Igor-PMF
Overview and Tutorial

Ingrid Ulbrich*,†
9th AMS Users Meeting
University of Manchester

*With much programming by the fabulous, the Sueper, Donna
†With many thanks to members of the Jimenez Group, Manjula, Sally, and Qi for their beta testing and suggestions

---

**Tutorial Outline**

PMF Intro, Run PMF \( (60\ min.) \)
- Resources for Help!
- What PMF does and how it does it
- Our Software Approach
- Set Up our Sample Case

**LUNCH** \( (60\ min.) \)

II. Viewing Results \( (60\ min.) \)
- Do the factors make physical sense?
- Do the factors give a satisfactory fit to the data?

PMF Real Case \( (10\ min.) \)

III. Compare with External Tracers \( (30\ min.) \)
- Are the factors consistent with tracers?
A word about learning…

- First introduction to material \textit{(lecture)}
- Time to digest and practice \textit{(homework)}
- Review of material \textit{(next lecture)}
  - Use of material to build next bit of knowledge
- Time to digest and practice \textit{(more homework)}

\textit{Present, repeat, repeat, repeat…}

A word about factor analytical methods…

- PMF is complicated, but it’s not enough on its own, or in all situations!
- More complicated situations require more complicated methods (See Lanz et al. ES&T 2008)
- \textit{The Fine Print}
  - This software is described in Ulbrich et al. ACPD 2008. Please cite this work (or the ACP version) when publishing results using this software. This software does not incorporate every PMF feature. This open source software can be tweaked at your will. This software is shared freely with you so that you may also advance your cause for world domination. factorization. I hope you get more than you pay for. No critters (squirrel, pika, ape) were harmed in the making of this software.
Where to go for More!

- All Panel Documentation
  http://cires.colorado.edu/jimenez-group/wiki/
  index.php?title=PMF-AMS_Analysis_Guide
  *(under development)*

- Software and Sample Data
  http://cires.colorado.edu/jimenez-group/PMFResources/
  *(also under development)*

**Generic Receptor Model Schematic**

A data matrix is decomposed into an **arbitrary number** of factors, each of which is represented by a **constant mass spectrum** and a contribution **time series**. There is usually some **residual** of fit.
The output will look like this…

… and you can look at…

View Other Solutions

Total Quality of Fit

Factors

Residuals
How does PMF solve for the factors?

\[ \text{Time} \times \text{Contribution (µg/m}^3\), Component 1 \times \text{Constant Profile, Component 1} + \text{Contribution (µg/m}^3\), Component 2 \times \text{Constant Profile, Component 2} + \ldots \]

Least-Squares Fit

Have some points…

… and fit a line…

\[ \min \sum_i^1 \sum_i^2 \]
Least-Squares Fit, Quantified

Could choose a different line that has a different fit…

The PMF Approach: Weighted Least Squares

• in many dimensions, called

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

estimated error of each point

• If all points are fit as well as we know them, the \( Q \)-contribution from each point is \( \sim 1 \)

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

and the total \( Q \) is \( \sim \) the size of the matrix.
The PMF Approach: Weighted Least Squares

- Let’s normalize the total $Q$ to $Q_{\text{exp}}$, then if all points are fit to $\sim \sigma$, $Q/Q_{\text{exp}} \sim 1$

$$Q/Q_{\text{exp}} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left( \frac{\sim \text{resid}_{ij}}{\sigma_{ij}} \right)^2 / m \times n$$

- Adding factors decreases $Q$, but may not give meaningful factors!

... and you can look at...

Q vs. # of Factors
PMF2 solution space seen in 2 dimensions

Start w/ random values

Each possible solution has a quality-of-fit $Q (\chi^2)$ – minimize $Q$ during iterative fit

Factor Space in 2 dimensions

FPEAK explores a subset of the possible rotations

One way to consider rotations of solutions…

If you only had two species and were trying to reconstruct some data…

Any of these pairs could be used.

Fig. 1. Simulated data showing multiple possible source “profiles” that could be used to fit the data.

Paatero et al., Chemometrics and Intelligent Lab Systems, 2002
PMF2 solution space seen in 2 dimensions

Start w/ random values

Factor Space in 2 dimensions

Each possible solution has a quality-of-fit $Q (\chi^2)$ – minimize $Q$ during iterative fit

Start w/ many random values

SEEDs

FPEAK explores a subset of the possible rotations

Adapted from Penti Paatero
2 Ways to Explore PMF2 Solutions for a certain number of factors

Each possible solution has a quality-of-fit $Q (\chi^2)$ – minimize $Q$ during iterative fit

Run a range of SEEDs

Run a range of FPEAKs

USER RESPONSIBILITY TO DO BOTH!

Factor Space in 2 dimensions

Tutorial Outline

PMF Intro, Run PMF (60 min.)
- Resources for Help!
- What PMF does and how it does it
- Our Software Approach
- Set Up our Sample Case

LUNCH (60 min.)

II. Viewing Results (60 min.)
- Do the factors make physical sense?
- Do the factors give a satisfactory fit to the data?

PMF Real Case (10 min.)

III. Compare with External Tracers (30 min.)
- Do the factors fit with what we already know?
Software Approach

0. Prepare your data and error matrices and remove empty rows, columns (Wiki!)
   • Run all combinations of ranges of
     \[ \text{FPEAKs} \quad \text{or} \quad \text{SEEDs} \]
     # of Factors \( \begin{cases} \text{and} & \text{FPEAKs} \\ \text{or} & \text{SEEDs} \end{cases} \)
   • Wait…. and wait… and maybe come back tomorrow…
   • View output and examine solutions

Let’s set up our sample case!

1. Put \texttt{mypmft.ini} in your PMF Executable Folder \( \text{(only the first time you run the Panel)} \)
   Must include \texttt{mypmft.ini} \texttt{pmf2.key} \texttt{Pmf2wtst.exe}

2. Open the Tutorial Template

3. Recommend Organizing your Directories:
   + root
     + DataTemplate \( \rightarrow \text{to copy for new versions} \)
     + DataForPMF\_v1 \( \rightarrow \text{e.g., your basis case} \)
     + DataForPMF\_v2 \( \rightarrow \text{e.g., with different error est.} \)
     + External\_MassSpectra \( \rightarrow \text{for use with the Scatter Panel} \)
     + External\_Tseries

** Data for PMF can only be in root: or a subdirectory of root:, not in any lower levels! **
Sample Case, cont.

4. Load and compile the 3 ipfs
   - PMF_Execution_v2_0.ipf
   - PMF_ViewResults_v2_0.ipf
   - PMF_Scatter_v2_0.ipf

5. From the new “PMF” pull-down menu, choose

   Perform PMF Analysis *Step 1*

The PMF Execution Panel *Step 1*

1. Choose the directory where you have the PMF Executable
1a. Check the DataFolder Result

The PMF Execution Panel *Step 1*

1. Choose the directory where you have the PMF Executable
2. Choose the folder where you have the Data and Error matrices
The PMF Execution Panel *Step 1*

1. Choose the directory where you have the PMF Executable
2. Choose the folder where you have the Data and Error matrices

---

The PMF Execution Panel *Step 1*

1. Choose the directory where you have the PMF Executable
2. Choose the folder where you have the Data and Error Matrices and choose those matrices

Read about Model Error in the Wiki!
The PMF Execution Panel *Step 1*

1. Choose the directory where you have the PMF Executable
2. Choose the folder where you have the Data and Error Matrices and choose those matrices
   Read about Model Error in the Wiki!
3. Choose the Number of Factors (we like to include 1 factor)

Recall: Have a Choice of FPEAKs or SEEDs

Each possible solution has a quality-of-fit Q ($\chi^2$) – minimize Q during iterative fit

Run a range of FPEAKs

Run a range of SEEDs

Factor Space in 2 dimensions

USER RESPONSIBILITY TO DO BOTH!

... but you can only do one at a time

Adapted from Pentti Paatero
The PMF Execution Panel *Step 1*

1. Choose the directory where you have the PMF Executable

2. Choose the folder where you have the Data and Error Matrices *and choose those matrices*

   *Read about Model Error in the Wiki!*

3. Choose the Number of Factors *(we like to include 1 factor)*

4. Choose FPEAKs or Seeds
   - Calculated Range *or* Provide a wave -1.5 to 1.5 with delta 0.5

5. Press the button!
Last iteration saved in Log file
Data was read properly
Iterations calculating

Advanced Topics
The Complete $Q$ Expression

\[ Q(E, G, F) = \sum_{i=1}^{m} \sum_{j=1}^{n} (E_{ij}/\sigma_{ij})^2 - \alpha \sum_{i=1}^{m} \sum_{h=1}^{p} \log G_{ih} - \beta \sum_{h=1}^{p} \sum_{j=1}^{n} \log F_{hj} + \gamma \sum_{i=1}^{m} \sum_{h=1}^{p} G_{ih}^2 + \delta \sum_{h=1}^{p} \sum_{j=1}^{n} F_{hj}^2. \]  

(6)


Execution Panel: “Model Error”

- PMF recalculates the provided error values as

\[ Error'_{ij} = Error_{ij} + ModelError \times Data_{ij} \]

- Igor-PMF software writes a file `STD_DEV_PROP.DAT` that is used by PMF and has all elements = ModelError
Running in non-Robust Mode

• Need to change the flag in mypmft.ini

• You may want to make a new PMF folder that contains this .ini so that you don’t forget which mode you’re running in.

Running a range of FPEAKs at non-0 Seed

• Currently not possible from the panel directly

• Can change the default seed in the Execution ipf