Downweighting “weak” and duplicate $m/z$’s in PMF Analysis

AMS Clinic
23 February 2010
Ingrid Ulbrich

Error calculations are our best estimate of uncertainty...

…and there are some $m/z$’s we know better than others! (They have better SNR.)
Low $m/z$'s generally have better SNR than high $m/z$'s

SNR < 2 is “weak”: signal $\approx$ noise

In this dataset, most $m/z$'s > 180 are “weak”
Why care about “weak” m/z’s?

• Paatero and Hopke (2003) show that inclusion of “weak” variables (m/z’s) can obscure small factors in the fit.

• But they may still contain information, so we’d like to include them in the fit...

... but count them less than the better points

*Paatero & Hopke, Discarding or downweighting high-noise variables in factor analytic models, Analytica Chimica Acta, 2003.

Handling “bad” and “weak” m/z’s

• Recall that the PMF fit is error-weighted

\[
Q = \sum_{i=rows} \sum_{j=columns} \left( \frac{\text{residual}_{ij}}{\text{error}_{ij}} \right)^2
\]

→ Increasing the error would decrease the Q-contribution or “downweight” the influence of these points

Downweighting “bad” and “weak” m/z’s

0.2 < SNR < 2 “weak” Downweight (x 2-3)

SNR < 0.2 “bad” Remove or
strongly downweight (x 5-10)
Downweighting “weak” m/z’s

- Calculate SNR with function `pmf_err_SNRwv`
  - Check the graph for “bad” m/z’s and remove them if necessary

- Downweight “weak” m/z’s with function `pmf_err_DwntWeakColumns`

* Functions in `pmf_errPrep_AMS_v2_3A.ipf*

---

m/z’s directly proportional to m/z 44

- PMF fits correlated patterns and doesn’t care about relative magnitudes

<table>
<thead>
<tr>
<th>m/z</th>
<th>frag_organic</th>
<th>New (frag_organic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td><strong>1*frag_organic[44]</strong></td>
<td><strong>0.225*frag_organic[44]</strong></td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>frag_organic[44]</td>
</tr>
</tbody>
</table>

* Aiken et al., ES&T 2008*

These m/z’s (and m/z 44) repeat the information of m/z 44 many times

- Same as putting 6 (or 7) copies of m/z 44 in the matrix
- Tells PMF that fitting m/z 44 is really (6x) important
- But m/z 44 isn’t really (6x) important
Prefer to weight $m/z$ 44 only once

- Could remove $m/z$’s 16-20, 28 before running PMF and insert them again afterwards
  
  or

- Downweight these $m/z$’s so that they contribute the $m/z$ 44 signal only once

(Sally Ng calculation, Supp. Info. of Ulbrich et al., ACP 2009)

* I remove $m/z$’s 19 and 20 because they have negligible mass.

Important: If you change the standard frag table, you may need to do this for other $m/z$’s.

---

Downweighting $m/z$ 44-related $m/z$’s

- Make a wave that has the $m/z$’s you want to downweight
  - E.g. `make/N=4 mz44peaksWv={16,17,18,44}`

- Use function `pmf_err_dnwt44peaks` to reweight these $m/z$’s

* Function in pmf_errPrep_AMS_v2_3A.ipf
What’s being shown in the panel?

Only place we tell PET about errors (uses the fully-prepared version)

... so Q here is based on the downweighted version

Should correct this for publications!

• `pmf_calc_Qmx_noSNRdnwt(dataMx, stdDevMx, SNRweightWv)`

• `SNRweightWv` must be made manually, based on `SNRwv` created by `pmf_err_SNRwv`
  – Recommend:
    ```
    make/O/N=dimsrize(noNaNs_amus)
    SNRweightWv = SNRwv<= 2 ? weightVal : 1
    ```
  And manually change m/z 44-related points in `SNRweightWv`