PAUL SCHERRER INSTITUT

PSI

Wir schaffen Wissen – heute für morgen

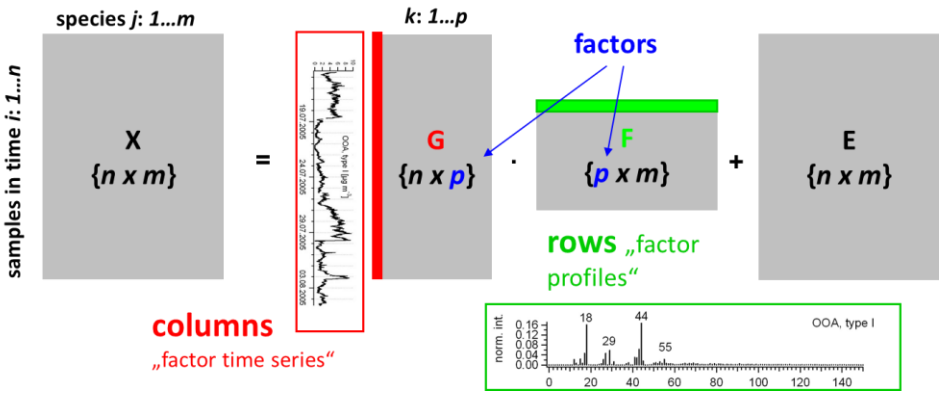
Paul Scherrer Institute

F. Canonaco, C. Bozzetti, K. Dällenbach, A. Tobler, J. G. Slowik, I. ElHaddad, M. Crippa, U. Baltensperger, **A. S. H. Prévôt** and many more

The Source Finder (SoFi)

Email to: francesco.canonaco@psi.ch

Source apportionment technique - PMF



bilinear PMF model

- rows of matrix F represent factor profiles
- columns of matrix G represent factor time series
- minimizing Q

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

advantages

- values in **G** & **F** are non-negative
- factors represent sources (POA) / aging (SOA)

disadvantages

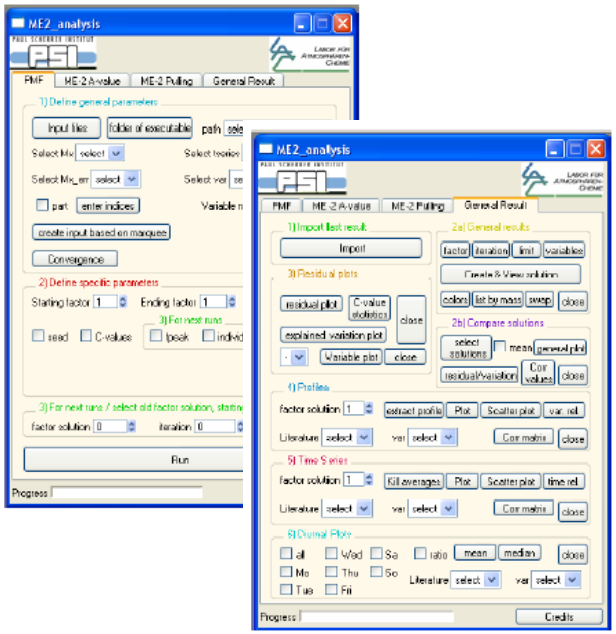
- constant factor profiles (mass spectra) over PMF run
- assess number of factors
- assess statistical error (e.g. resampling strategy)
- assess amount of rotational ambiguity

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

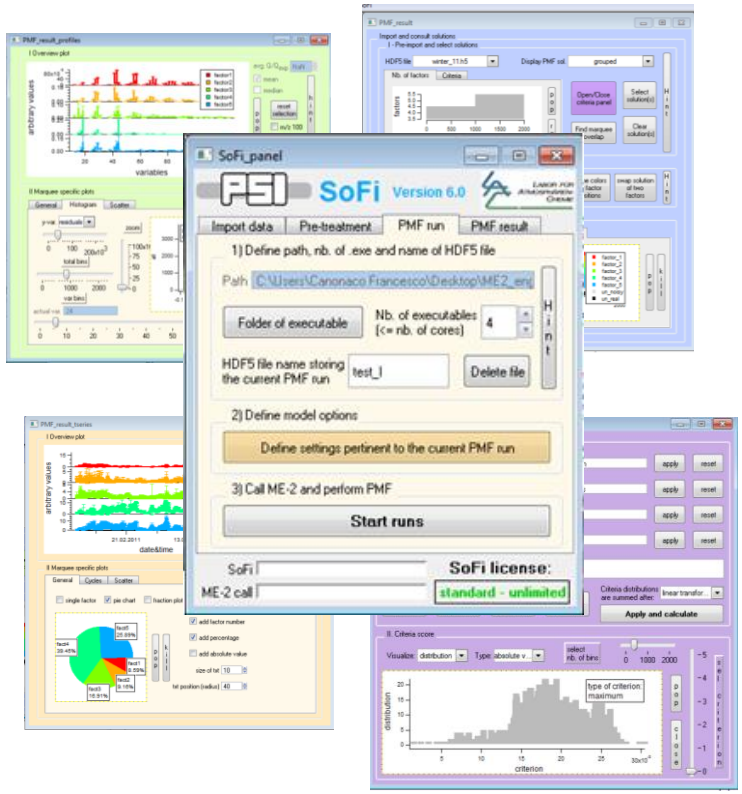
Paatero et al. 1994 / 2002 / 2008

SoFi – The Software Package

Canonaco et al., 2013



SoFi 6.4 current version



messy interface evolved into a…

fast, robust and very user-friendly panel-oriented software

SoFi – The Software Package

SoFi is

- IGOR-based, with the advantage of exploiting all good features (excellent and dynamic grafical quality, strong statistical package, HDF technology, threadsafe fcts., etc.)

…and the drawback to be updated to newer IGOR versions.

current SoFi package is **NOT** compatible with IGOR 7, nor IGOR 8. Upgrade directly to IGOR 8 planned during the next 6 months

- a modular software package currently with two modules, i.e. the standard SoFi and SoFi Pro (SoFi Pro runs only if standard SoFi is also compiled).
  In near future a third module for the real-time source apportionment (RT-SA or AuRo-SoFi) planned

SoFi – The Software Package

SoFi standard

- stable SoFi version, no changes in architecture
- compatibility is guaranteed from 6.0 onwards
- freeware but in collaboration (at least 2 peer-reviewed papers per scientific group)
- fast (exploits multiprocessors for PMF calc. and during the data treatment in IGOR)
- storage of PMF runs in HDF files avoids memory problems in IGOR and allows to perform and treat many PMF runs (thousands to millions)

SoFi – The Software Package

SoFi Pro

- statistical uncertainty can be quantified, e.g. using the **bootstrap resampling strategy** (see later)
- **Rolling mechanism** allowing source profiles to vary over time (see later)

➔ These two approaches generate thousands of PMF runs that must be sorted and classified

- Inspecting relevant features of the PMF run (e.g. specific *m/z*'s or correlation values) by providing **user-defined criteria** (see later) allows to compare thousands of PMF runs
- SoFi Pro evaluates averages of user-selected PMF runs and all SoFi standard features are also available for the **averaged PMF solution** including the statistics on the average

- Many more features to come in SoFi Pro (PMF input panel for treating non-IGOR data, pulling panel, control for trilinear PMF, dynamic marquee panel (DATADESK), etc.)

SoFi – The Software Package

SoFi standard and SoFi Pro homepage

- sign up to the Google Group **SoFi_ME2 under (https://groups.google.com/forum/#!forum/sofi_me2)** to receive news on SoFi, ME-2 and to get the current password for extracting the latest SoFi code

- Consult our new SoFi homepage (https://www.psi.ch/lac/sofi-sourcefinder) for

Manual

The manual for the released SoFi version can be found **here**.

Support

https://docs.google.com/spreadsheets/d/1W_kY3UJVl1TfxPzsZnrOHfiS6qUi6YfduUyir4hv7PI/edit?usp=sharing

Download

download of ME2 folder (2016/03/11) **here**.

download of SoFi standard (vers. 6.4, 2018/07/06) **here**.

download of SoFi Pro (vers. 6.4, 2018/07/06) **here**.

SoFi – The Software Package

Commercial aspects of SoFi Pro

- License-based system (beta version is for free until the end of the year)

The following table summarizes the costs for a SoFi license.

	1 PC per year	5 PCs per 1 years	1 PC per 5 years	5 PCs per 5 years
costs per PC and year / €	1000	750	750	500

costs required to guarantee the existence of a spin-off company in Switzerland for

- maintenance and further development of the SoFi package
- troubleshooting service

- In addition, SA service offered to institutes with no SA expertise

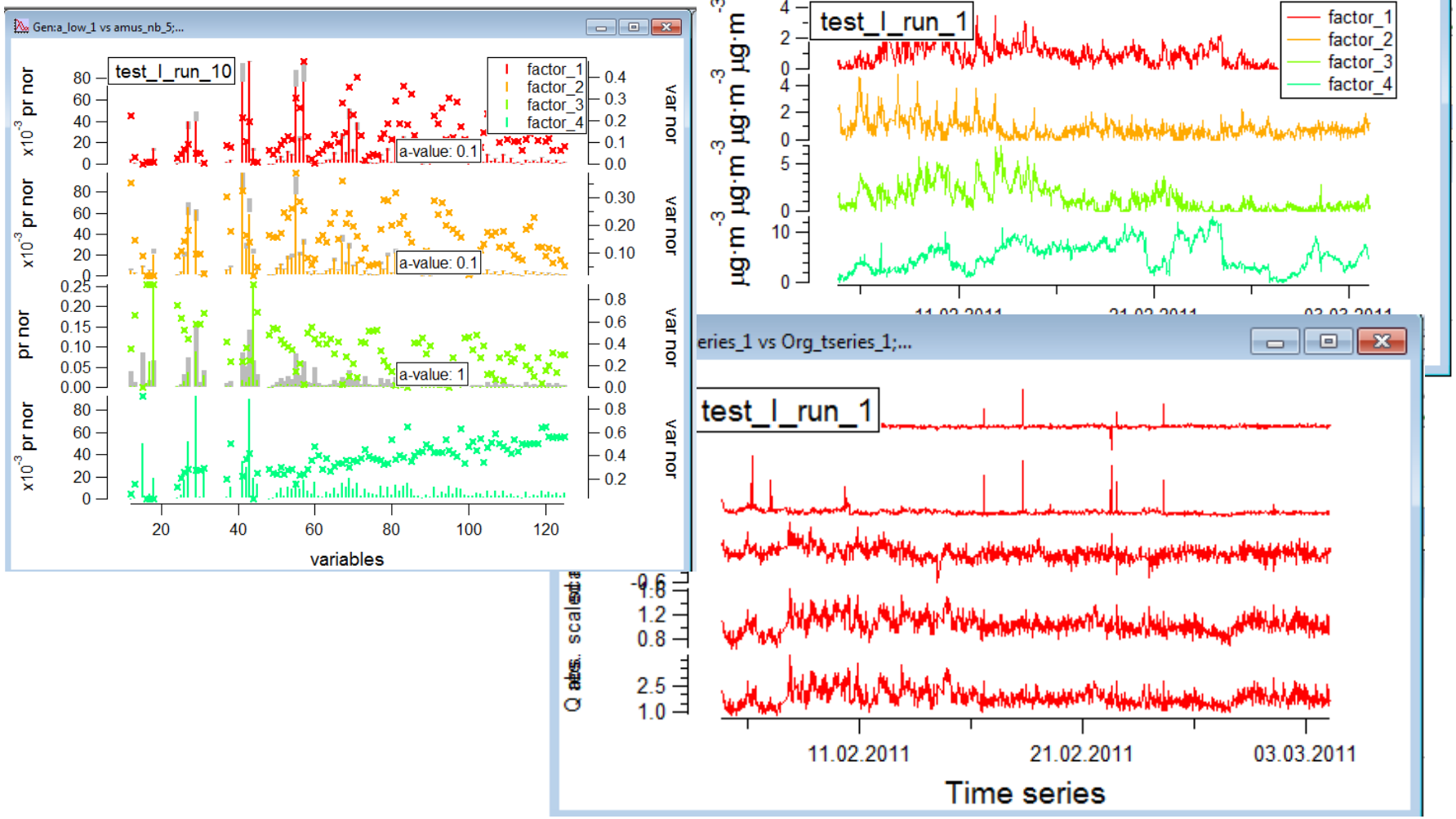
SoFi standard

Impressions of possible result plots

SoFi standard 6.4

results generated per mouse click from SoFi standard

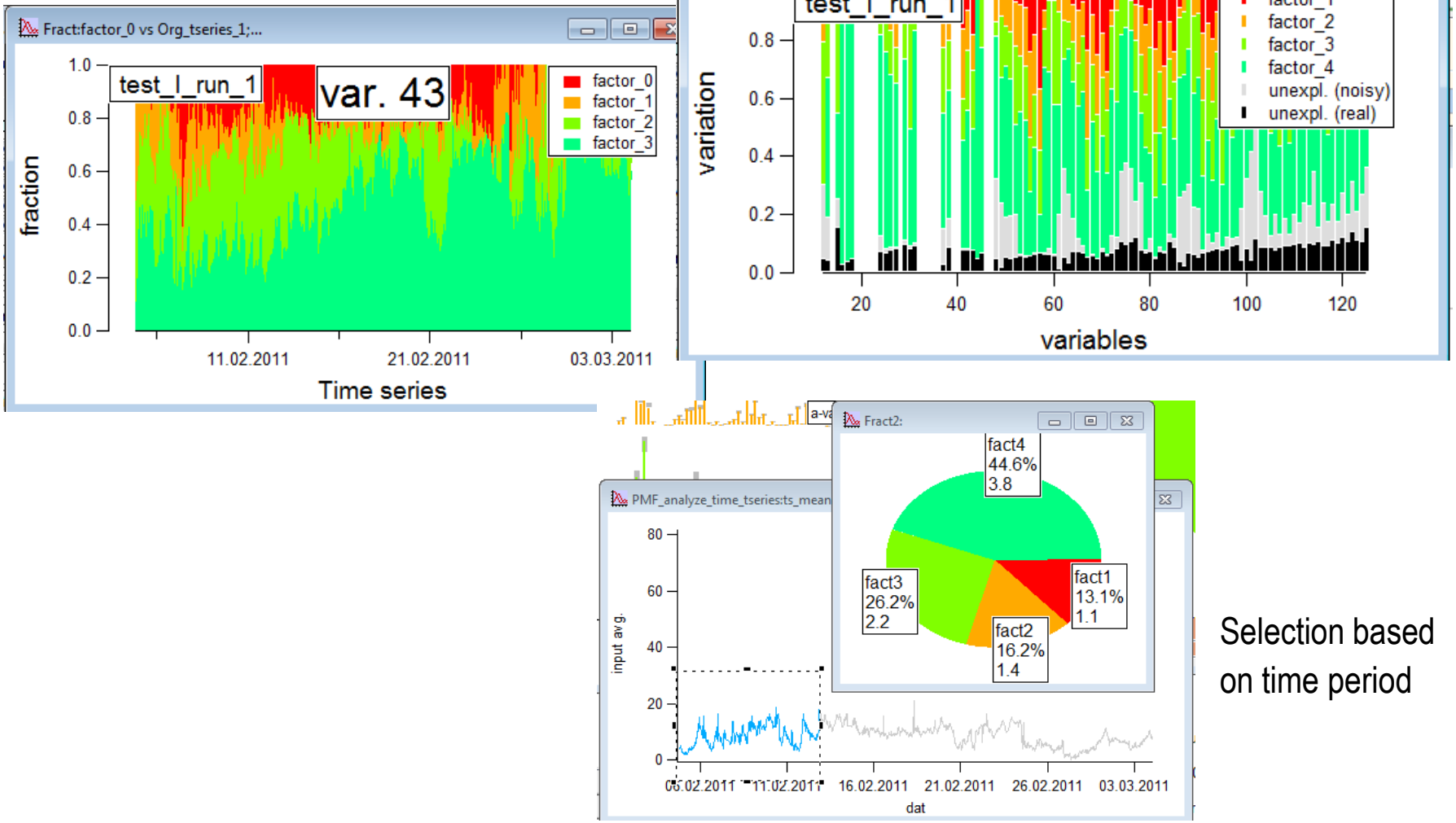
- Overview plots



SoFi standard 6.4

results generated per mouse click from SoFi standard

- Fraction plots



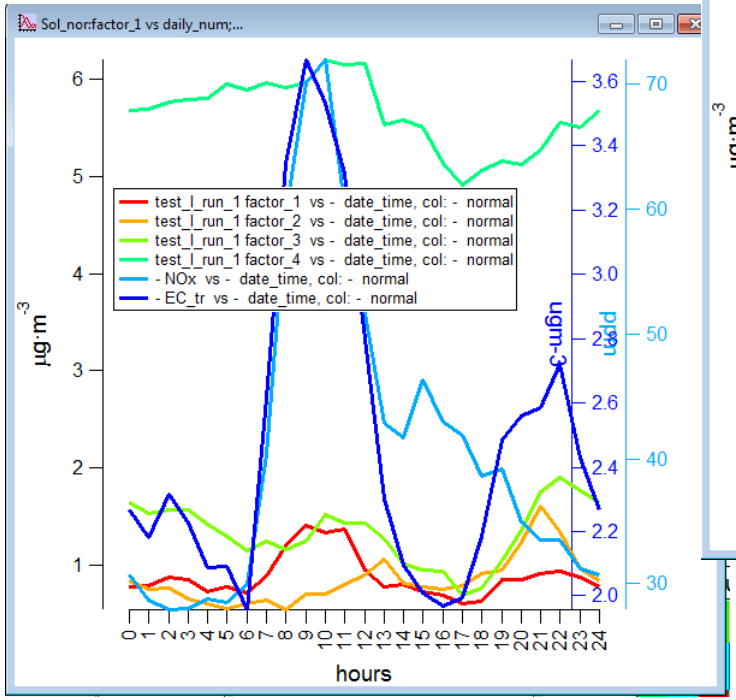
Selection based on time period

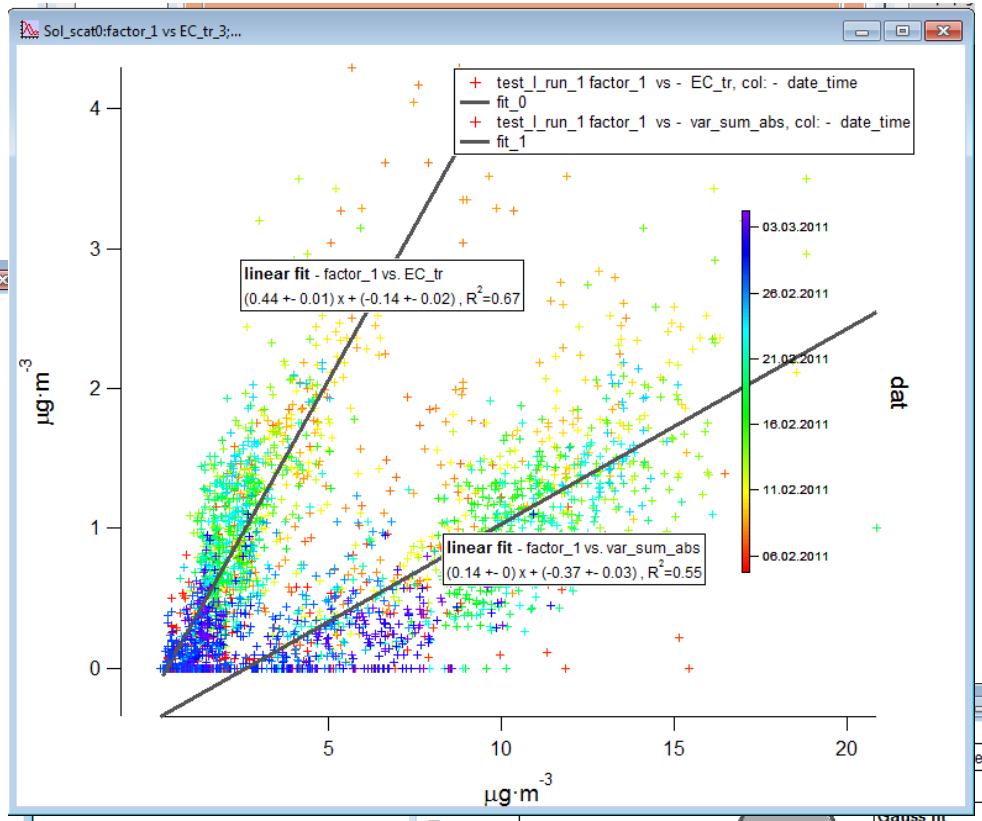
Email to: francesco.canonaco@psi.ch

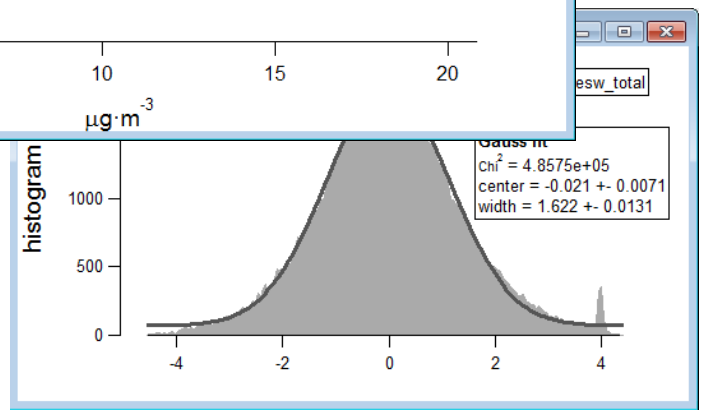
Standard SoFi 6.3

results generated per mouse click from SoFi standard

- Normal, scatter, residual plot



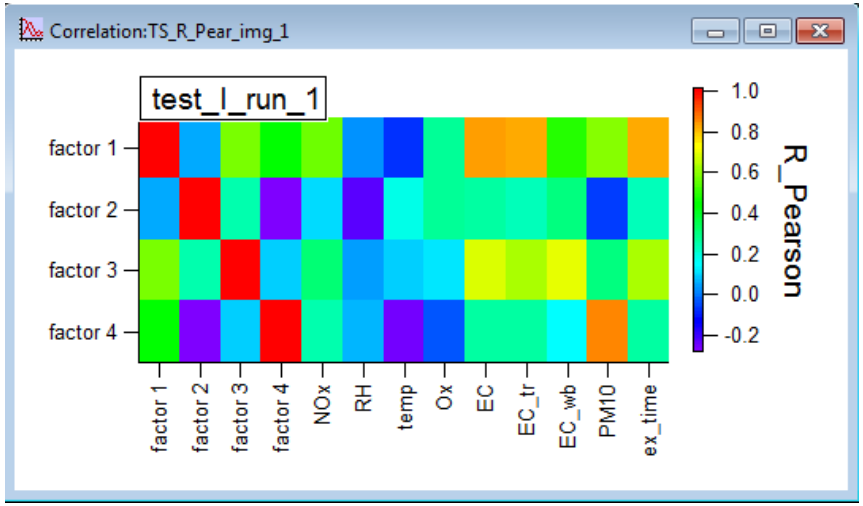


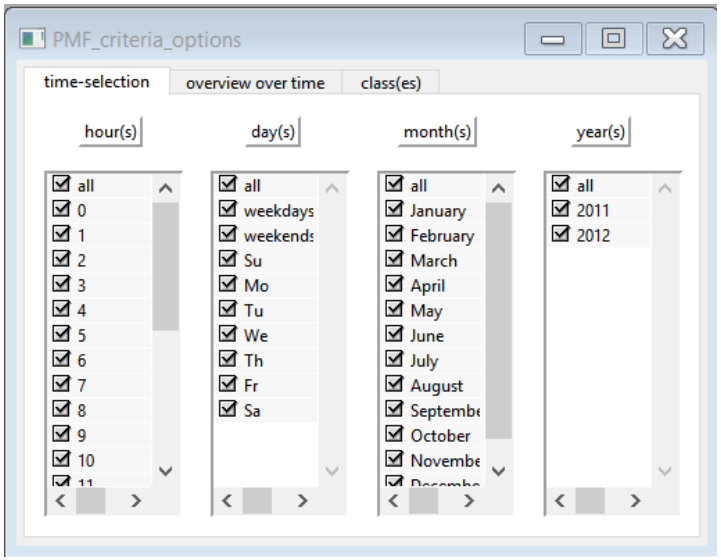


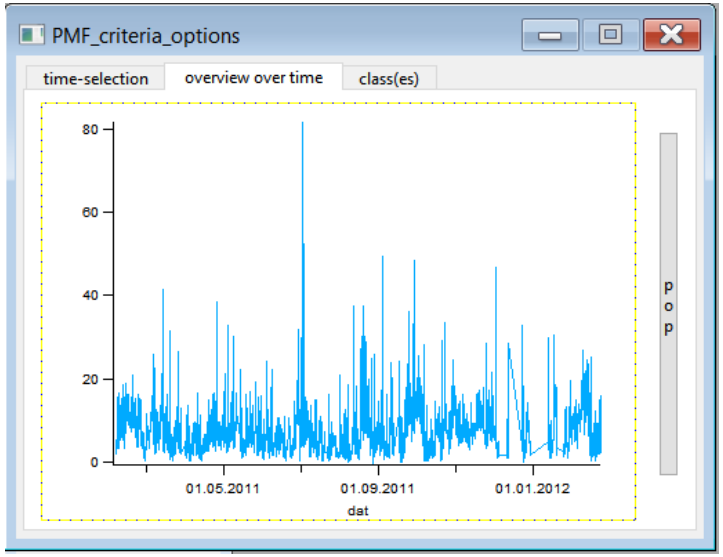
results generated per mouse click from SoFi standard

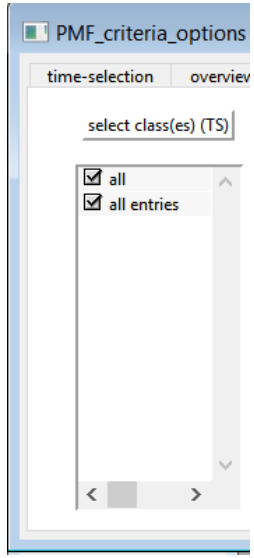
- Correlation matrix

For this table but also for all other results presented earlier user-selection (see below) can be easily applied





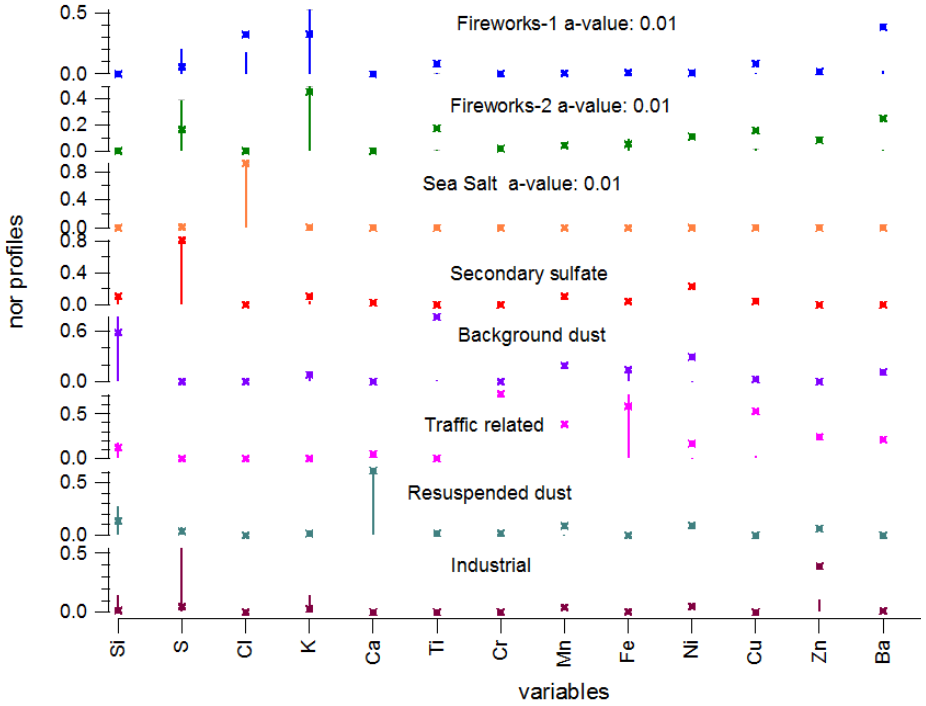


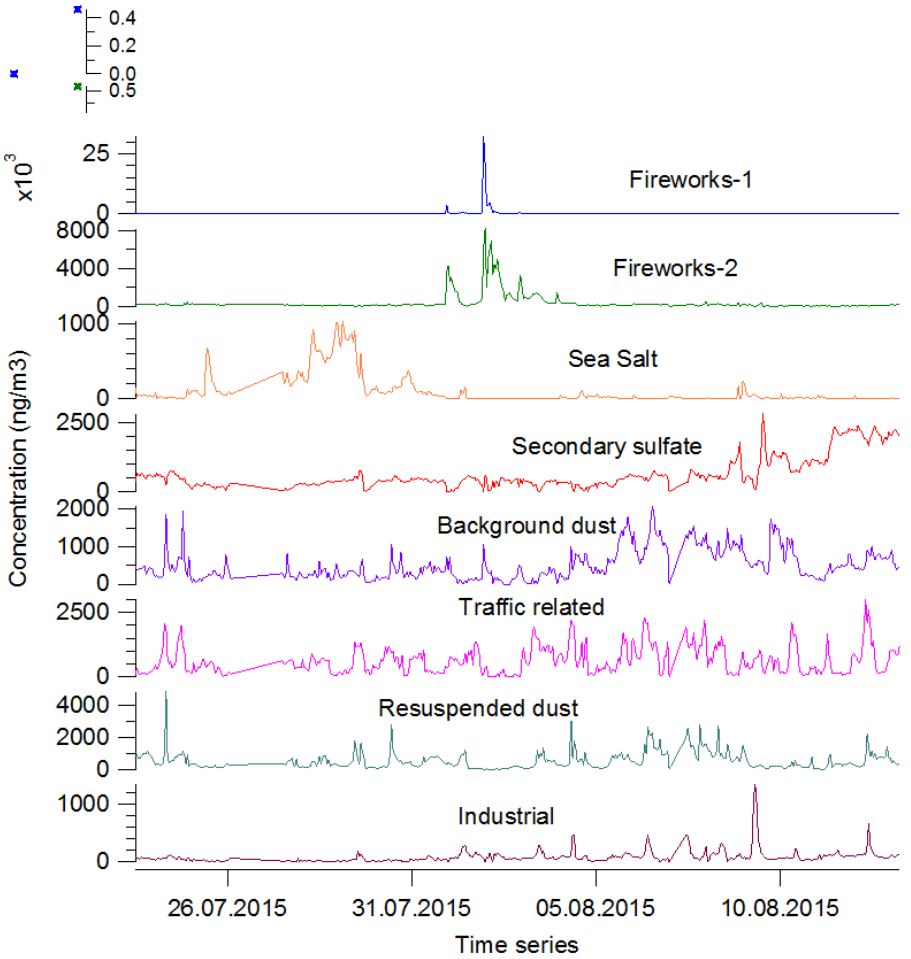


SoFi standard 6.4

results generated per mouse click from SoFi standard

- SoFi can also treat other kind of data, e.g. elemental data





Email to: francesco.canonaco@psi.ch

SoFi Pro

main features with some impressions

SoFi Pro 6.4

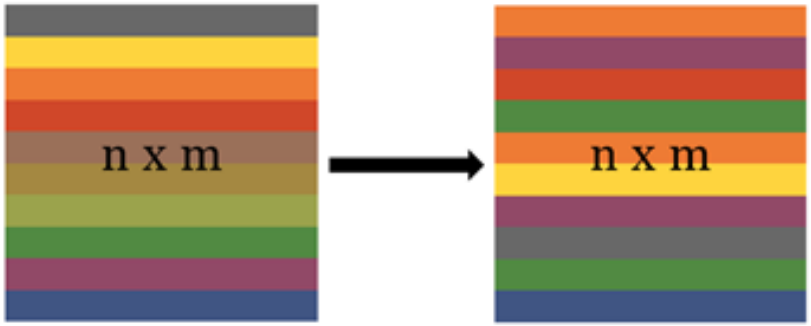
Main features

- full inspection of the averaged PMF solutions (results presented in standard SoFi before including statistics on the average accessible per mouse click)
- *Bootstrap (BS) resampling strategy*
- *Rolling window mechanism*
- *Definition of relevant properties as criteria for inspection and selection of thousands of PMF runs resulting from BS / Rolling*
- automated relative weight of errors, e.g. when combining AMS with PTR-MS data (C-value approach)
- Definition of classes (variables, e.g. different size-fractions or points in time, e.g. various stations) can be defined and results inspected separately

(points in *italic* are discussed)

Bootstrap (BS) resampling strategy

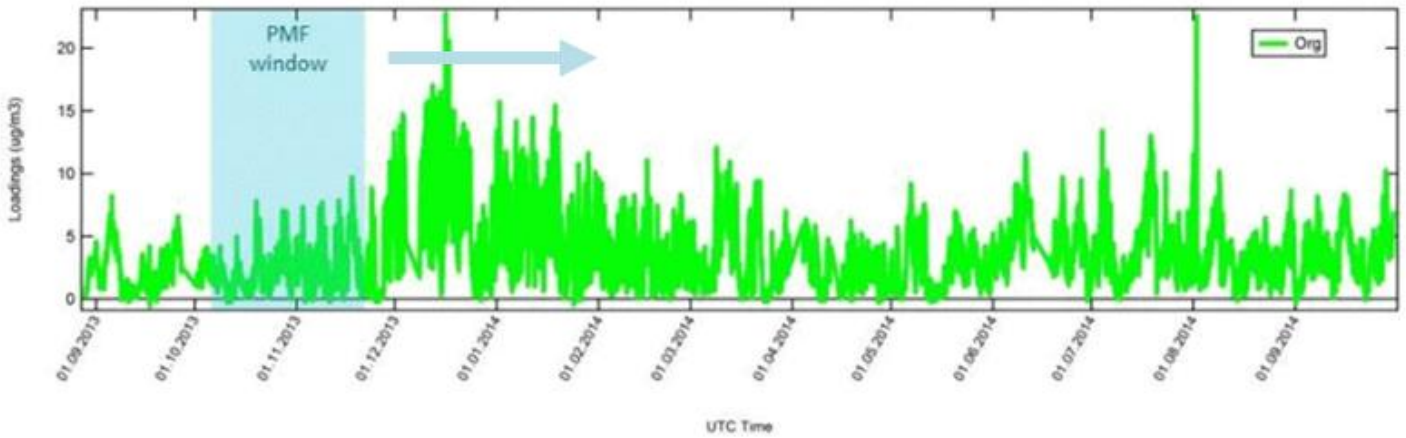
- BS generates a set of new input matrices for analysis from random resampling of the original input data by creating replicates of some points while excluding others (see below).
- Given that a sufficient number of resamples has been performed, the variation within the identified factors across all bootstrapped runs allows to estimate the statistical uncertainty.



Efron, 1979

Rolling mechanism

- PMF windows shifted over the PMF input
- after every shift the PMF runs are reinitialized (seed, *a* value, bootstrap, fpeak, etc.)
- length of window and shift are user-defined parameters



SoFi manual for 6.4
Canonaco et al., in prep.

SoFi Pro 6.4

Criteria-based approach

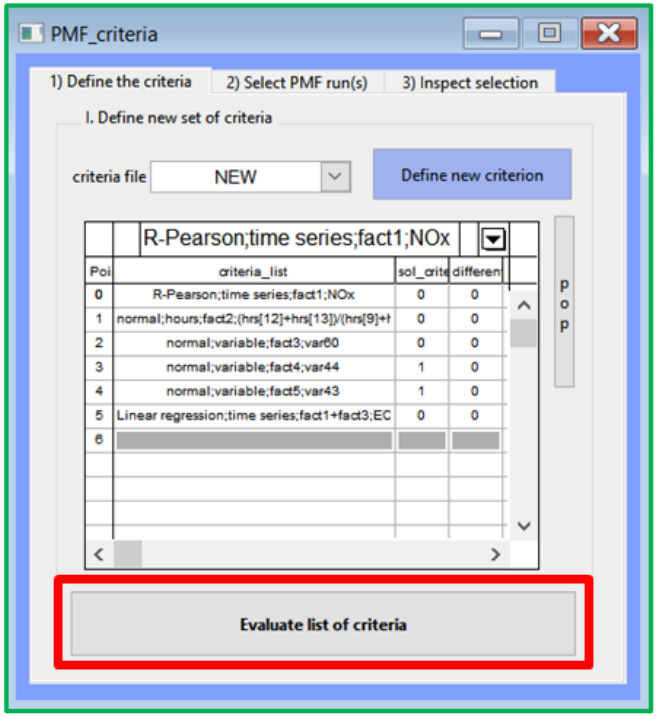
- User defines tracers/proxies (one or more) to be monitored for each factor (implementation on the right)

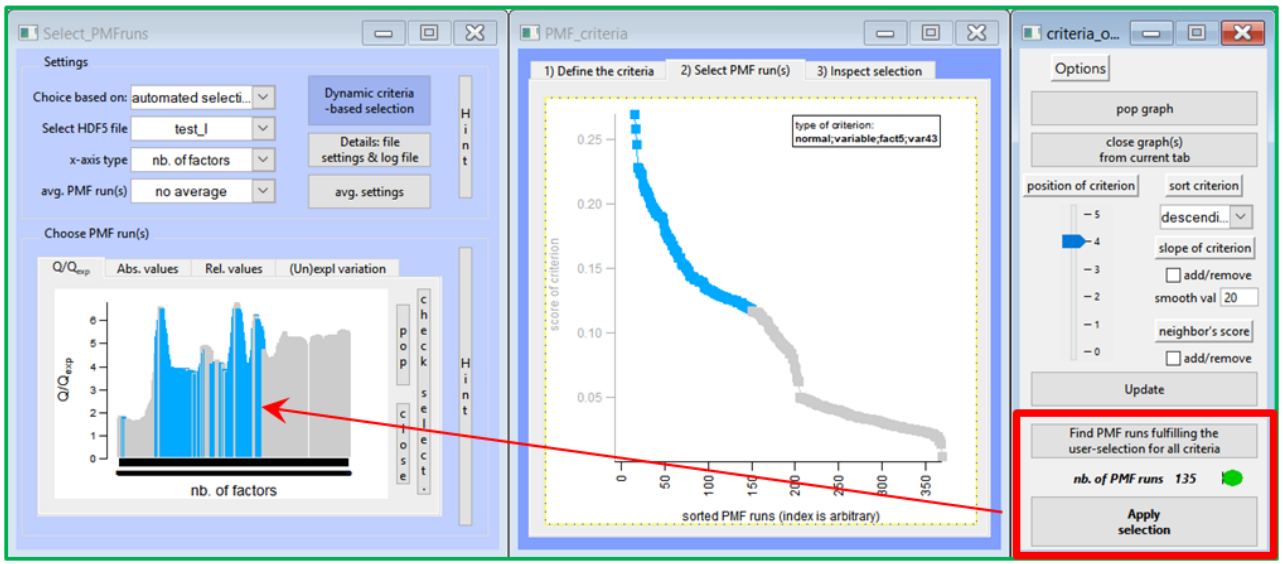
e.g. Correlation with external data (NOx with BCtr)

Minimum fraction of a certain m/z in factor (e.g. f60 in BBOA)

Clear diurnal in cooking

Multi-linear regression (e.g. BC with HOA/BBOA/CCOA)





SoFi manual for 6.4

Canonaco et al., in prep.

SoFi Pro 6.4

First publication on SoFi Pro using Rolling window on Zurich data submitted shortly

Canonaco et al., AMTD

PSI Positions

- **Tenure Track position** .. Towards a **permanent Scientist** Position.. Somebody that can handle many instruments/students/ ideally high source apportionment skills, experience in field/lab studies

Very likely:

- **Postdoc position** on on-line EESI applications (especially ambient in China, some lab studies)

- **Postdoc position** on Real-Time source apportionment (especially for measurements in China) using combination of ToF-ACSM, Xact (elements), Aethalometer (black/brown carbon)

- **PhD student position** on off-line AMS/EESI analyses (especially for filters in China)