Datalystica Ltd. at Park InnovAARE in Villigen, Switzerland

Datalystica Ltd.

The Source Finder (SoFi)

F. Canonaco, A. Tobler, C. Bozzetti, Y. Sosedova, A.S.H. Prévôt and many more

- SoFi is a user-friendly and powerful software package for the efficient and in-depth source apportionment analysisapplying the positive matrix factorizatoin (PMF) algorithm

- Datalystica Ltd. is the official distributor for the source finder software (SoFi) and of the multilinear engine (ME2)
- We also offer source apportionment service

- Almost **200 SoFi Pro licenses** issued in total and SoFi comunity counts more than **300 registered SoFi members over all continents**.

DR. FRANCESCO CANONACO
CEO, board member & Software engineer

DR. CARLO BOZZETTI
Board member & Software engineer

DR. ANNA TOBLER
Software engineer & Support

DR. YULIA SOSEDOVA
Data analysis

Specifications

IGOR-based software

- SoFi 6.E working on IGOR 6.37 (only bug fixing!)
- SoFi 8.0 beta working on IGOR 8.04 (strong development of new features)
- SoFi standard (freely available), SoFi Pro (license-based) and SoFi RT (license-based, will be released soon)
- Licenses require stable communication with the internet, or offline license, if strictly necessary

- **ME-2 solver**
- group license

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

SoFi license system

- Policy using SoFi standard only
  - collaboration for at least 1-2 source apportionment studies (A. Prevot and PSI / Datalystica staff supporting your analysis). This guarantees sufficient knowledge transfer of SoFi and the constraining techniques
  - cite the SoFi paper in AMT (Canonaco et al. 2013)
- SoFi Pro
  - license-based can be purchased from Datalystica Ltd. (visit https://datalystica.com)
  - No collaborative restrictions

Pricing list

The following table summarizes the costs in Euro (€) for a SoFi Pro license. Prices are given **per PC and year** and are without VAT/Sales tax. VAT/Sales Tax may be added to your purchase totals as required by your state and local laws.

25 % discount is applied for the purchase of multi-user or multi-year licenses.

All prices are subject to change without prior notice.

	1 PC	multi-user (≥3 PCs)	multi-year (1 PC and 3 years)	multi-user & -year (≥3 PCs and 3 years)	perpetual ME-2 license
costs per PC & year in Euro (€)	1000	750	750	500	800

Main Features of SoFi Pro

- (**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 averaged solutions
- (**rolling technique**) user-based PMF sub-window moves over the entire PMF input allowing for time-dependent factor profiles. Especially relevant for long-term SA studies, where factor profiles are supposed 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 values
- ...and much more, consult the SoFi manual for more details.

SoFi – overview

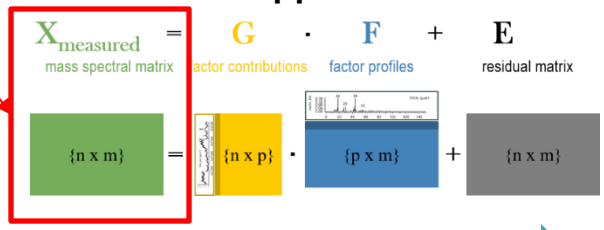
Why SoFi?

Ambient air measurements

Source apportionment

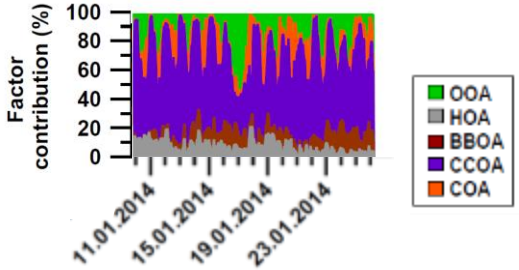
$$X_{measured} = G \cdot F + E$$

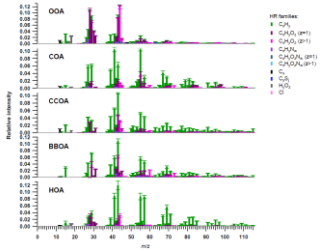
mass spectral matrix — factor contributions — factor profiles — residual matrix



SoFi
(SourceFinder)

Source specific chemical fingerprint and contribution over time



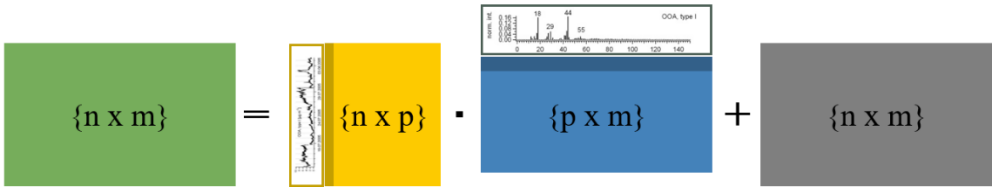


Elser et al. (2016), ACP

SoFi (Source Finder)

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

mass spectral matrix — factor contributions — factor profiles — residual matrix

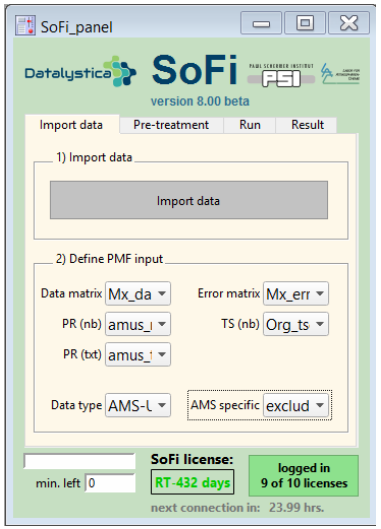


But SoFi is more than just PMF!

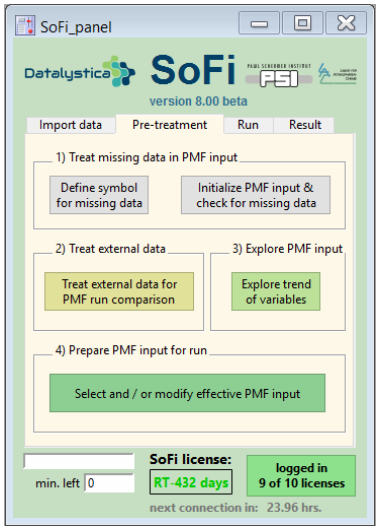
SoFi offers:

- Explorative analysis
  - exploration of a “base case” using a priori information, constrain what is known (e.g. profiles of POA) to possibly extract the more information on complex part (e.g. SOA) (iterative way)
- Assessment of rotational and statistical uncertainties of final solution
  - rotational uncertainties using random *a*-values (with pulling equations, like for displacement in EPA PMF) and statistical uncertainties using bootstrap
- Extremely user-friendly interface to inspect PMF solutions, uncertainties, correlations with external data and much more

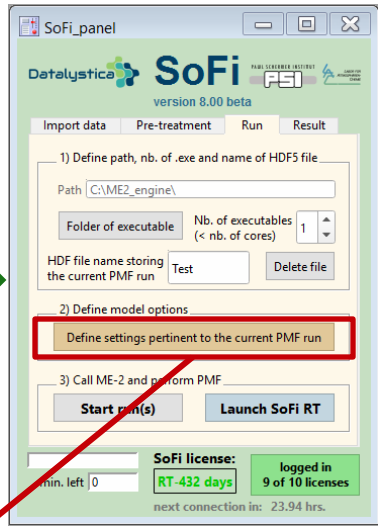
SoFi: prepare and run PMF



Define input data
Add external data



Pre-treatment of the input data



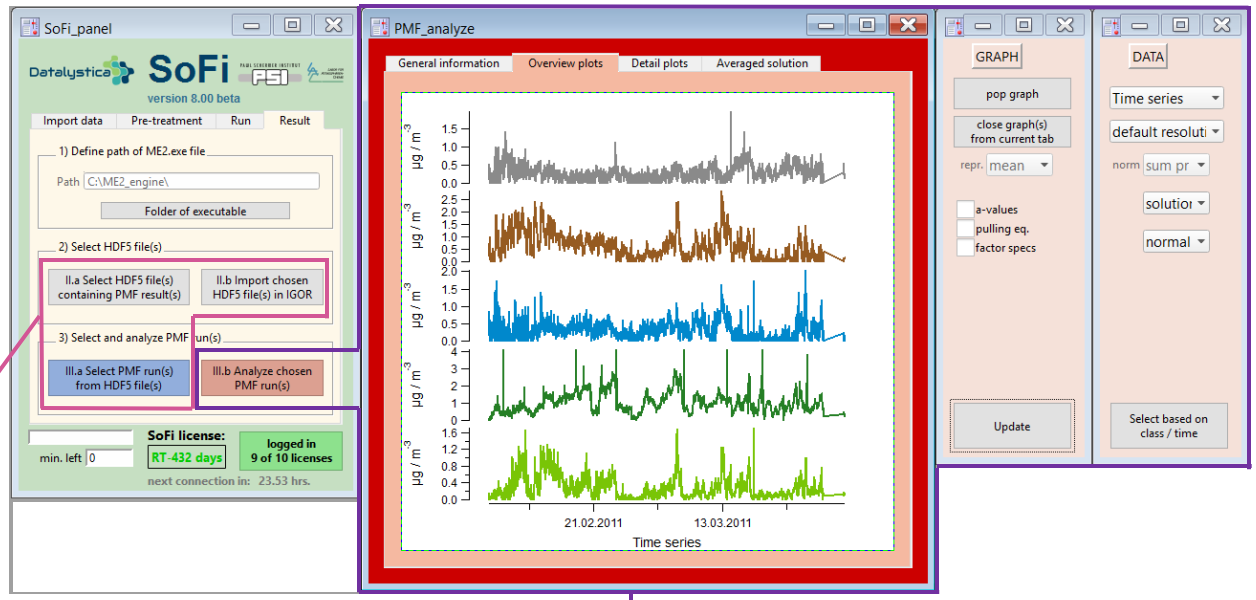
ME2.exe

Define settings for PMF run

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

SoFi: analyze PMF results

Result stored in HDF file(s)



Select and import PMF runs for analysis

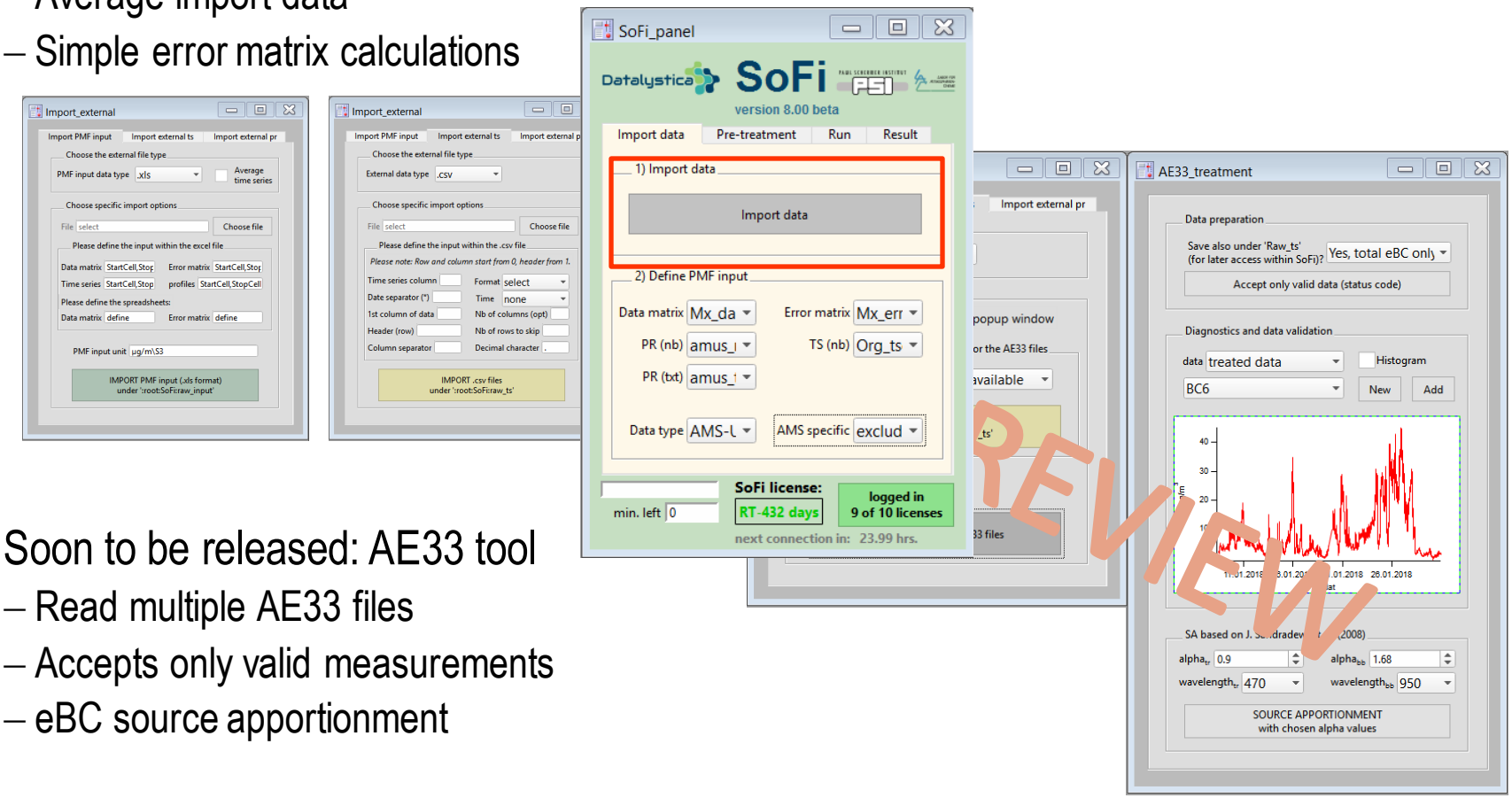
Tools for analyzing selected PMF runs

- 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)
    - etc.

Main features of SoFi

Import PMF input

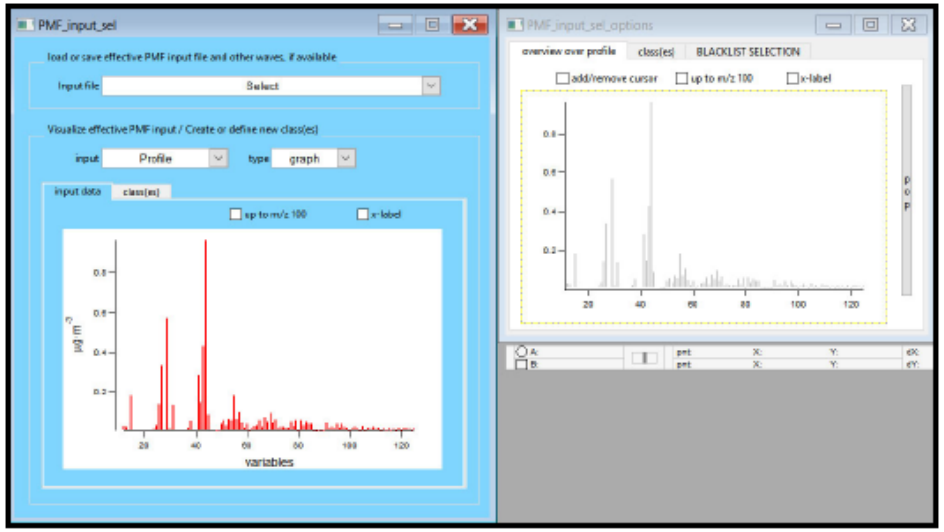
- Easily import data for your PMF input as well as external ts & pr from the most common file types (.itx, .xlsx, .csv, .dat, .txt)
  - Average import data
  - Simple error matrix calculations

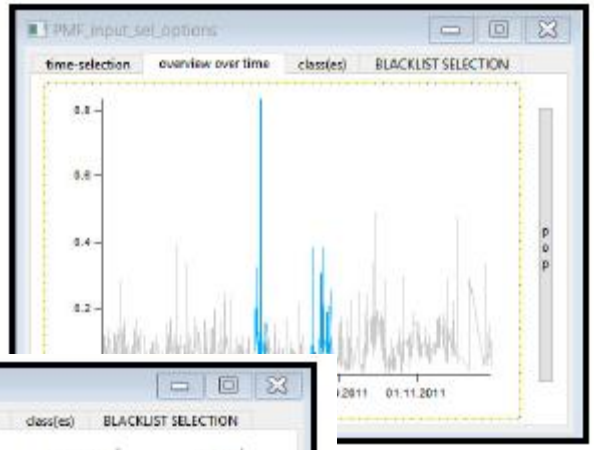


- Soon to be released: AE33 tool
  - Read multiple AE33 files
  - Accepts only valid measurements
  - eBC source apportionment

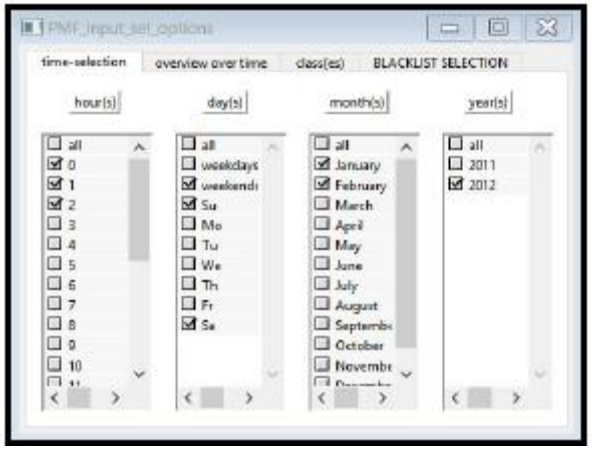
Modify PMF input for further analysis

- Default case considers the full PMF input





- Generate sub selection within seconds for further in-depth analysis

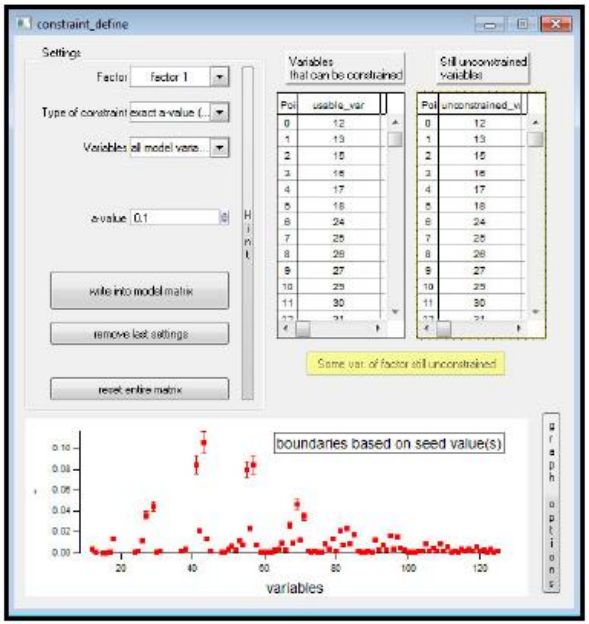


Full and easy control over sofisticated model parameters

- a value technique (random and systematic) with up to 5 independent dimensions for sensitivity analyses

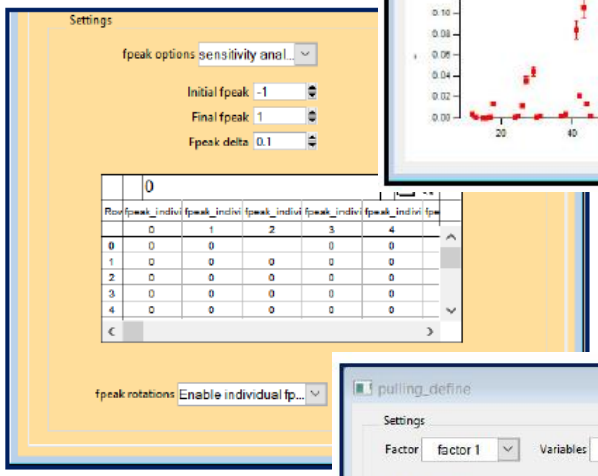
$$f_{j,\,solution} = f_j \pm a \cdot f_j$$

$$g_{i,\,solution} = g_i \pm a \cdot g_i$$

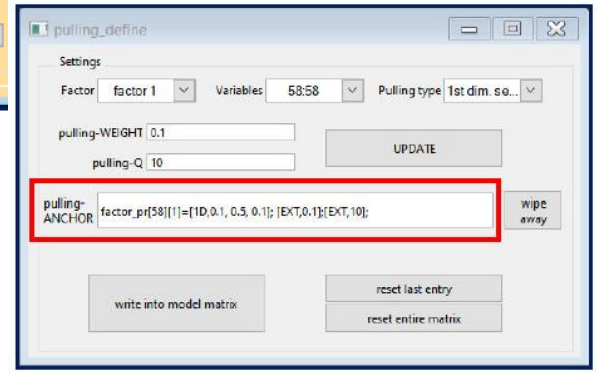


- Global and individual fpeak control (random and systematic) analysis

$$\bar{G} = GT \text{ and } \bar{F} = T^{-1}F$$

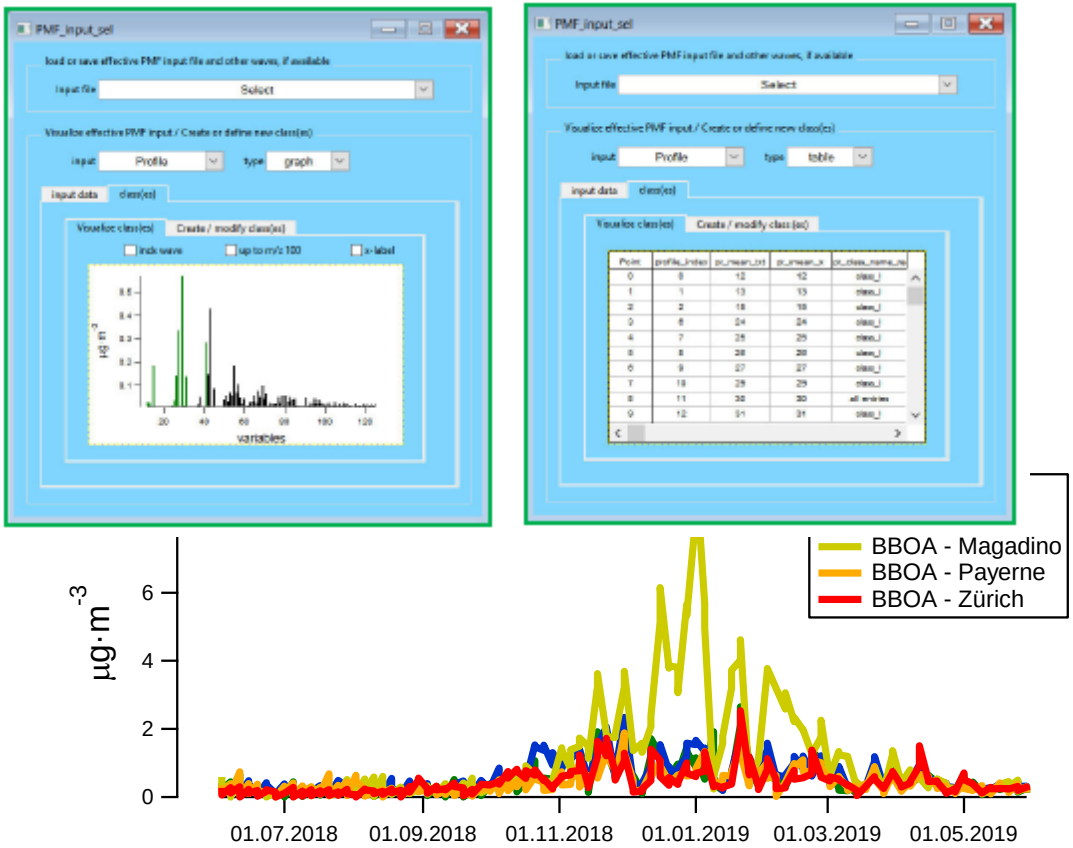


- Pulling equations (random and systematic) with up to 5 independent dimensions for sensitivity analyses for the various pulling parameters

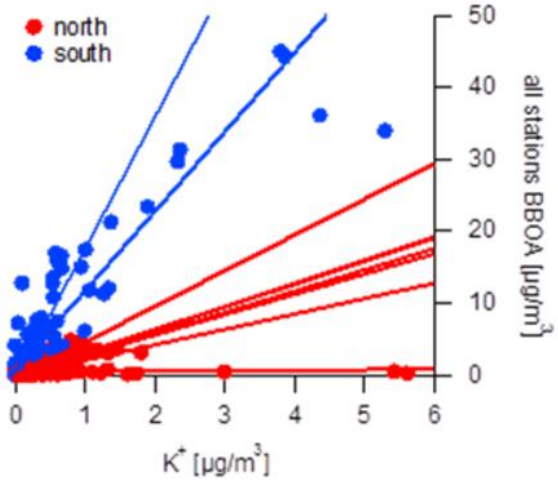


Use of classes as additional metric

- Inspect PMF result based on additional information over variables / time
  - e.g. PMF run over data from different stations, size fractions etc. but inspect them separately



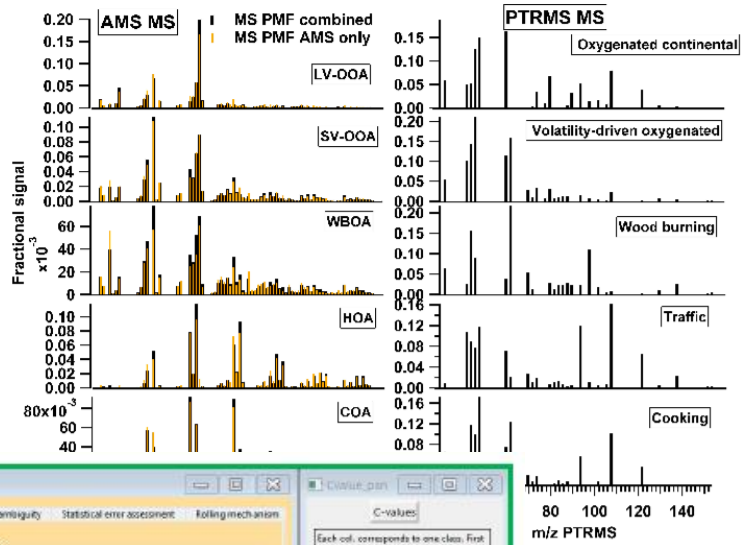
Ortiz et al., in prep.

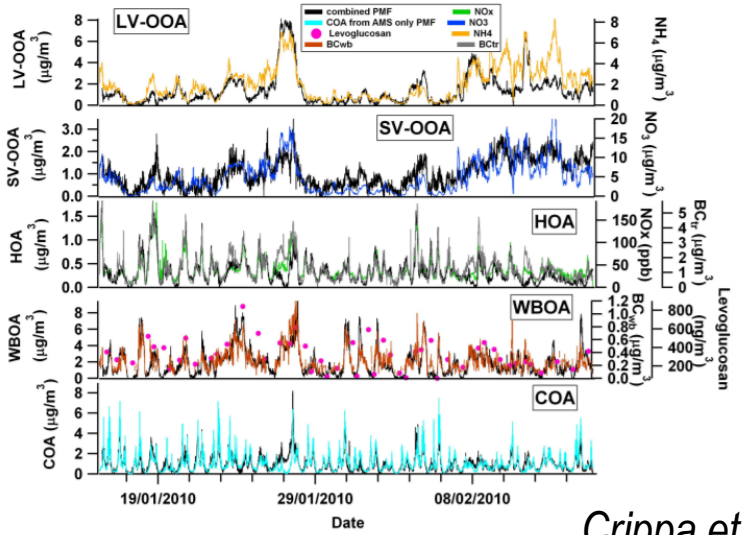


Daellenbach et al., Atmos. Chem. Phys., 17, 13265–13282, 2017

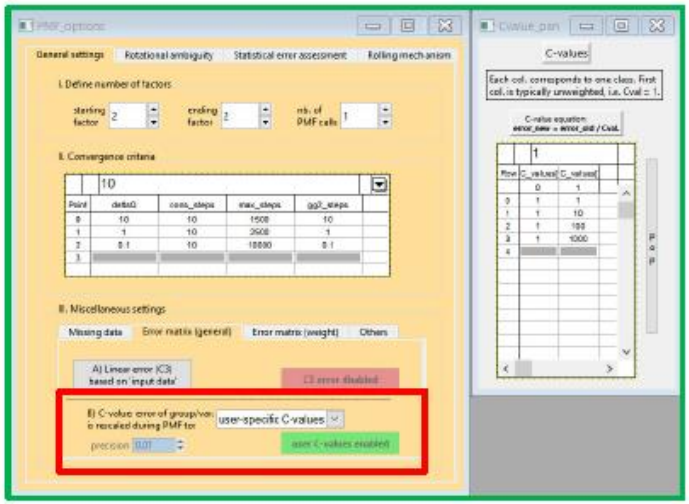
C-value approach

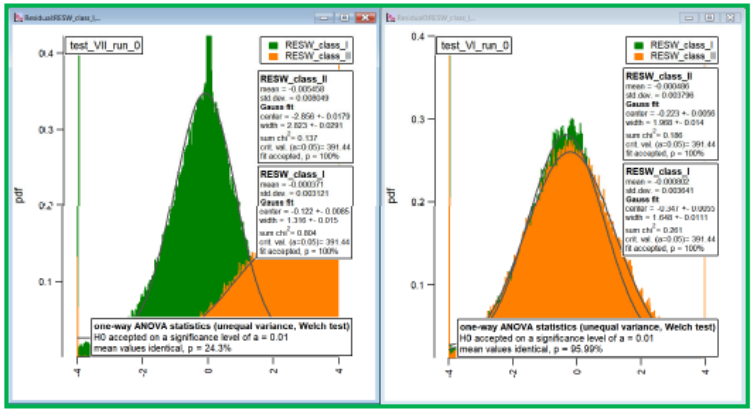
- Relative weight of errors, e.g., when combining AMS with PTR-MS data





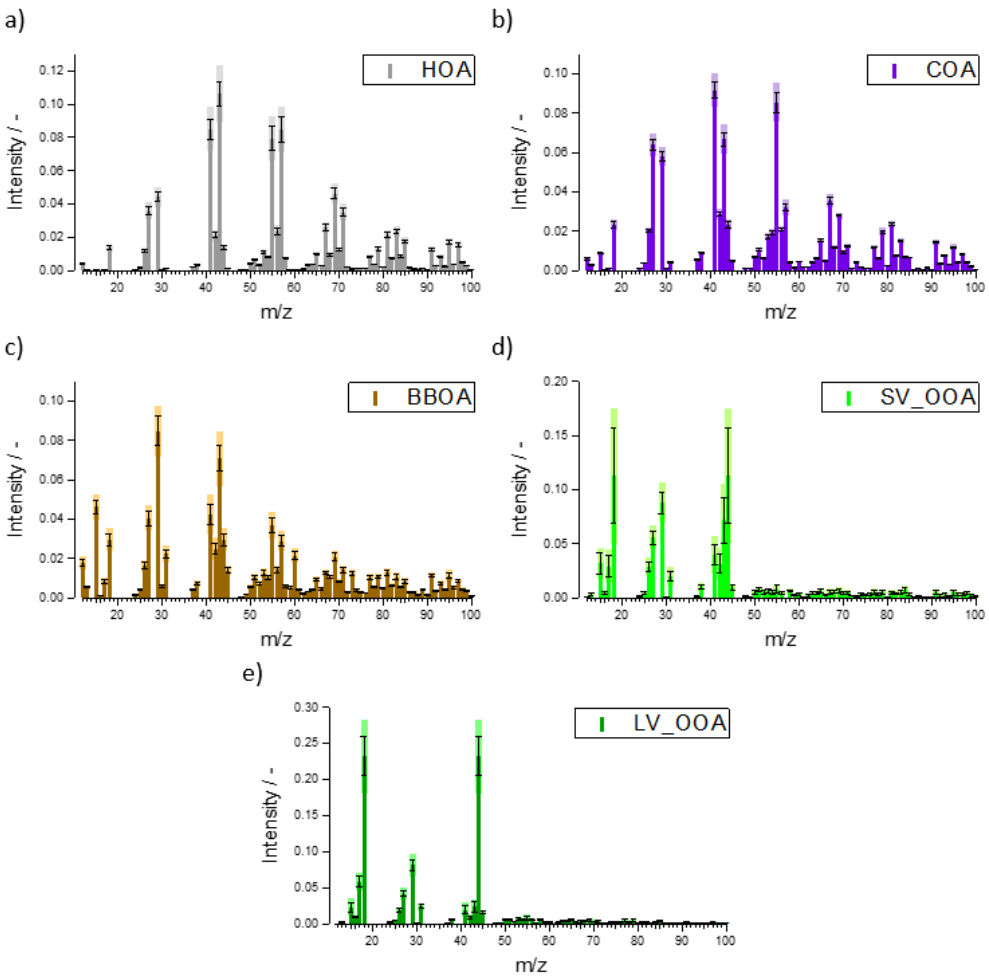
Crippa et al. 2013





Uncertainties of PMF solutions

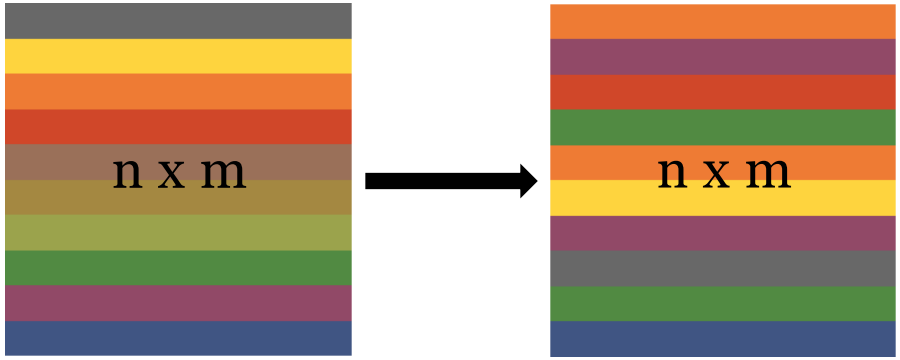
- **PMF solution**
  - PMF solution contains all PMF results that are environmentally reasonable (!not only one PMF run!)
  - rotational uncertainty (amount of rotational ambiguity) is assessed using e.g., random a value or pulling techniques
  - statistical uncertainty, for ambient ACSM data mainly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



Canonaco et al., AMT, 2021 (in a few days)

PMF uncertainty estimation – bootstrap

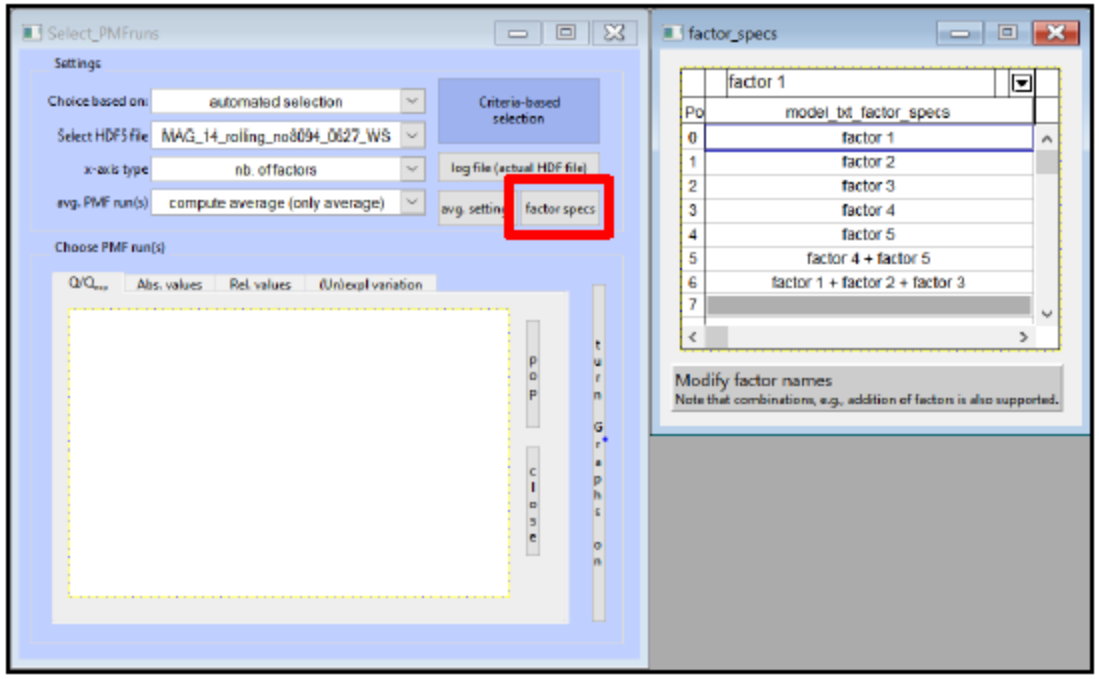
- Uncertainty estimation / variability in PMF solution
  - Randomly selecting rows or blocks of consecutive samples
  - Create new data set with dimensions of the original data set
  - PMF runs on resampled data set



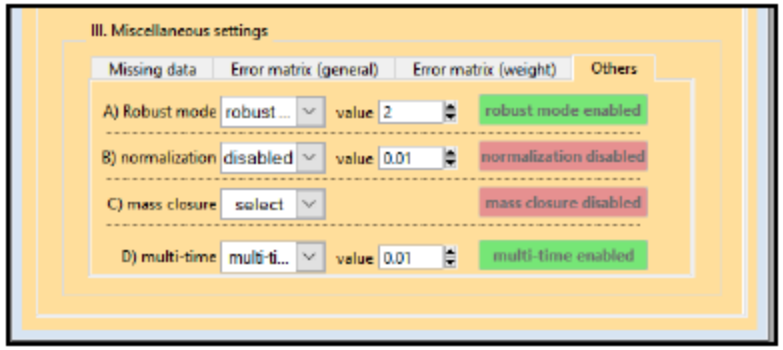
Efron, 1979 (BS technique) and Ulbrich et al., 2009 (application on AMS)

Combining PMF results together

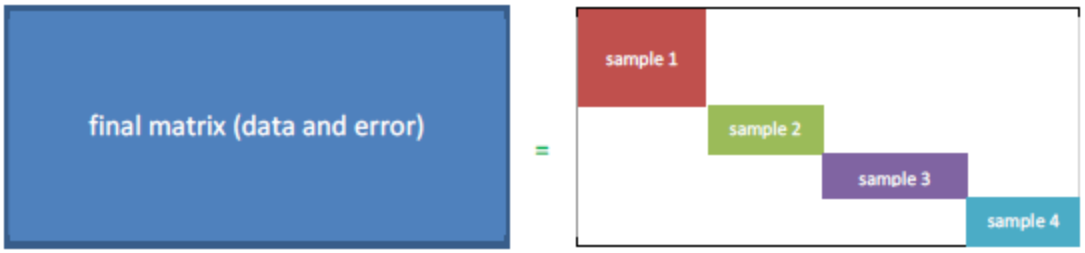
- Combining two and more PMF results from various HDF files can be easily achieved in SoFi by specifying names of the factors
- Also, combinations of factors can be easily represented in the result panel using the same name wave



Other miscellaneous settings, e.g., multi-time



- Combining data with different time resolution can be directly imported into SoFi



- multi-time equations are coded in SoFi

$$x_{sj} = \frac{1}{(t_{s2} - t_{s1} + 1)} \sum_{k=1}^{p} \left(f_{jk} \sum_{i=t_{s1}}^{i=t_{s2}} g_{ik} \eta_j \right) + e_{sj}$$

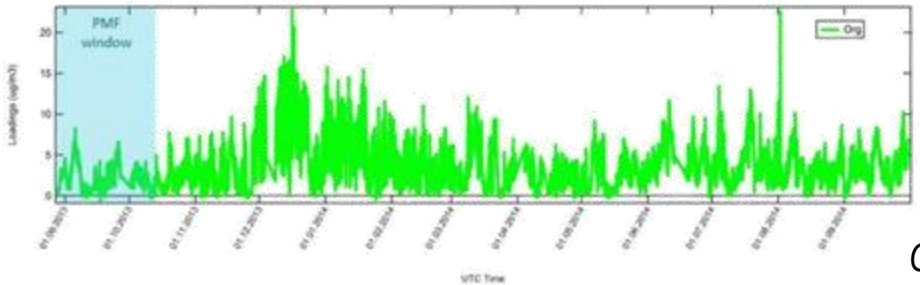
Ogulei et al., 2005

Automatic rolling PMF window

- limitation of PMF: factor profile is constant over the PMF run

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

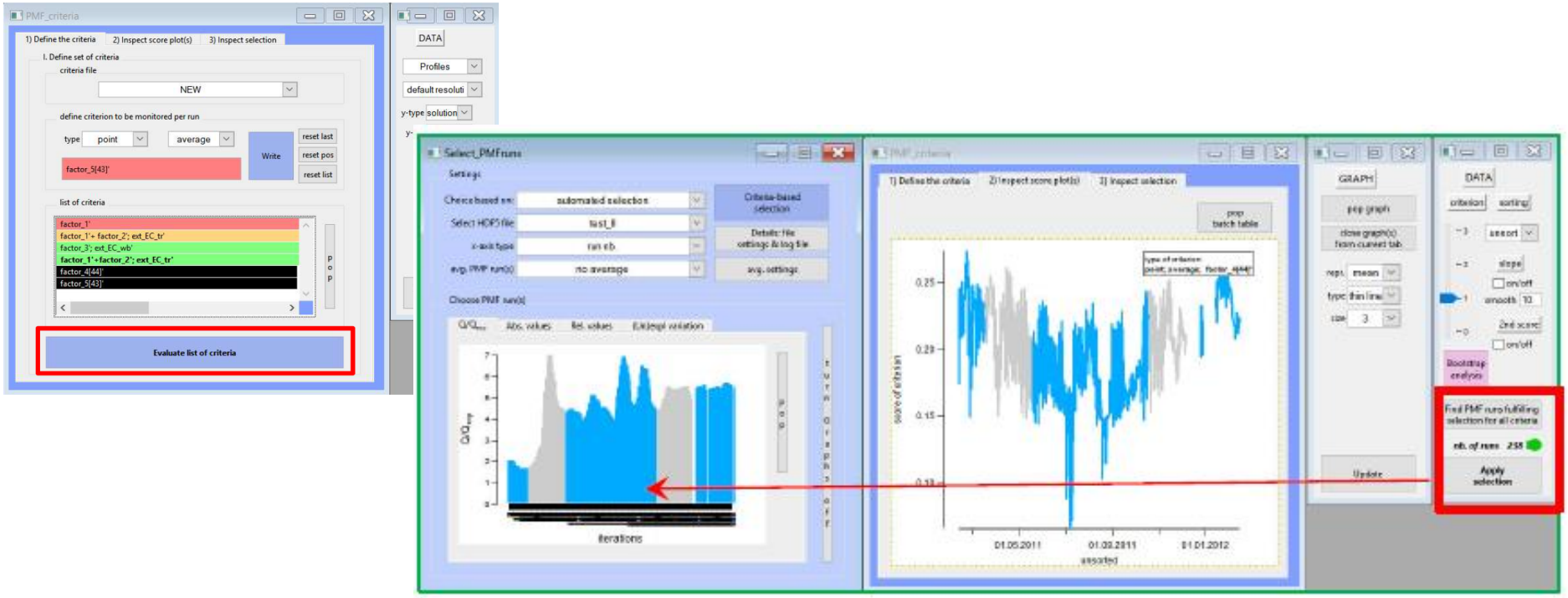
- PMF algorithm is run repeatedly on a short subset of data for a defined period
  - Assumption of constant aerosol sources during that time
  - after every shift the PMF runs are reinitialized (seed, *a*-value, fpeak, bootstrap, etc.)
- PMF window is subsequently shifted by defined period
- Thousands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



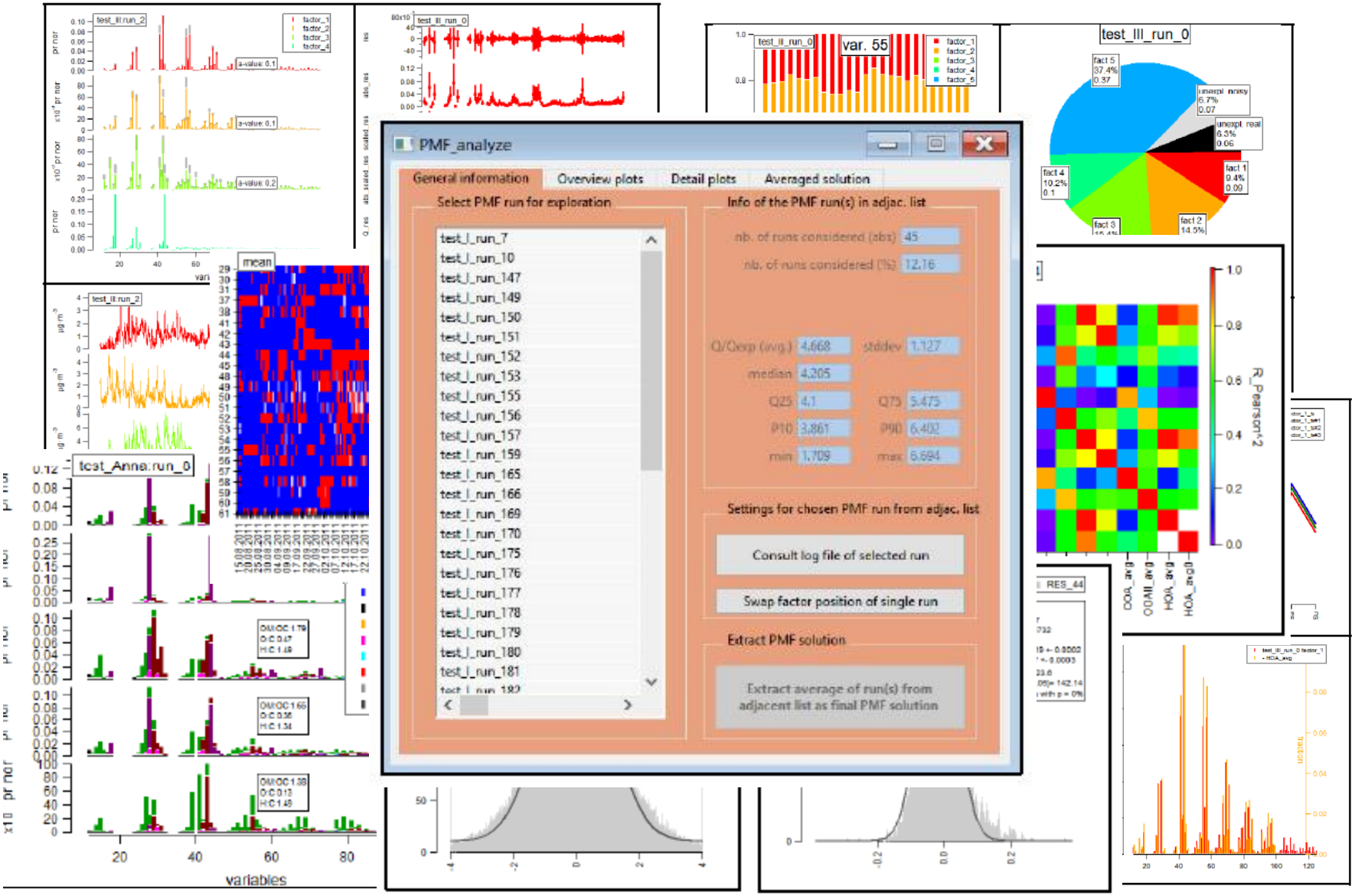
Canonaco et al., AMT, 2021 (next week)

Criteria-based selection of PMF-runs

- Large number of PMF runs can be easily inspected
  - User defines criteria for the factors, correlation coefficients, contributions, fractions, etc. to be monitored over the PMF runs
  - The scores of the criteria are evaluated for the PMF runs
  - PMF runs are selected based on statistical significance tests



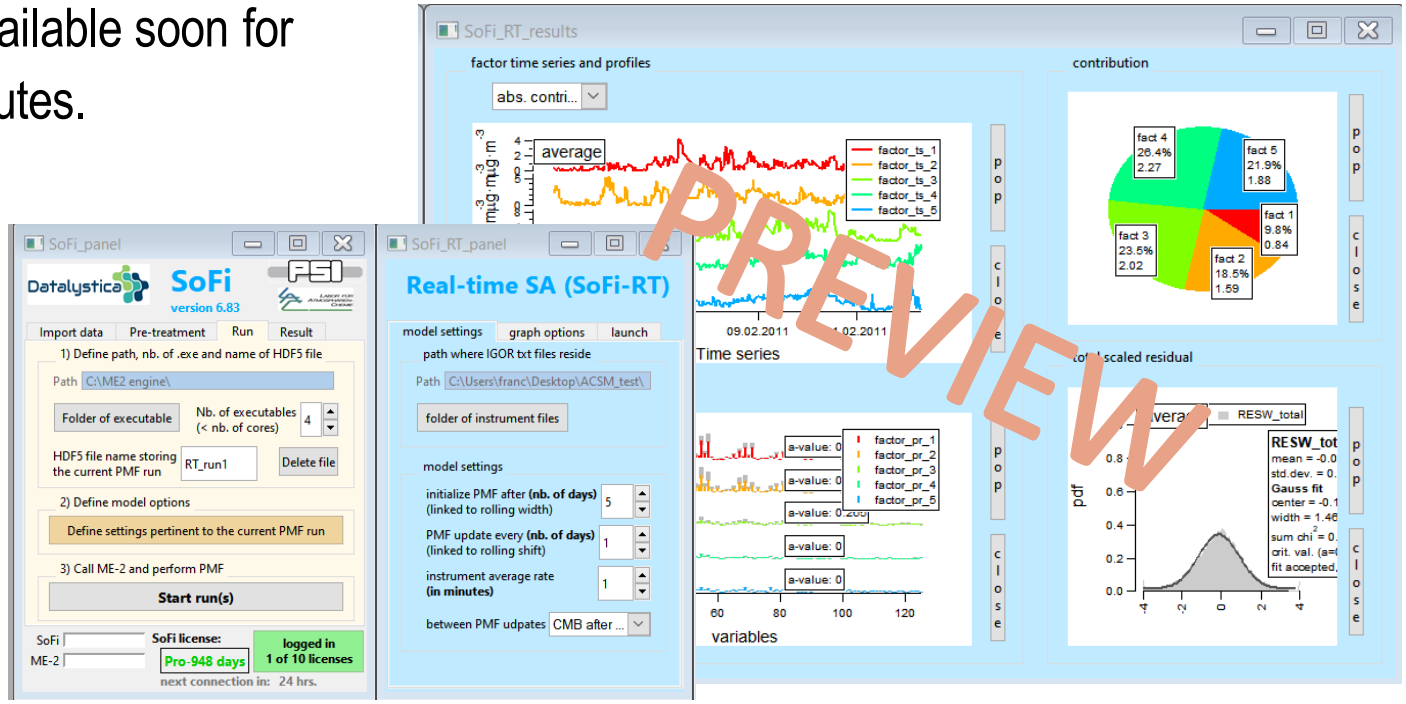
Visualize all necessary plots for SA analysis



Real-time source apportionment

- **Real-time source apportionment with SoFi (SoFi RT):** SoFi conducts an automated rolling PMF on data that is continuously collected by e.g., a Q-ACSM

- under development and testing for Q-ACSM, Xact and Aethalometer
- SoFi RT comes as additional module that requires SoFi and SoFi Pro to run on the PC
- beta version available soon for interested institutes.



Possible SoFi users meeting in April

- Online SoFi users meeting in April 2021, if there is enough interest across the community

- Short Survey: **https://speakneon.com/u6ZcaHxgl**