Defining Errors for HR-ToF-PMF

Manjula Canagaratna

Some slides from James Allan, Ingrid Ulbrich, Aki-Matti Kortelainen
OA Components: Multivariate Analyses

\[ \text{ORG} = C \times \text{MS} + E \]

Mass Spectrum

Mass Spectrum, Component 1

Mass Spectrum, Component 2

Mass Spectrum, Component 3

Contribution (\(\text{ug/m}^3\)), Component 1

Contribution (\(\text{ug/m}^3\)), Component 2

Contribution (\(\text{ug/m}^3\)), Component 3

...
Why are errors important for PMF

- They prescribe relative importance of datapoint to a fit.

Mountain or molehill?
What do you think of when you hear….

“sunglasses”

- Beach
- Beautiful Orlando
Why are errors important for PMF?

Each datapoint is weighted in fit according to its error

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

PMF minimizes Q
Ideal fit: $\text{residual}_{ij} = \text{error}_{ij}$

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

$$Q_{\text{exp}} = N$$

$$\frac{Q}{Q_{\text{exp}}} = 1$$
Precision vs accuracy

What do we use as the ‘error’ for PMF?
What errors matters for PMF?

A) Precision

B) Accuracy

C) Both

D) Depends
\[ C_{j,t} \text{(ugm}^3) = \frac{10^{12} \cdot MW_{NO_3}}{CE_s RIE_s I_{NO_3} QNA} \quad I_{j,t} \text{(Hz)} \]

**Accuracy**
(Unit conversions, AB corrections, Transmissin Efficiency Corrections, Flow rate Corrections)
- Factors that multiply Entire Row or Column
- Don’t affect temporal correlations of m/zs

Change Factor MS? NO
Change Factor time trend? NO
(only differ by multiplier)

**Precision**
(Uncertainty in the Ion Count rate)
- m/z dependent
- Individual points affected differently

Change Factor MS? YES
Change Factor time trend? YES
Sources of Precision Error

1) Counting Error - Number of ions you generate for a given number of parent molecules is variable (POISSON STATISTICS)

2) Ion to Ion Variability - MCP response to ions is variable (Width of single ion histogram measurements)

3) Electronic Noise –provides minimum error, decreases with sqrt(t)
Other Errors Particular to ToF systems

- Effect of Baseline Subtraction
- m/z calibration shifts, peak width uncertainties
- Peak Fitting Errors (Effect of neighbor intensity and spectral distance on)

How good are the errors output by current PIKA program? Donna has new diagnostic to look at this. (See Jill’s presentation)
Diagnostics in PMF Tool

- Q/Q_{exp} contribution as a function of time
- Q/Q_{exp} contribution as a function of m/z

\[
Q = \sum_{i=\text{rows}} \sum_{j=\text{columns}} \left( \frac{\text{residual}_{ij}}{\text{error}_{ij}} \right)^2
\]
Q/Q\text{exp} contribution as a function of m/z

Why are the scaled residuals so high?

72.021 C\text{3H4O2}
78.011 C\text{5H2O}
84.058 C\text{5H8O}
98.073 C\text{6H10O}
Why are scaled residuals high

- Need more factors?
- Do these ions have any significant information?
  - Look at the time series of ions in PMF panel
  - Check HR fits to make sure they are real and not tails
Issues to think about

• Do these ions have any significant information?
  - Look at the time series of ions in PMF panel
  - Check HR fits to make sure they are real and not tails
Both have their own contribution to signal
Signal is probably just a tail of C6H6
BASELINE IS WRONG!

98.073 C6H10O