ME-2 analysis of aerosol mass spectra using a new interface in IGOR Pro

28-March-2013 5th AMS Clinic in Boulder
The bilinear model

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

\( \{n \times m\} \)

\( \{n \times p\} \)

\( \{p \times m\} \)

\{n \times m\}

- The rows of the matrix F represent the profiles (factors) of these linear combinations
- The columns of the matrix G represent the time series of the linear combinations

Paatero and Tapper 1994
PMF evaluates the matrices $G$ and $F$ without any a priori information

$$\hat{X} = G \cdot F$$

If the profiles ($F$) are provided and only the time series are modeled, it is in the spirit of the chemical mass balance (CMB)

Paatero 1999, Paatero and Hopke, 2008
Rotations in the bilinear model:

- Multiplication of \( G \) by a factor and \( F \) by its inverse; product of \( GF \) does not change (pure rotation)

\[
\bar{G} = GT \quad \text{and} \quad \bar{F} = T^{-1}F
\]

- ME-2 also allows approximate rotations (limited changes in \( GF \) and \( Q \)).

How to explore?

- PMF run
  - unrotated solution
  - rotated solution with, global fpeak (\( \phi \))

\[
T_{fpeak, p=3} = \begin{bmatrix}
1 & \phi & \phi \\
\phi & 1 & \phi \\
\phi & \phi & 1
\end{bmatrix}
\]
Rotations in the bilinear model:

- Multiplication of $G$ by a factor and $F$ by its inverse; product of $GF$ does not change (pure rotation)

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

- ME-2 also allows approximate rotations (limited changes in $GF$ and $Q$).

How to explore?

- **ME-2 run**
  - unrotated solution
  - rotated solution with, global fpeak ($\phi$)
  - rotated solution with individual fpeak matrix ($\phi$)

$$T_{p=3} = \begin{bmatrix} 1 & 0 & \phi \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, T^{-1}_{p=3} = \begin{bmatrix} 1 & 0 & -\phi \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
General approach: use *a priori* information to constrain rotations.

→ Constrain factor profiles and/or time series to yield environmentally-meaningful results

Two implementations in ME-2 GUI: *a*-value and pulling equations

**a-value**

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

and/or

\[ g_{i, \text{solution}} = g_i \pm a \cdot g_i \]

*f* and *g* are anchor profiles/time series

*a* gives tightness of constraint (0 to 1)

**Pulling equations**

\[
\begin{align*}
\arg \min_{G,F} (Q) &= \arg \min_{G,F} (Q^m + Q^{aux}) \\
Q^{aux} &= \sum_{k=1}^{K} \left( \frac{r_k}{S_k} \right)^2 \\
r &= \text{difference between anchor and actual values} \\
S &= \text{softness of pull}
\end{align*}
\]
Rotational control in ME-2

**a-value**

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

and/or

\[ g_{i, \text{solution}} = g_i \pm a \cdot g_i \]

\( a \) gives tightness of constraint (0 to 1)

\( f \) and \( g \) are anchor profiles/time series

**Pulling equations**

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

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

\( r \) = difference between anchor and actual values

\( k \) = softness of pull

---

<table>
<thead>
<tr>
<th><img src="diagram.png" alt="Diagram" /></th>
<th><img src="diagram.png" alt="Diagram" /></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchor</td>
<td>Distance from anchor</td>
</tr>
<tr>
<td>~~~Effect on Q</td>
<td>~~~Distance from anchor</td>
</tr>
<tr>
<td>Distance from anchor</td>
<td></td>
</tr>
</tbody>
</table>
Implementation: $a$-value

Common to all ME-2 implementations

Option to load reference data as anchor

Can constrain some factors, leave others free

$a$-values can be constant across a factor or variable-dependent
Implementation: pulling

Option to load reference data as anchor

Can constrain some factors, leave others free

\((dQ\text{ parameter determines the maximum effect of } Q^{\text{aux}} \text{ on the total } Q)\)

Pulling equations can be constant across a factor or variable-dependent
Implementation: PMF and viewing results
Case Study: Zürich winter

PMF solution yields mixed factors

→ m/z 60 split between COA, HOA, and BBOA

→ strong m/z 44 contribution to HOA

→ mixing in COA/HOA diurnals

ME-2 approaches

→ CMB (all 5 factors fixed)

→ a-value and pulling approaches with primary factors (COA, HOA, BBOA) anchored

*Canonaco et al., in preparation*
Overview of solutions

PMF – 5 factors free

CMB – 5 factors fixed (HOA, COA, BBOA, LV-OOA, SV-OOA)

Pulling and a-value: anchored HOA, COA, and BBOA; others free

*Constraints applied to factor profiles ONLY; time series are completely free*

Much higher Q for CMB, not much difference between the others

Q very slightly increases with increasing constraint (approximate rotations...)

*Canonaco et al., in preparation*
Constrained solutions outperform both CMB and (unconstrained) PMF.

A range of environmentally-reasonable solutions is found, having similarly-good correlations with external data

Canonaco et al., in preparation
Explained variation and apportioned mass

CMB again sticks out: high unexplained variation, low apportioned mass

Range of results obtained for “good” solutions
→ can evaluate uncertainty of model solution

Canonaco et al., in preparation
Not much variation in primary factor profiles (mostly because of constraints)

Largest changes in SV-OOA (esp. m/z 43)

Does NOT imply similar variations in the factor time series... Canonaco et al., in preparation
Accepted solutions: factor time series (diurnals)

BBOA and SV-OOA show greatest variation (most model uncertainty).

→ Effect of specific anchor spectrum (BBOA)?

→ Distinction between SV-OOA and LV-OOA is probably largely arbitrary

Canonaco et al., in preparation
New IGOR interface developed for implementation and analysis of the bilinear model using the ME-2 engine

- Reduce rotational ambiguity and optimize solutions via anchoring of factor profiles and/or time series
  - Implementations: a-value (hard) and pulling equations (soft)
- Case study on ACSM data from Zürich (winter 2011 and 2012)
  - Anchored solutions outperform unconstrained PMF
    - Reduced mixing in factor profiles
    - Better correlations with external data
    - Total explained variation comparable to PMF
- Frequently obtain range of acceptable solutions → allows estimation of model uncertainties
Source apportionment

a) Unconstrained run (PMF)

b) Constrain the primary sources (HOA, BBOA, COA) within the a-value or pulling approach and resolve the secondary sources (LV-OOA, SV-OOA) freely

c) Compare the change of the model (residuals, weighted residuals, residual/uncertainty image plot, explained/unexplained variation, correlations with external time series, diurnal, etc.)

d) Do a sensitivity test for the a-value or pulling parameters and report the range of reasonable solutions (model uncertainty)

e) Alternatively to point d) various factor profiles might also be tested
The validation of the model depends on

- residual ($E$) and weighted residual ($Q$) analysis
  
  not the absolute $Q$ is relevant!!! (uncertainties not well known)
  
  the change of $Q$ compared to various model runs should be monitored
  
  ($Paatero$ and Hopke, 2008)
  
  $Q$ depends on size of data and on number of factors

- normalize by $Q_{exp} = (n \cdot m - p \cdot n - p \cdot m)$ and monitor $Q / Q_{exp}$

- For AMS / ACSM data empirically $Q / Q_{exp} \sim 1$

- explained / unexplained variation (model variation scaled by measured variation) ideally $\sim 1$

- (comparison with external mass spectra), only for PMF

- comparison with co-located measurements

- diurnal cycle
Least-squares algorithm (Conjugate gradient)

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

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

\[ x_{ij} = \sum_{p=1}^{P} g_{ip} f_{pj} + e_{ij} \]

\[ e_{ij}: \text{measured} - \text{model} \]
\[ \sigma_{ij}: \text{uncertainty (measure, model)} \]

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

\[ a_k = \overline{f_k} + r_k \]

\[ a_k: \text{anchor for } k\text{-th model position in } G \text{ or } F \]
\[ f_k: \text{ } k\text{-th iterative model value} \]
\[ r_k: \text{ } k\text{-th difference} \]
\[ s_k: \text{ } k\text{-th weight} \]

Paatero 1999, Paatero and Hopke, 2008
GUI: general results
GUI: global residual plots
GUI: residuals for single variable

green lines: 25-75 percentiles; residual = (measured - model)
GUI: factor profile analysis

Apportionment of variables

Correlation matrix vs. reference spectra

Individual profile comparisons
GUI: time series comparisons

Apportionment summary

Correlation matrix vs. external data

Comparison of individual time series
GUI: diurnals
GUI: compare solutions

Show all solutions

Show means and spread
GUI: compare solutions (explained variation & residuals)
GUI: compare solutions (correlation with external data)

![Graph showing correlation values]