ME-2: interface updates and analysis strategies


Paul Scherrer Institute, Laboratory of Atmospheric Chemistry
Matrix rotation problem
Can insert product of transformation matrix and its inverse (= identity matrix)
\[ GF = GTT^{-1}F \]
Define new solution matrices
\[ G' = GT, F' = T^{-1}F \]
Many mathematically equivalent solutions
\[ GF = G'F' = G''F'' = G'''F''' = \ldots \]
…And we also must consider approximate rotations
\[ GF \sim G'F' \sim G''F'' \sim G'''F''' \sim \ldots \]

PMF2: one (randomly chosen) rotation + 1-dimensional exploration via fpeak
→ others inaccessible!
ME-2: Every solution accessible

Out of n x m dimensions
As expected, ME-2 improves model performance vs. PMF2

Cost:
- More computationally intensive
- More user care required

Canonaco et al., 2013
ME-2 overview

Challenge: every solution is accessible in ME-2
→ need computationally reasonable way to explore the solution space

Current approach: a-value
→ require 1 or more factors to return a profile (or TS) within predefined range
\[ f_{p,j,solution} = f_{p,j} \pm a \cdot f_{p,j} \]

Caution: a-value variation not necessarily sufficient for exploration
→ local minima or inflections in Q-space

Recommend varying both a-values AND profiles
→ impractical to rely on “user inspection” for solution selection/validation
→ new tools – systematic & objective exploration/evaluation
New analysis paradigm... SoFi v5+

Optimized for multi-run execution & post-analysis
  → programmable exploration of a-values and anchor profiles
  → programmable criteria for selecting/visualizing acceptable solutions

Pre-analysis tools
  → quick look at data, downweighting tools

Lots of miscellaneous improvements
  → handle missing data, define outliers for robust mode, specific instrument support
A quick tour… model setup

II. Define output file

IV. Run and Consult PMF solutions

Start runs

Open/Close result panel

Solution space investigation

I. Define exploration

seed

II. Define seed type

ran...

III. nb. of iterations

IV. Open/close model entries

Delete data
A quick tour… a-value method

June 20, 2015
Jay G. Slowik
Require 1 factor to be between HOA_avg (Ng et al.) & HOA_Paris (Crippa et al.)

Model varies factor profiles within these boundaries

Other configurations possible: conventional a-value, a-value scans…
Analysis strategies

Anchor profile selection
- Literature
- Extraction from dataset

Selection/validation methods
- External tracers
- Multiple criteria types (tracers + diurnals + profile characteristics + …)

Analysis of long-term datasets
- Rolling window analysis
Extracting profile from a dataset

ToF-ACSM data at Jungfraujoch
Most pollution advected from valley with boundary layer movement

Organic data punctuated by short-duration spikes → propagate into ME-2 factors (esp. HOA)

Solution: PMF on spike periods → extract profile related to cigarette smoke

Fröhlich et al., in prep
Extracting a profile from a dataset

Final Jungfraujoch solution

Good separation of HOA & POA_{loc} (both constrained), despite similar profiles

Fröhlich et al., in prep
Case #1 -

Summer/winter “offline AMS” data – Switzerland
Constrain COA & HOA

Three sets of tracer/factor criteria (“x”)

For each, calculate:

\[
\frac{sdev(\log(x))}{mean(\log(x))}
\]

Sum all criteria to select solution

Dällenbach et al, in prep
Solution shows systematic winter/summer differences. 
→ Is variability well captured, or are their systematic biases?

Dällenbach et al, in prep
Case #2: combining various criteria

- HR-ToF-AMS data from online measurements in China (2013-2014)

Unconstrained solution: 5 factors

- High m/z 60 (also in 25 factors solution)
- High HOA/BCtr (from literature we find that it can go up to 1 but PMF gives 1.5)
- Constrain HOA Paris winter (Crippa., 2013)
- High m/z 44
- Diurnal gets bad if fixing HOA (Background and structure over night)

* CCOA: Coal combustion organic aerosols

Elser et al., in prep
Criteria for a-value optimization

» HOW TO OPTIMIZE THE CHOICE OF THE a-VALUES?

Step 1- Run all possible a-value combinations:

This work: a-value: 0 to 1 with steps of 0.1

Step 2- Establish criteria to choose good solutions:

This work:

1) Hard criteria (YES or NO):
   - Minimize m/z 60 in HOA
   - Optimize diurnal COA

2) Soft criteria: correlation with externals
   - $\text{NH}_4 \text{ vs OOA}$
   - $\text{BC}_{\text{tr}} \text{ vs HOA}$
   - $\text{BC}_{\text{wb}} \text{ vs BBOA}$
   - $\text{PAH}_{\text{fitted}} = a*\text{BBOA} + b*\text{CCOA} \text{ vs PAH}^*$

* PAH from HR-AMS
  If PMF with OA+PAH → PAH in CCOA and BBOA
  Linear regression of PAH vs. BBOA ,CCOA → same result

Elser et al., in prep
Hard criteria (1)

- **Low m/z 60 in HOA**
  
  Accepted limits? 
  
  $a COA = 0$
  
  $a COA = 1$

- **Optimize COA diurnal**
  
  Accepted limits? 
  
  Cluster analysis 

---

June 20, 2015

Jay G. Slowik
When to stop?

Take red cluster in 3 clusters solution

Elser et al., in prep
Hard criteria (3)

- Low m/z 60 in HOA

- Optimize COA diurnal

Do 100 repetitions and use only combinations that are in same cluster at least 95% of the times

Elser et al., in prep
Soft criteria

\[ f_{\text{ext}} = \frac{\text{Measured-Fitted}}{\text{Measured}} \]

\[
\begin{align*}
\text{PAH fitted } (t) &= a \cdot \text{BBOA}(t) + b \cdot \text{CCOA}(t) \\
\text{NH}_4 \text{ fitted } (t) &= (\text{NH}_4/\text{OOA})_{av} \cdot \text{OOA}(t) \\
\text{BC}_{\text{tr}} \text{ fitted } (t) &= (\text{BC}_{\text{tr}}/\text{HOA})_{av} \cdot \text{HOA}(t) \\
\text{BC}_{\text{wb}} \text{ fitted } (t) &= (\text{BC}_{\text{wb}}/\text{BBOA})_{av} \cdot \text{BBOA}(t)
\end{align*}
\]

Minimize the error (\(\sigma_{f_{\text{ext}}})\):

\[
\sigma_{\text{ALL}} = \sqrt{(\sigma_{\text{PAH}})^2 + (\sigma_{\text{NH}_4})^2 + (\sigma_{\text{BC}_{\text{tr}}})^2 + (\sigma_{\text{BC}_{\text{wb}}})^2}
\]

Take solution with minimum: \(a_{\text{HOA}} = 0.5\) & \(a_{\text{COA}} = 0.5\)

Perturb input Org matrix inside the errors

Variability of solution for 10 perturbed runs:

\[
\sigma_{\text{ALL}}(0.5&0.5) = 1.04 - 1.17 \rightarrow \sigma = 11.26 \%
\]

We take all solutions with:

\[
\sigma_{\text{ALL}} < \sigma_{\text{ALL}}(0.5&0.5) + 2\sigma
\]

Elser et al., in prep
"Final" solutions

Elser et al., in prep

June 20, 2015

Jay G. Slowik
Profile variation

XRF measurements of trace elements in London
→ Already determined number of factors, appropriate constraints, etc.

Now, test profile sensitivity
→ 10,000 runs, randomly varying all profiles within +/-20% of base, with \( a = 0 \)
→ accept only solutions satisfying criteria below (25-65% of solutions)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM(_{10-2.5}) Relative intensity in brake wear factor of Cu + Zn + Zr + Mo + Sn + Sb + Ba</td>
<td>&gt; 75%</td>
</tr>
<tr>
<td>Al/Si in resuspended dust factor</td>
<td>±40% of 0.3(^a)</td>
</tr>
<tr>
<td>Mg/Na in aged sea salt factor</td>
<td>±40% of 0.12(^b)</td>
</tr>
<tr>
<td>PM(_{2.5-1.0}) Relative intensity in brake wear factor of Cu + Zn + Zr + Mo + Sn + Sb + Ba</td>
<td>&gt; 70%</td>
</tr>
<tr>
<td>Al/Si in resuspended dust factor</td>
<td>±40% of 0.3(^a)</td>
</tr>
<tr>
<td>Mg/Na in aged sea salt factor</td>
<td>±40% of 0.12(^b)</td>
</tr>
<tr>
<td>PM(<em>{1.0-0.3}) Avg. Cl/Na in mean good solutions of PM(</em>{10-2.5}) and PM(_{2.5-1.0}) in sea/road salt factor</td>
<td>Cl/Na ±20% of avg.</td>
</tr>
<tr>
<td>All ( Q/Q_{exp} )</td>
<td>≤ 10% of min. ( Q/Q_{exp} )</td>
</tr>
</tbody>
</table>

Want a significant fraction of rejected solutions to ensure space is completely explored

Aged sea salt

Visser et al., 2015
Analysis of long-term datasets

SOA f44 vs. f43 for 1 year of data in Zürich

f43/f44 variation always roughly on a line… but different slope in winter vs. summer

Cannot represent yearly OOA with just two factors!

Option 1: Add more OOAs → interpretation (esp. for ACSM)?

Option 2: ME-2 in rolling window → “easy” to implement
Open question whether to sort chronologically or by temperature

Same data analyzed in multiple windows → Uncertainty estimates
Model setup - evaluation

<table>
<thead>
<tr>
<th>Factor</th>
<th>Correlation</th>
<th>Criterion</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary OA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOA</td>
<td>Time series, diurnal cycle</td>
<td>$R^2$(HOA, NOx)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$R^2$(HOA, BC_traffic)</td>
<td>1</td>
</tr>
<tr>
<td>BBOA</td>
<td>Time series, diurnal cycle</td>
<td>$R^2$(BBOA, BC_wood_burning)</td>
<td>2</td>
</tr>
<tr>
<td>Secondary OA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SV-OOA</td>
<td>Time series, diurnal cycle</td>
<td>$R^2$(SV-OOA, NO$_3^-$)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$R^2$(SV-OOA, Temperature)</td>
<td>1</td>
</tr>
<tr>
<td>LV-OOA</td>
<td>Time series, diurnal cycle</td>
<td>$R^2$(OOA, NH$_4^+$)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$R^2$(LV-OOA, SO$_4^{2-}$)</td>
<td>1</td>
</tr>
</tbody>
</table>

- For each PMF window $\rightarrow k$ solutions for $a$-value within the range [0, $a$]
- To estimate the best solution for a PMF window, we calculated the Pearson’s correlation coefficients between the factor solutions and inorganic species measured by ACSM or an external parameter (NOx, black carbon, T)
- **Only one solution** that maximizes the sum of all the criteria $\sum R^2$ is to be chosen!!

Will change
Evaluation – quick look at criteria

Magadino  
(temperature-sorted)

Ispra  
(chronologically-sorted)

Different criteria are more/less important depending on season
Modelled HOA and BBOA factors were constrained using "reference" profiles obtained from PMF analysis of the winter Magadino data, within a \( [0, 0.5] \) value range. 

A local factor was constrained using a LOA profile from PMF run on the summer data, a \( [0, 0.5] \) value range. From aethalometer data, a trend is observed. 

From aethalometer data, there is a trend observed.
Conclusions/Outlook

ME-2 panel & analysis is rapidly evolving

→ Emphasis on systematic, objective methods for exploring/evaluating the solution space
  → a-value exploration
  → profile variation
  → acceptance of multiple “best” solutions

Additional discussion/development:

→ SoFi download: www.psi.ch/acsm-stations/me-2
→ ME-2 mailing list
→ ME-2 workshops (probably next winter)