Updates to the PET

(PMF Evaluation Tool)

Ingrid Ulbrich

2 Oct 2011

12th Users Meeting, Orlando
**What do I do, and in what order?**


<table>
<thead>
<tr>
<th>Steps for preparing and choosing the best solution from PMF analysis of AMS datasets.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Matrix</strong></td>
</tr>
<tr>
<td>1. Calculate data and error matrices</td>
</tr>
<tr>
<td>2. Further data and error treatments</td>
</tr>
<tr>
<td>2a. Apply minimum error</td>
</tr>
<tr>
<td>2b. Remove anomalous spikes, if desired</td>
</tr>
<tr>
<td>2c. Smooth data, if desired</td>
</tr>
<tr>
<td>2d. Downweight low-SNR data</td>
</tr>
<tr>
<td>2e. Downweight repeated information (m/z 44 and related m/z’s)</td>
</tr>
<tr>
<td>3. Run PMF for a range of number of factors ($P$) and random starts (SEEDs). Examine $Q/Q_{\text{max}}$ vs. number of factors in solution (Fig. 5a). A steep change in slope indicates the minimum number of factors to consider for a good solution.</td>
</tr>
</tbody>
</table>
Panel for Error Prep Steps

1. Isotope removal for IR
2. Remove ∞ cols (SqPrint)
3. Don't have peaks by test (CO2, O2, etc.)
4. Check for StdDev Value < 0
5. Delete bad SRM m/z's (col. tv = hit (don't reset) opt for bad (don't reset or delete)
6. Check for m/z's with 1 max < 0
7. Check for runs with low signal < 0

XO 2D constr (α, β)
in ME2

10 2D constr (α, β)

Panel for Error Prep Steps

1. Isotope removal for IR
2. Remove ∞ cols (SqPrint)
3. Don't have peaks by test (CO2, O2, etc.)
4. Check for StdDev Value < 0
5. Delete bad SRM m/z's (col. tv = hit (don't reset) opt for bad (don't reset or delete)
6. Check for m/z's with 1 max < 0
7. Check for runs with low signal < 0

XO 2D constr (α, β)
in ME2

Panel for Error Prep Steps

1. Isotope removal for IR
2. Remove ∞ cols (SqPrint)
3. Don't have peaks by test (CO2, O2, etc.)
4. Check for StdDev Value < 0
5. Delete bad SRM m/z's (col. tv = hit (don't reset) opt for bad (don't reset or delete)
6. Check for m/z's with 1 max < 0
7. Check for runs with low signal < 0

XO 2D constr (α, β)
in ME2

- Make things not TD
- Check std dev (col) vals > 0

- Bootstrap using EDA method for matching factors
- Load data for each iteration ML?
- Add Total to to modified Total Resid plot

- Scatter panel streamline (draw BES)
- Scatter panel color-by 3rd cov

- Scatter plot all of factors vs each other in same sub (manually/say code?)

- Find for truncating long WNs

- Multithreading BES
- Free space
- Use index names (e.g., in Reid's lab)

- Gen assign Dim labels (e.g., List, label list, dim list)

- Binary search interp for Select Species

- Way to name factors, color same in all 5 plots? (hard to implement)

- Sort by panel w/ cyt. Factor (Feb 25)
What do I do, and in what order?


Table 1. Steps for preparing and choosing the best solution from PMF analysis of AMS datasets.

<table>
<thead>
<tr>
<th>Step</th>
<th>Data Matrix</th>
<th>Uncertainty Matrix</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Calculate data and error matrices</td>
<td>X</td>
<td>X</td>
<td>Squirrel/Pika/ACSM</td>
</tr>
<tr>
<td>2. Further data and error treatments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a. Apply minimum error</td>
<td></td>
<td></td>
<td>PET Step 1</td>
</tr>
<tr>
<td>2b. Remove anomalous spikes, if desired</td>
<td>X</td>
<td>X</td>
<td>“prep” tab</td>
</tr>
<tr>
<td>2c. Smooth data, if desired</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>2d. Downweight low-SNR data</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>2e. Downweight repeated information (m/z 44 and related m/z’s)</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>3. Run PMF for a range of number of factors (P) and random starts (SEEDs). Examine $Q/Q_{\text{max}}$ vs. number of factors in solution (Fig. 5a). A steep change in slope indicates the minimum number of factors to consider for a good solution.</td>
<td></td>
<td></td>
<td>PET Step 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>“run” tab</td>
</tr>
</tbody>
</table>
Error Prep in Execution Panel

2. Further data and error treatments
   2a. Apply minimum error
   2b. Remove anomalous spikes, if desired
   2c. Smooth data, if desired
   2d. Downweight low-SNR data
   2e. Downweight repeated information (m/z 44 and related m/z’s)

Will work for HR fragments
Any drives for PMF and experiment

- No longer need to be on same drive
Advanced Controls in Execution Panel
What, you don’t want to carry a 17” laptop?
New Time Averages for Time Series
HR features for spectra

- Stack and color by family
- Elemental ratios
Label factors in each solution, then sort
New graph capability for Scatter Panel

Color correlations by a different variable
Some behind-the-scenes work to improve speed

- Streamlining panel updates to only calculate once, reduce unnecessary calculations
- All time average (diurnal, month, etc.) and scaled residual box/whisker values calculated after execution and stored
Panel for Error Prep steps (Tab of Step 1 panel)

- Include removal for EIR
- Remove cols. (Sq/Rh)
- Don't highlight peaks by text (C0, C0, C0)
- Check for StdDev value < 0
- Delete bad NMR (col. t), if (data exist) opt for bad (chart exist or delete)
- check for mis. with 1 miss < 0?
- Check for runs with/without signal < 0?

No 2D constr (α, β)

5. Can num rows, num cols, max P be global vars? Does it matter? [BTS]

1. Streamline main panel update - usually calc + redraw twice

2. Updated Basic Classes version (2)
   - Initiate unique name

3. Calc uncond. & display check?
   - Delete NMR takes optional t-value

2. Advanced tab (see 3D)

- Button to export peak to AIMS4 (code in Sp)

1. Elemental analysis of HMS (code exists from Donna)

2. Color by MS Family (code exists from Donna)

- Scale read box, whiskers
  - Calc divalued value at end of execution & store (code partial)
  - Usually for PDA, day of week, weekday/day, Income/median, check/hide only?

2. Give speak check hide min.

3. Func for Axis height [BTS]

4. Func for delete - log. but files

- Binary search; interp for selecting species

4. Way to name factors, color same in all plots? (hard to implement)

1. Sort by panel w/ cyt. Factors (from 2D)
Longer-term Developments (next year)

- “CMB hybrid” approach to constraining factor mass spectra
  - Lanz et al., ES&T 2008 (2D, $\alpha$ parameter)
  - Ulbrich et al., AMTD 2011 (3D, $\beta$ parameter)