PAUL SCHERRER INSTITUT

PSI

Datalystica

Paul Scherrer Institute

The Source Finder (SoFi)

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

PAUL SCHERRER INSTITUT

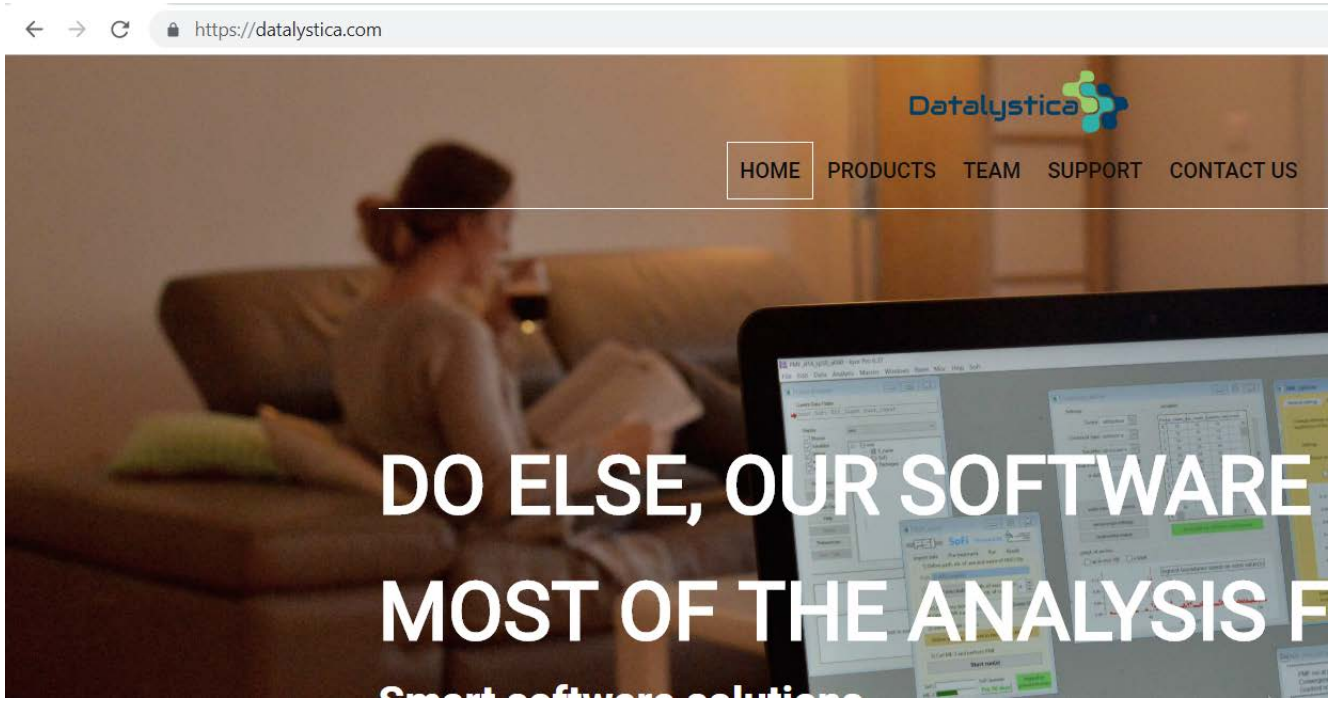
PSI

PSI, 22.08.2019

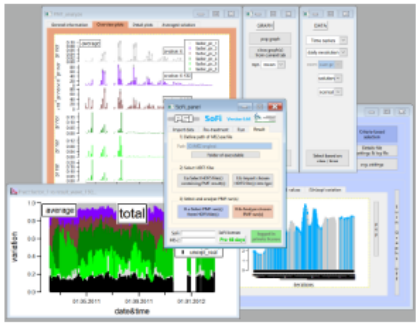
General information

Datalystica

- responsible for the technical support, maintenance and further development of SoFi
- visit https://datalystica.com for more details



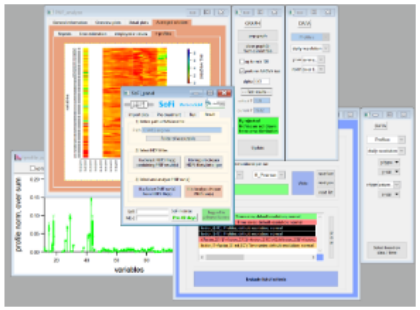
General – Datalystica



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.



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.

Data analysis service

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

Read more information on our data analysis service.

JOIN OUR SOFI MAILING LIST

under:
https://datalystica.com/products
sign up to the mailing list:
info@datalystica.com
to receive news on SoFi, SoFi Pro and SoFi RT

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 can be purchased from the responsible company Datalystica Ltd. (visit https://datalystica.com)
  - No collaborative obligations

Pricing list

The following table summarizes the costs in US Dollars 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)
costs per PC & year in USD	1050	788	788	525

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.

IGOR-based software

- currently working with IGOR 6.37
- will be made compatible with IGOR 8 this winter (➔SoFi 8 family)

Main concept in SoFi

- 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 (soon also with pulling equations, like for displacement in EPA PMF) and statistical uncertainties using bootstrap

Canonaco et al., Atmos Meas. Tech., 6, 3649–3661, 2013
Crippa et al. Atmos. Chem. Phys., 14, 6159-6176.
Canonaco et al., Atmos Meas Tech., in prep.
A. Vlachou *et al*., Atmos. Chem. Phys., 18, 6187-6206, 2018.
P. Rai *et al*., Atmos. Chem. Phys. Discuss., in review, 2019.

PMF – Useful resources

- Francesco Canonaco @ datalystica: francesco.canonaco@datalystica.com
- Carlo Bozzetti @ datalystica: carlo.bozzetti@datalystica.com
- Anna Tobler; Ian Chen @ PSI: anna.tobler@psi.ch ; gang.chen@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 https://datalystica.com)

- visit the SoFi poster presenting the main features of SoFi Pro, P1-073, poster session on Monday between 13.00 and 15.00

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/

PAUL SCHERRER INSTITUT

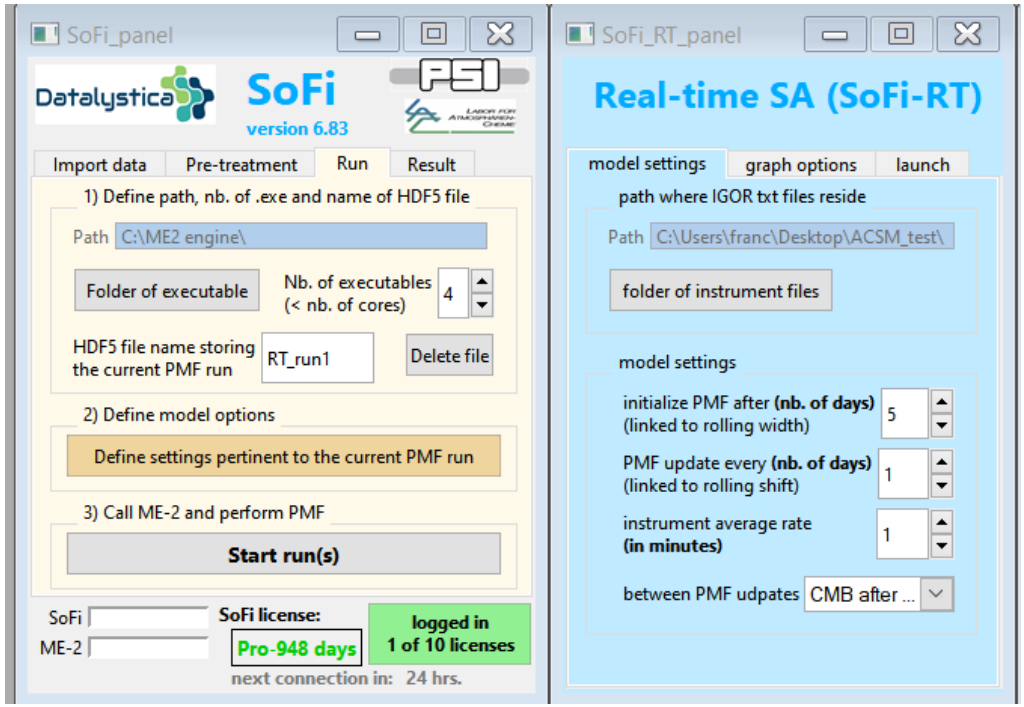
PSI

PSI, 22.08.2019

Real-time Source Finder (SoFi RT)

Real-time SoFi (SoFi RT)

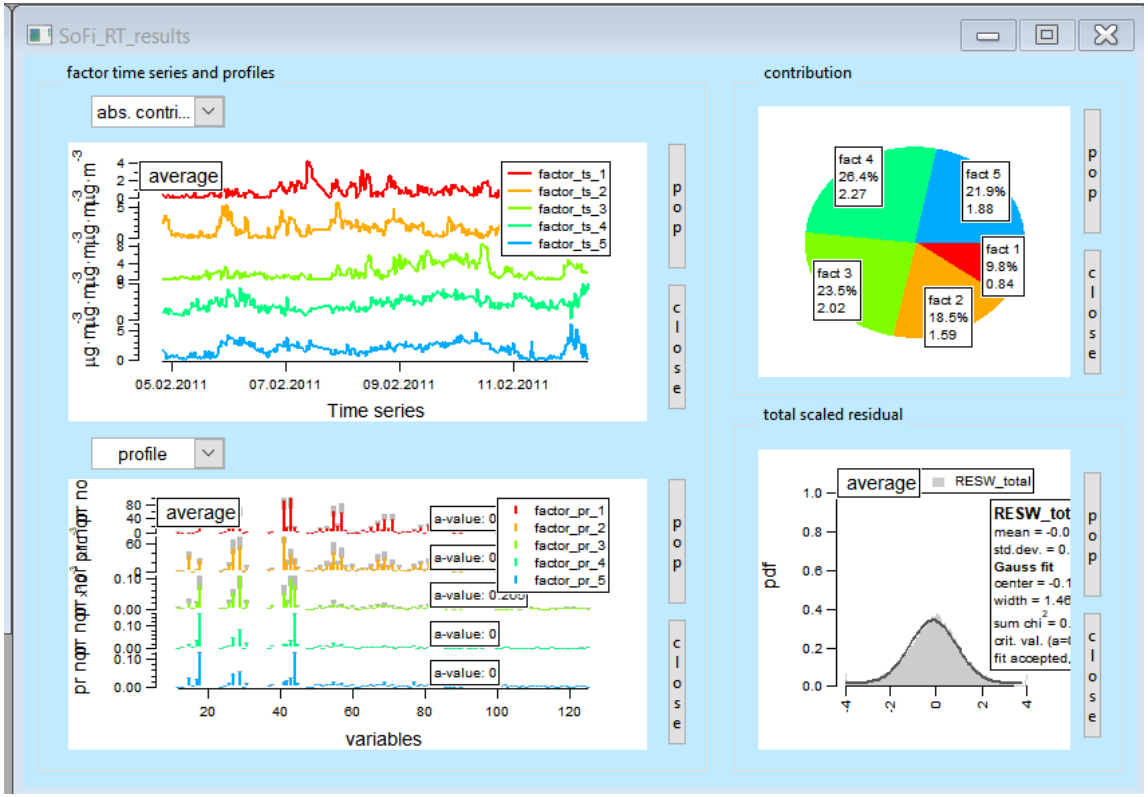
- Automatic real-time source apportionment is currently under development and offline testing.
- SoFi RT comes as additional module that requires SoFi and SoFi Pro to run on the PC
- A beta version will be released very soon.



Real-time SoFi (SoFi RT)

Main idea

- SoFi conducts an automated rolling PMF (e.g. two weeks window with a daily shift) on data that is continuously collected by a Q- or ToF-ACSM
- In these PMF runs the POA factor profiles are constrained to keep factor identification simple
- Based on uncertainty analysis similar sources, e.g. SV-OOA and LV-OOA are merged together to get robust results
- Real-time update of the model is achieved by running a CMB on the incoming scans



PAUL SCHERRER INSTITUT

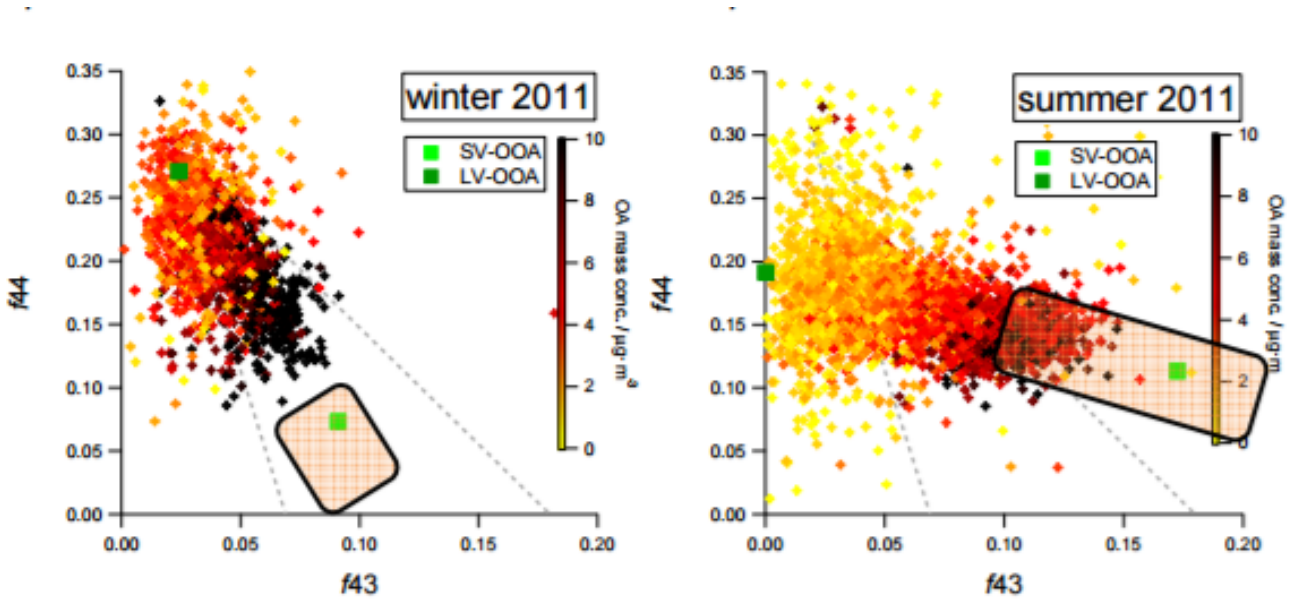
PSI

PSI, 22.08.2019

Main SoFi Pro features

Variability of factor profiles

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

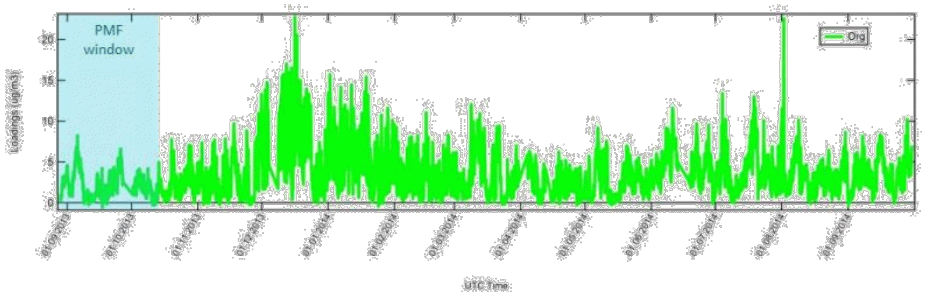


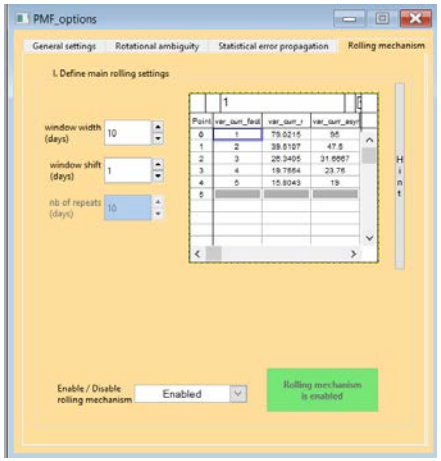
Canonaco et al. 2015

Automatic rolling PMF window

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

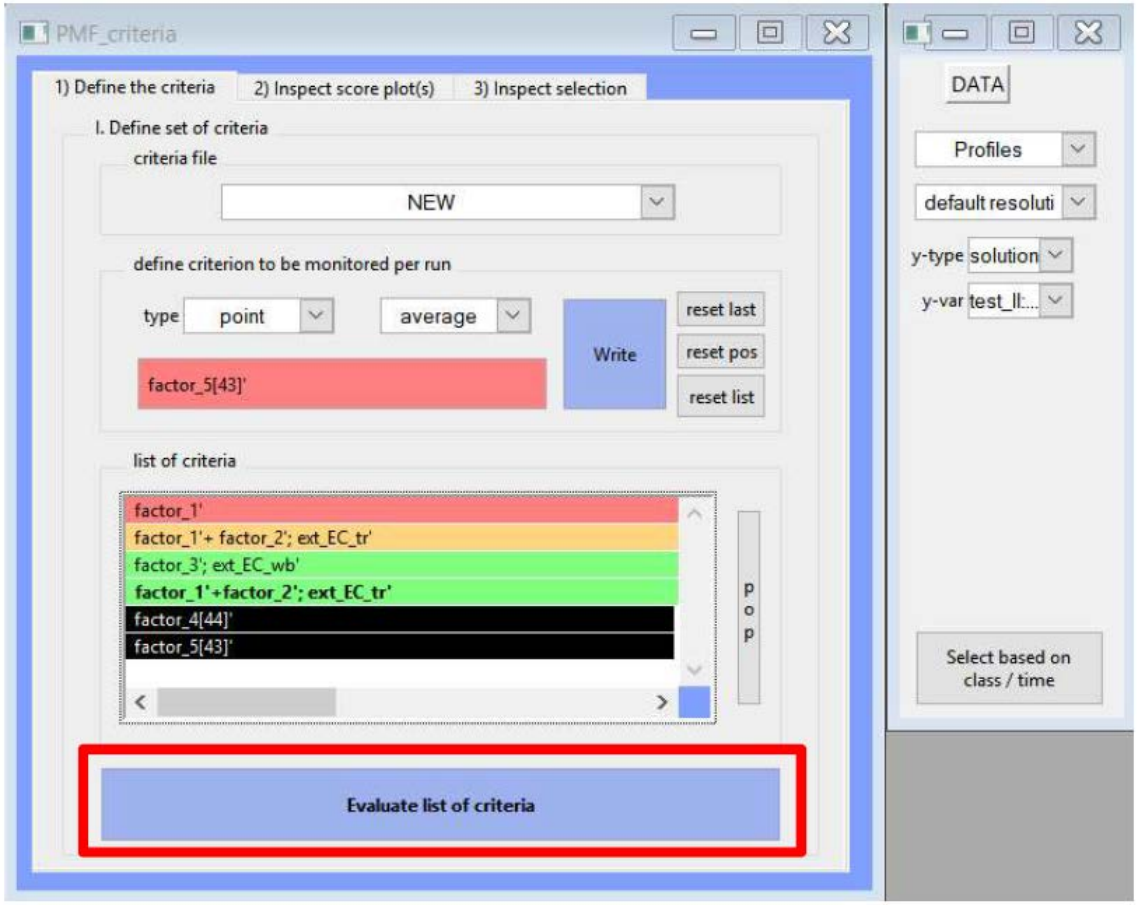
- PMF algorithm is run repeatedly on a short subset of data for a defined period
  - Assumption of constant aerosol sources during that time
  - after every shift the PMF runs are reinitialized (seed, *a*-value, fpeak, bootstrap, etc.)
- PMF window is subsequently shifted by defined period
- Thousands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





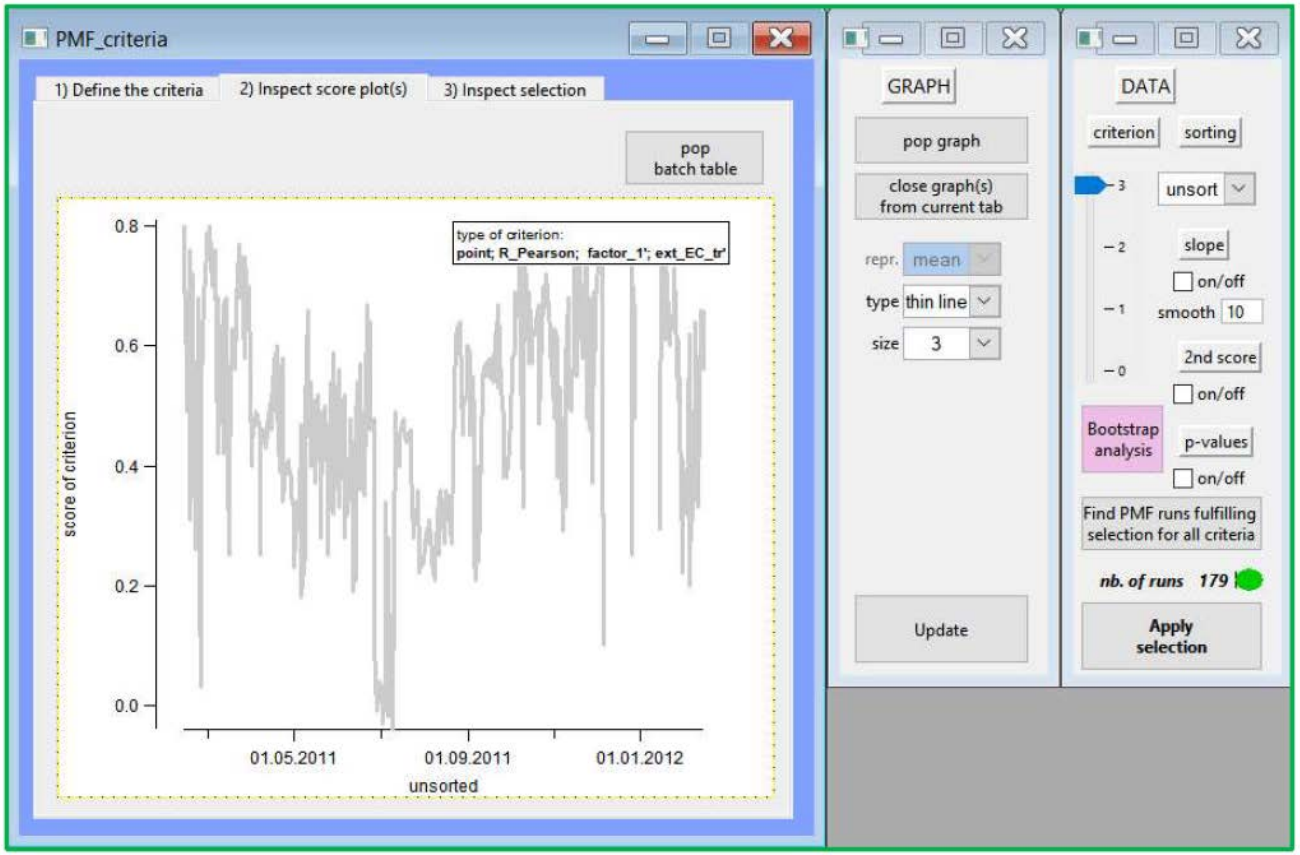
Automated criteria-based selection of PMF-runs

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



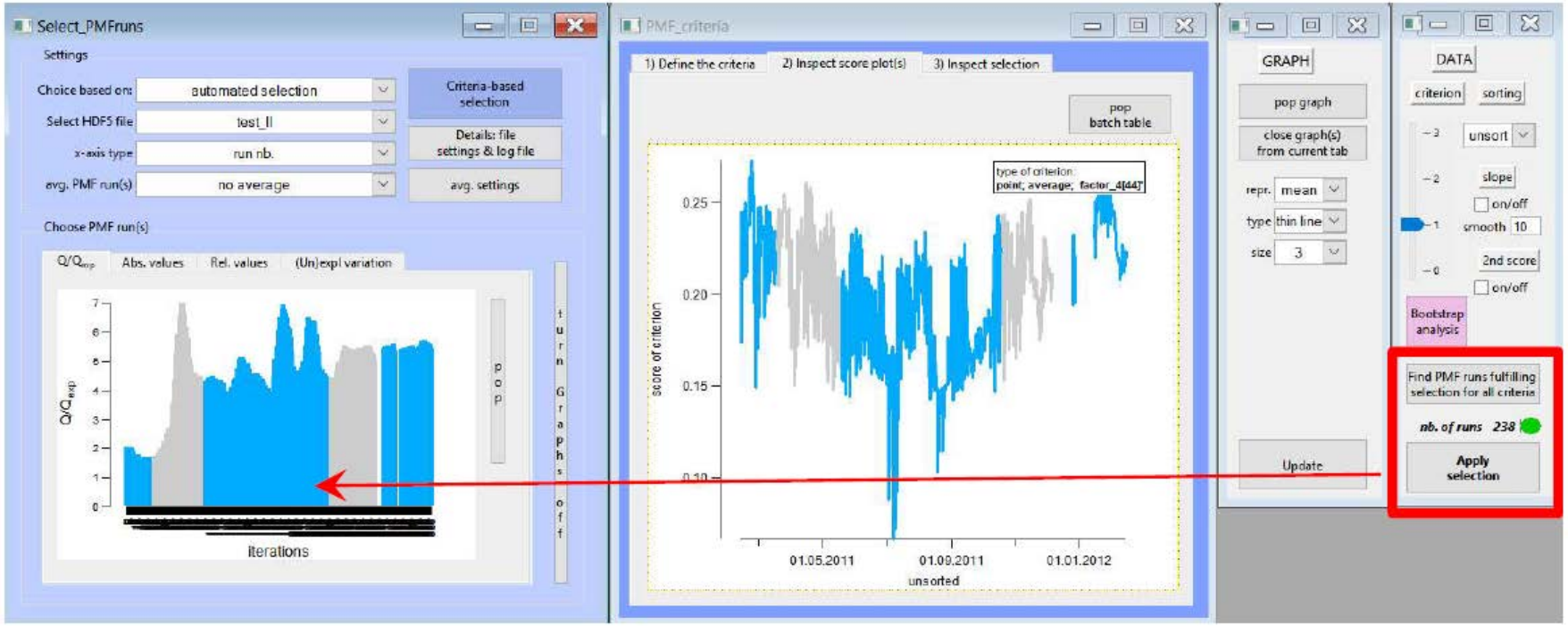
Automated criteria-based selection of PMF-runs

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



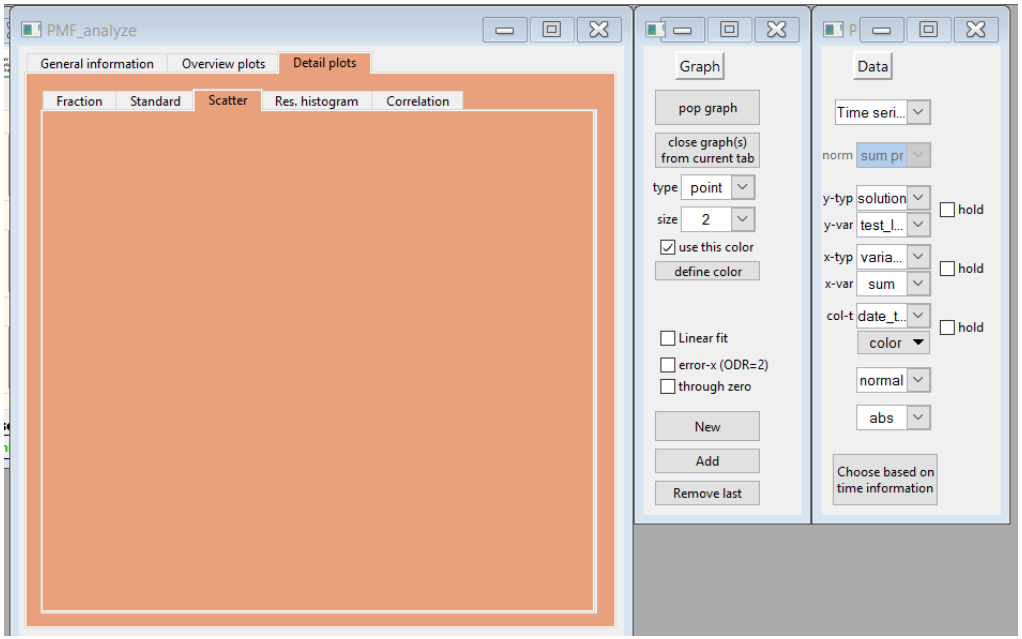
Automated criteria-based selection of PMF-runs

- 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.



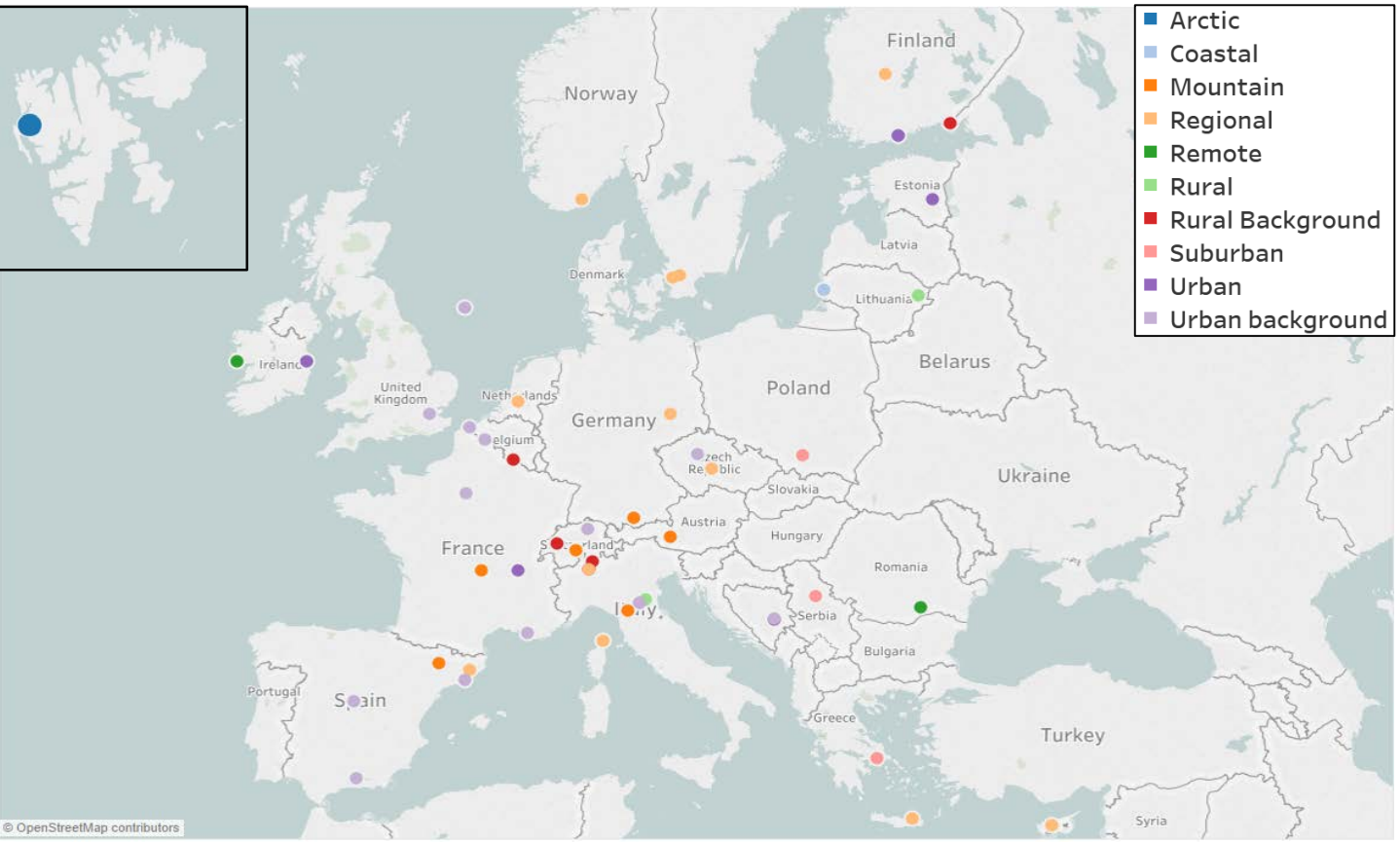
Include additional information

- 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

Developing a coherent approach using moving window applications in SoFi pro for stations in Europe



(adapted from the ACSM manual)

13:15
Room R31, 3rd floor, Tower 1, Gothia Towers.

https://www.psi.ch/acsm-stations/overview-full-period

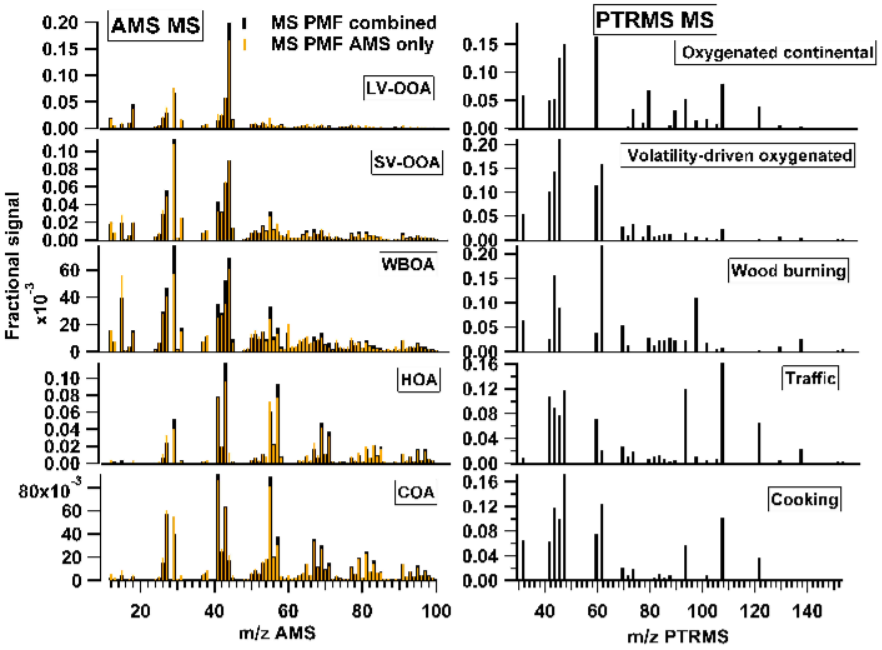
>40 monitoring stations across Europe have or had an Aerosol Chemical Speciation Monitor (ACSM) over a longer time period

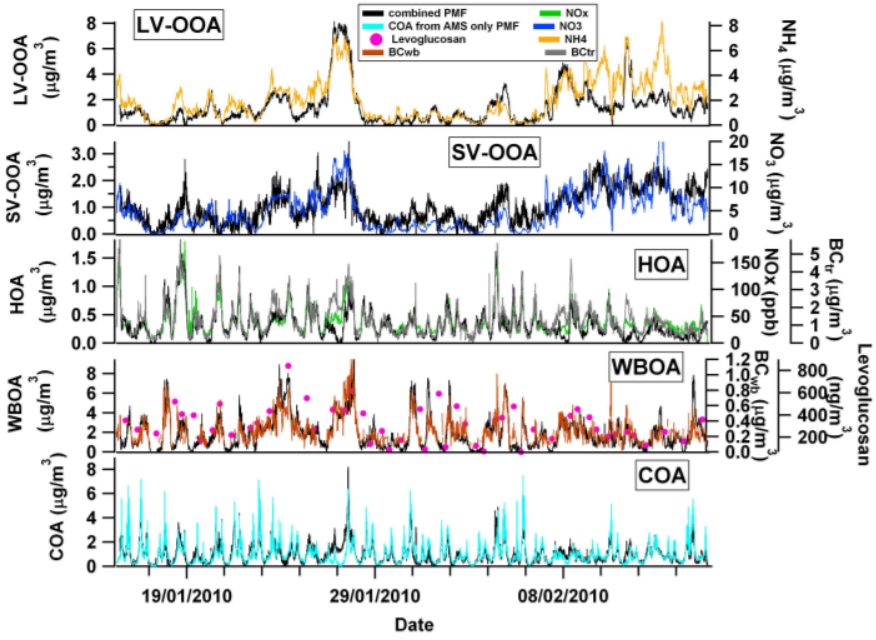
Activity within COLOSSAL / ACTRIS

francesco.canonaco@datalystica.com; andre.prevot@psi.ch

C-value approach

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

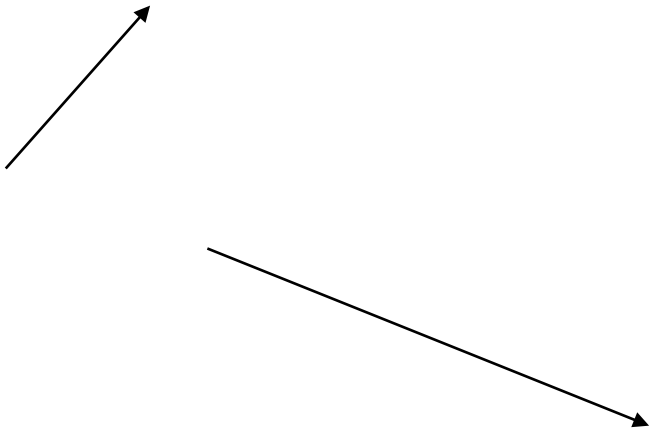




Crippa et al. 2013

C-value approach

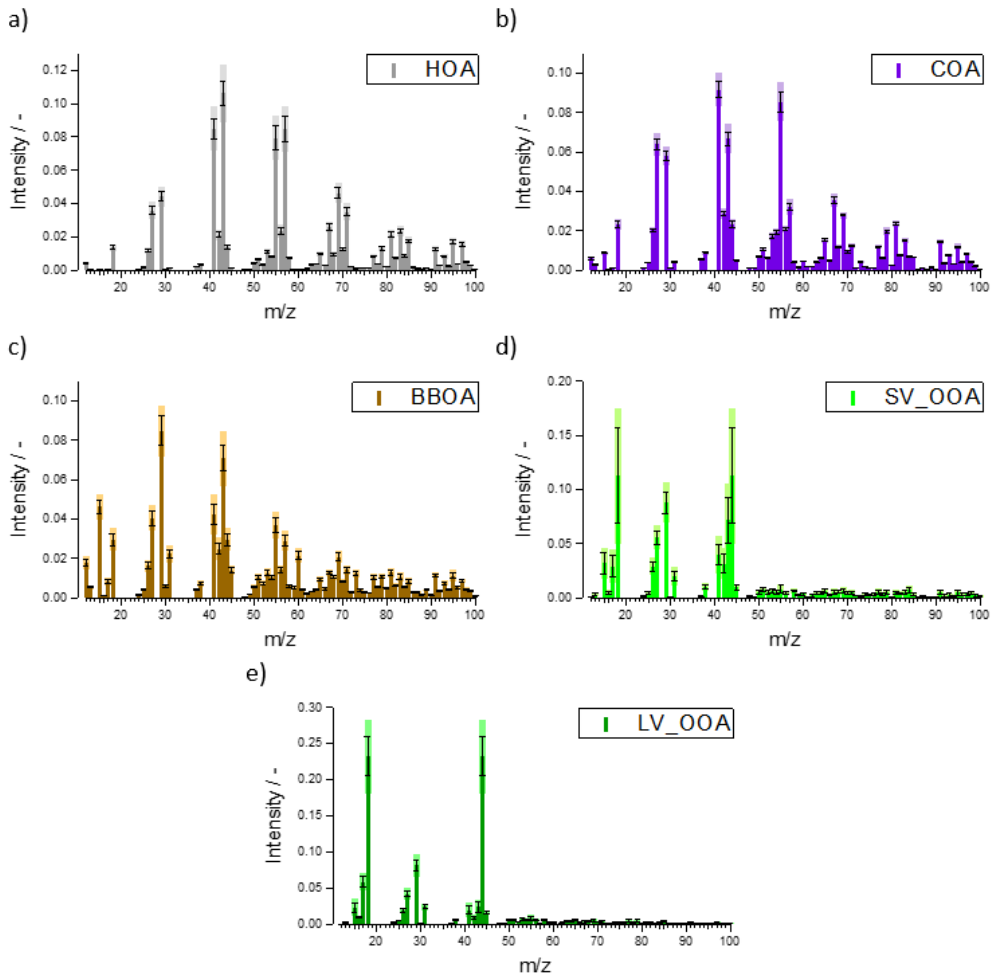
- automated weight of errors, e.g. when combining AMS with PTR-MS data



Crippa et al. 2013

PMF – uncertainties of PMF solutions

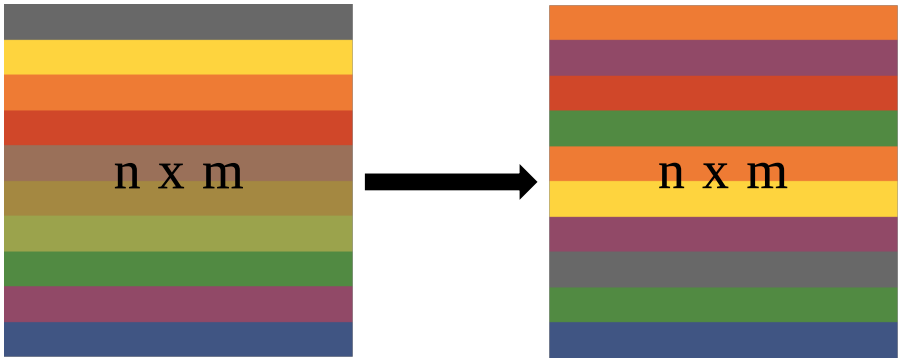
- PMF solution
  - PMF solution contains all PMF results that are environmentally reasonable (!not only one PMF run!)
  - rotational uncertainty (amount of rotational ambiguity) is assessed using e.g. the a value technique
  - statistical uncertainty, for ambient ACSM data mainly the daily variation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



Canonaco et al. in prep

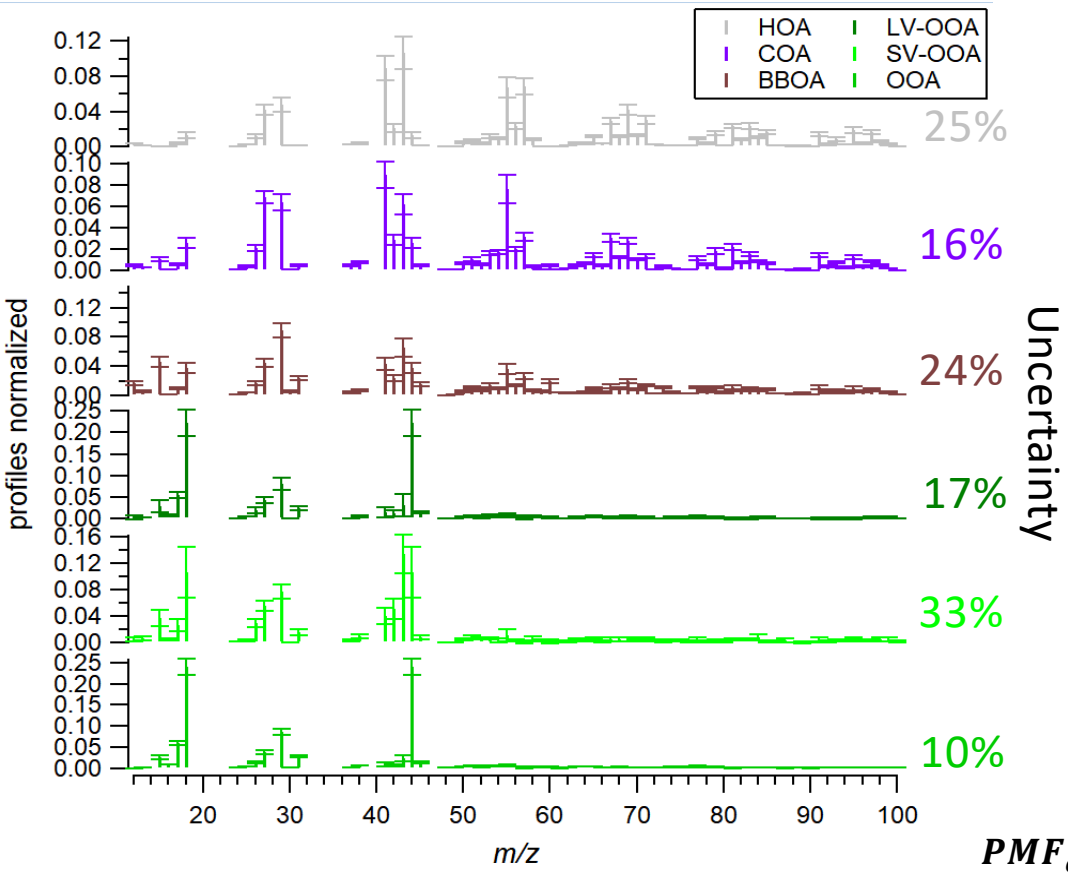
PMF – bootstrap BS

- 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



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

Yearly averaged Profile for Zurich



$$PMF_{error} = IQR(factor_i)\ vs\ mean(factor_i) = \frac{Slope}{2}$$

(Canonaco et al., in prep.)

Clear separation of factor profiles;

Uncertainties of each factor can be estimated in SoFi Pro