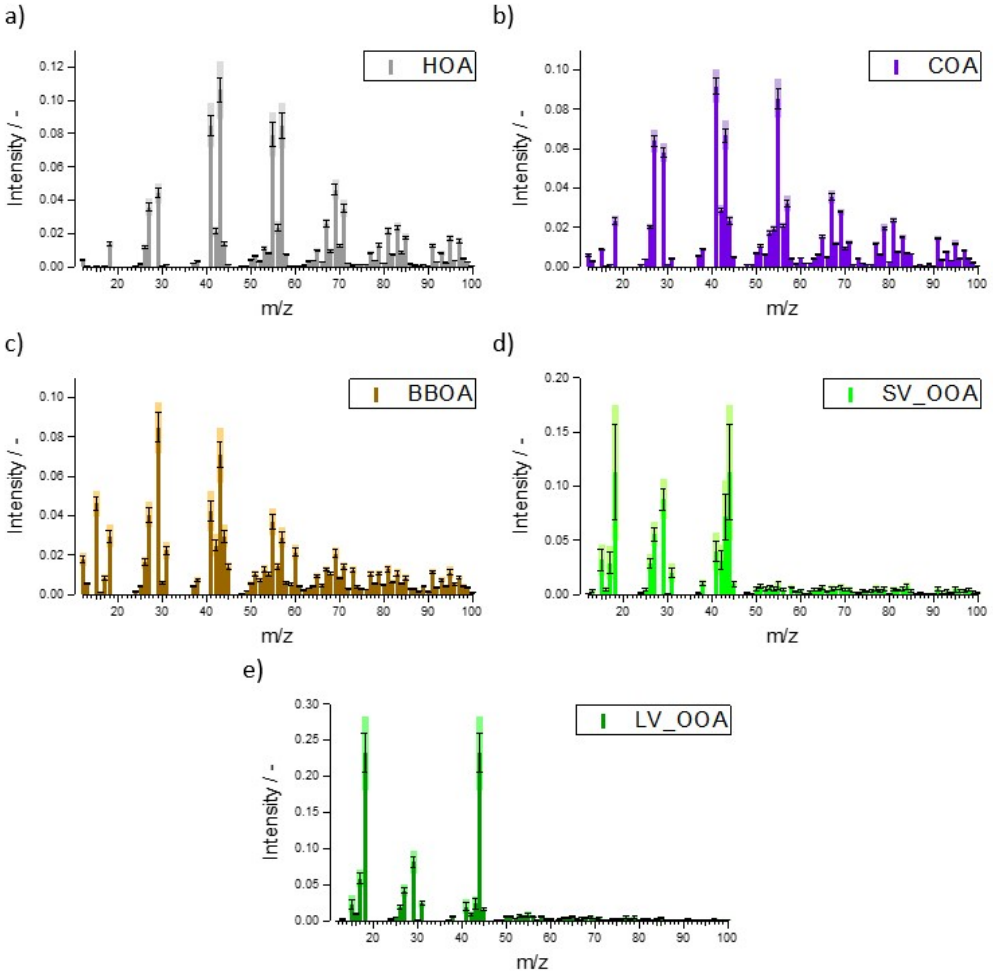
$$s_{ij,\text{new}} = \frac{s_{ij}}{C_{\text{PTR}}}.$$



Crippa et al. 2013

PMF – uncertainties of PMF solutions

- **PMF solution**
  - PMF solution contains all PMF results that are environmentally reasonable (!not only one PMF run!)
  - rotational uncertainty (amount of rotational ambiguity) is assessed using e.g. the a value technique
  - statistical uncertainty, for ambient ACSM data mainly the daily variation of the sources, is assessed using the resampling strategy «bootstrap»

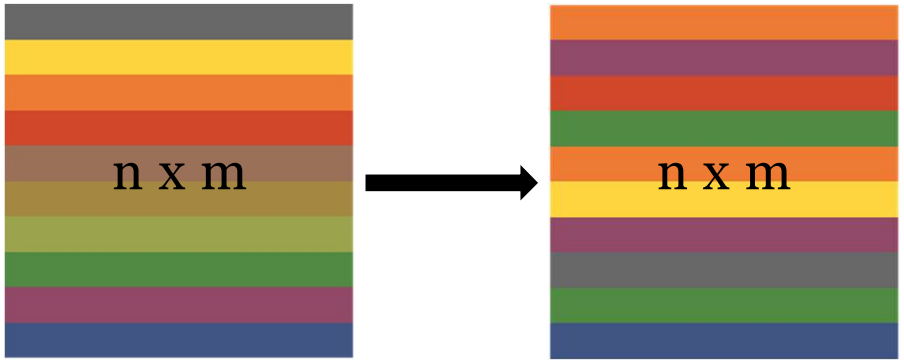


Canonaco et al. in prep

PMF – bootstrap BS

- Uncertainty estimation / variability in PMF solution
  - Randomly selecting rows or blocks of consecutive samples
  - Create new data set with dimensions of the original data set
  - PMF runs on resampled data set

- Does not properly account for rotational ambiguity



Efron, 1979 (BS technique) and Ulbrich et al., 2009 (application on AMS)

PAUL SCHERRER INSTITUT

PSI, 01.10.2018

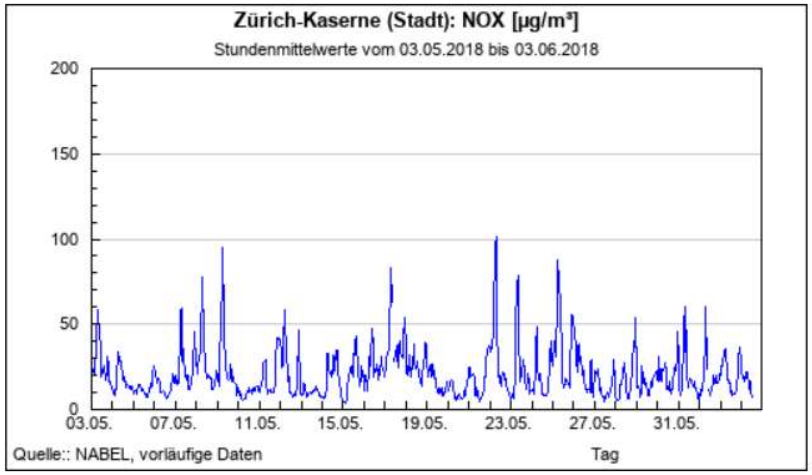
Tutorial – SoFi Pro

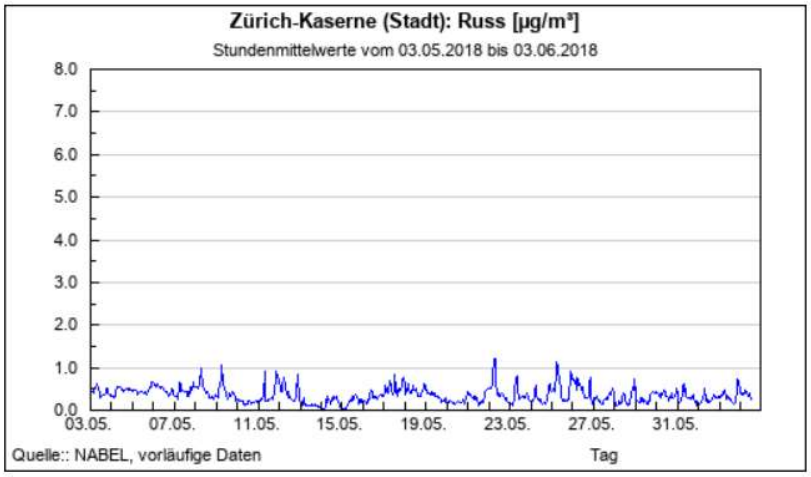
Learning goal
- apply the rolling strategy
- define and use criteria on PMF runs
- inspect averaged PMF solution

Model - Positive Matrix Factorization (PMF)

Data collection

Ambient measurement, e.g. aerosol, gas-phase, liquid data, etc.



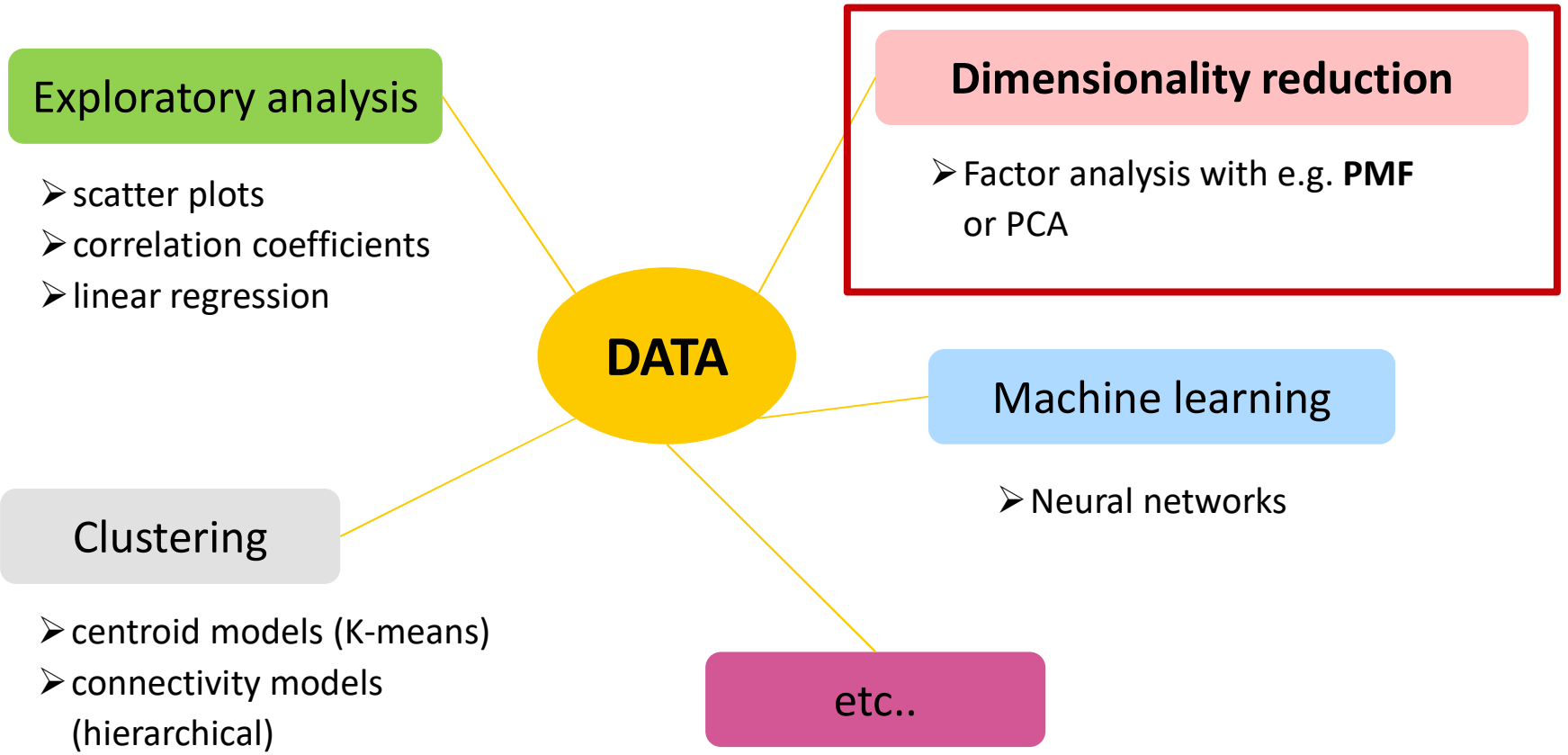


https://www.bafu.admin.ch/bafu/de/home/themen/luft/zustand/daten/datenabfrage-nabel.html (last visited on the 03,06,2018)

Model - Positive Matrix Factorization (PMF)

Data analysis

Several techniques can be applied depending on the type of data



https://en.wikipedia.org/wiki/Data_analysis (last visited on the 03,06,2018)

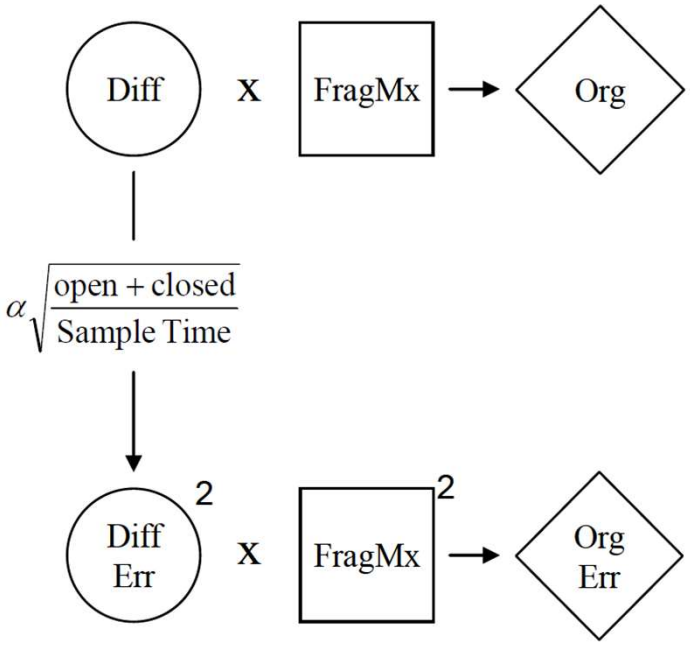
Required information

SoFi will need the following input

- data matrix
- error matrix
- numeric wave that has index values for your species (1, 2, 3, ...)
  - text wave that has the names of your species (short names are better) but can be created by SoFi
- a time wave

PMF – Error matrix

- Standard AMS error calculation



Allan et al., 2003

- Direct calculation in the softwares for AMS and ACSM

PMF – Error matrix

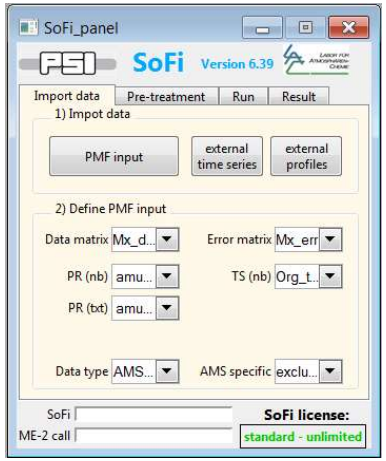
- **PTR Data Set**

$$Error = \frac{1}{I_{H3O+} * sensitivity} * \sqrt{\frac{I_{meas.}}{dwelltime_{meas.}} + \frac{I_{background.}}{dwelltime_{background.}}}$$ Crippa et al. (ACP, 2013)

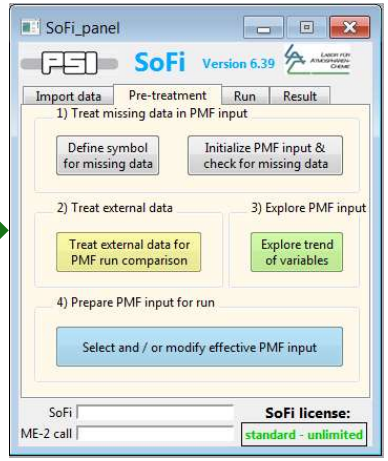
- **GC/MS Data Set (2 methods)**
  - **Polissar et al. (JGR, 1998)**
    - If C < DL replace C with LD/2 and $Error = \frac{5}{6} LD$
    - If C > DL $Error = \frac{1}{3} LD$
  - **Gianini et al. (Atmo Envi., 2012)**
    - $Error = \sqrt{DL^2 + (CV * C)^2 + (a * C)^2}$

where $a = 0.03$ and CV is the Standard Deviation average of replicates

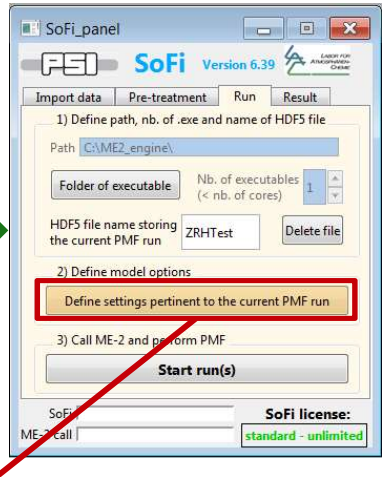
SoFi 6.39 – prepare PMF input



Define input data
Add external data



Pre-treatment of the input data

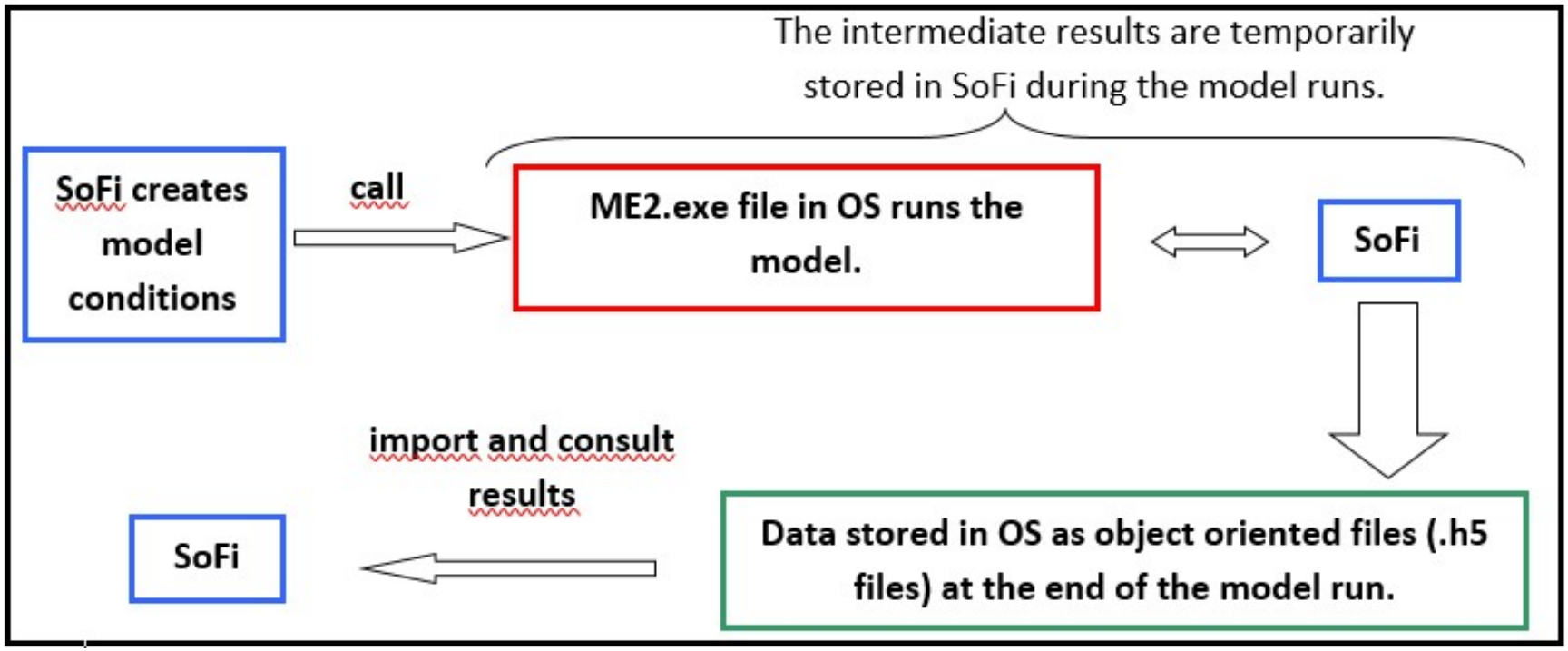


ME2.exe

Define setting for PMF run

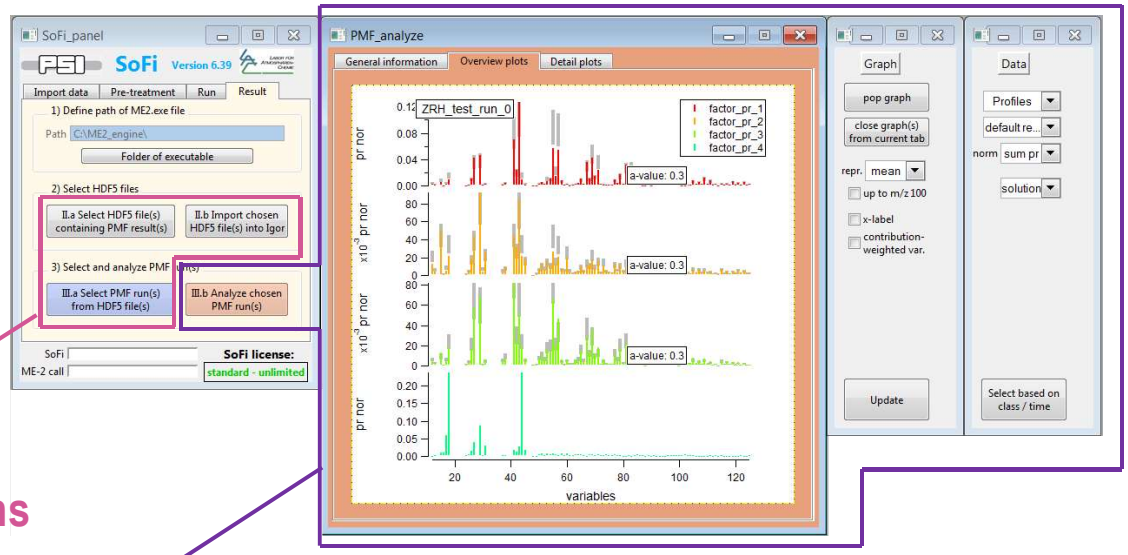
- General (#factors, missing data, etc.)
- Rotational ambiguity (seed, a-value, fpeak, pulling value)
- Statistical error propagation (add-on, released soon)
- Rolling mechanism (add-on, released soon)

SoFi/ME2 interaction



SoFi 6.39 – analyzing PMF results

ME2.exe



Select and import PMF runs for analysis

Tools for analyzing selected PMF runs

- Consult results quickly in preview window
  - General overview (time series, profiles)
  - Detail plots for time series and profiles
    - Fraction plots
    - Scatter plots
    - Correlation
    - (HR-family for HR-data)