Paul Scherrer Institute
F. Canonaco, M. Crippa, J. Slowik, A. Prevot, U. Baltensperger

SoFi (SourceFinder) based on ME-2
Source apportionment technique - PMF

\[ X_{\text{measured}} = \hat{X}_{\text{model}} + E_{\text{model}} \]

Bilinear model – Positive Matrix Factorization (PMF)
- The rows of the matrix \( F \) represent the factor profiles
- The columns of the matrix \( G \) represent the factor time series

Paatero and Tapper 1994
Source apportionment technique – 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)
- \( \sigma_{ij} \): uncertainty (measurement and/or model)

Advantages

- Values in G & F are non-negative
- Factors represent sources (POA) / aging (SOA)

Disadvantages

- Constant factor profiles (mass spectra)
- Validation of the solution \( \Rightarrow \) mostly b

Solver

- PMF2, only for unconstrained runs
- ME-2, unconstrained --- totally constrained runs (CMB) \textit{Paatero 1999, Paatero and Hopke 2008}
<table>
<thead>
<tr>
<th>PMF2 solver</th>
<th>ME2 solver</th>
</tr>
</thead>
<tbody>
<tr>
<td>runs with unconstrained G and F</td>
<td>runs with unconstrained G and F</td>
</tr>
<tr>
<td>global rotational tool (fpeak)</td>
<td>global rotational tool (fpeak)</td>
</tr>
<tr>
<td>Individual rotational tool (fpeak)</td>
<td>Constraining values (G and/or F) a-value/pulling</td>
</tr>
</tbody>
</table>

**a-value approach for F**

\[ f_{p,j,\text{solution}} = f_{p,j} \pm a \cdot f_{p,j} \]

**Pulling approach for F**

\[ f_{p,j,\text{solution}} = f_{p,j} + r_{p,j} \]

\[ a_j = \bar{f}_j + r_j \]

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

\[ Q^{aux} = \sum_{k=1}^{K} \left( \frac{r_k}{s_k} \right)^2 \]

\[ \arg\min_{G,F}(Q) = \arg\min_{G,F}(Q^m + Q^{aux}) \]

Paatero 1999,
Paatero and Hopke 2008
- Combined PMF, e.g. AMS with gas-phase data
ACSM winter 2011 / 2012 data in Zurich Switzerland analyzed with SoFi

Unconstrained case

- Evidence for HOA, COA, BBOA, SV-OOA and LV-OOA
- Mixed solution
- Global rotational tool (fpeak) didn’t enhance the solution

Canonaco et al., AMTD
Results – constraining full profiles

Canonaco et al., AMTD
Results – constraining full profiles

Canonaco et al., AMTD
Source apportionment strategy (main part)

- Unconstrained run (PMF)

- Constrain expected and well-known factors (e.g. HOA) and test various number of factors

- Investigate the structure (diurnal) and levels of f60 / CO → constrain BBOA

- Investigate the structure of the residuals to guess the presence of other sources not represented yet by the model (e.g. COA)

Crippa et al., ACPD
Results – constraining single m/z’s in a profile

PMF solution

Pulling parameters:
Anchors=0.08
dQ=1000
Weight=0.1

Total increase in Q
~0.06%

Canonaco et al., in prep.
Results – constraining single m/z’s in a profile

$R^2_{\text{pearson}}$ (HOA and BC): from 0.3 to 0.6

Canonaco et al., in prep.
SoFi – explained/unexplained variation

- User defines the threshold for the S/N ratio (default is 2) crucial for the variation plots

\[
\text{variation}^{\text{explained}} = \frac{\text{variation}^{\text{model}}}{\text{variation}^{\text{measured}}}
\]

\[
\text{variation}^{\text{unexplained}} = 1 - \text{variation}^{\text{explained}}
\]

gray = unexplained variation but noisy
black = unexplained variation
Reweighting of some variables, e.g. input data from different instruments

PMF successful for variables until 100

Consider C-value (reweighting factor) for which:

- the distribution of the unexplained variation is ~ unimodal
- spread is minimal
Monitor the correlation of a factor time series with external data, e.g. HOA with BC over different model runs
Compare different model runs together