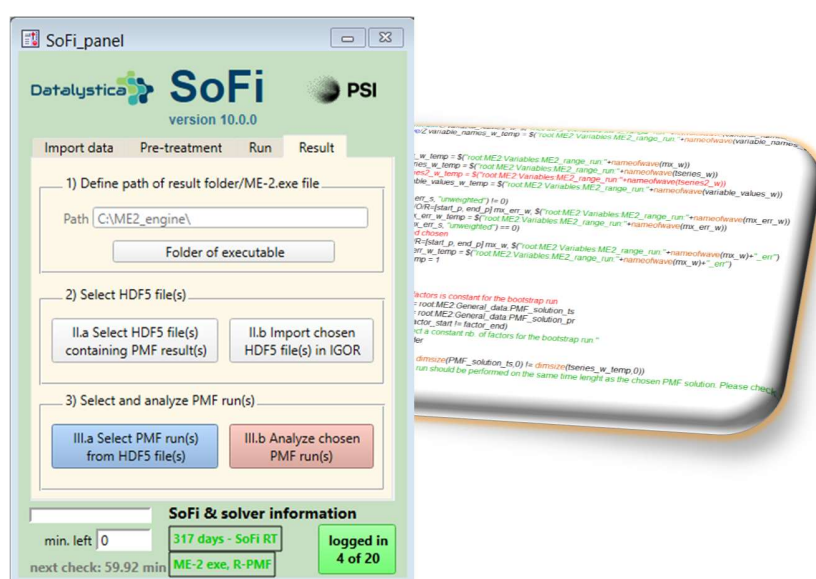


Source Finder (SoFi)

基于 IGOR 平台的 SoFi Pro 及 SoFi RT 用户手册



作者

Datalystica 有限公司

地址：瑞士 Villigen,

Parkstrasse 1, 5234

最近更新于:

8/21/2026

本手册适用于此版本:

SoFi 10.0.1

目录

1.	关于	9
1.1.	什么是 SoFi	9
1.2.	适用对象	9
1.3.	软件及个人电脑权限需求	9
1.4.	关于 SoFi 结构的更多细节	9
1.5.	SoFi 版本之间的差异	10
2.	PMF 理论	11
2.1.	双线性模型 – 概述	11
2.1.1.	一般双线性方程	11
2.1.2.	Robust 模式	12
2.2.	通过 Q 的期望值 (Q_{exp}) 对 Q 进行归一化	12
2.3.	模型解的旋转多义性	12
2.4.	ME-2 中的控制旋转多义性的技术	13
2.4.1.	初始种子运行	13
2.4.2.	全局 fpeak.....	13
2.4.3.	个体 fpeak.....	13
2.4.4.	a-值方法: 部分约束矩阵 F 和/或 G	14
2.4.5.	拉动方法: 部分约束矩阵 F 和/或 G	14
3.	概述	16
3.1.	HDF5 接口的安装 (Igor 8)	16
3.2.	SoFi 整体结构	16
3.3.	第一步	17
3.4.	SoFi 菜单	17
3.4.1.	初始化 SoFi	17
3.4.2.	删除未使用的数据 (例如, 已关闭的图表)	18
3.4.3.	重新创建面板, 即关闭图形、表格, 仅面板	18

3.4.4.	重新创建设置，除了文件夹“External_data”和“Reference_ts/pr”	18
3.4.5.	重新创建 SoFi，即整个 SoFi 文件夹	18
3.4.6.	将 SoFi 面板置于顶部	18
3.4.7.	打开集群面板	18
3.4.8.	主设置	18
3.4.9.	自定义高分辨率 HR 数据的 family	22
3.4.10.	用户手册 (Manual)	23
3.4.11.	历史记录 (History)	24
3.4.12.	条款和条件(Terms and conditions)	24
3.4.13.	致谢 (Credits)	24
3.4.14.	许可证信息 (License information)	24
3.4.15.	关闭 SoFi 包/密钥程序 (Close SoFi package/key procedure)	24
3.4.16.	执行通信测试 (Perform communication tests)	24
3.5.	SoFi 密钥激活步骤	24
3.6.	IGOR 中的数据层次结构	25
3.7.	弹窗提醒更新 SoFi 版本	26
4.	导入并定义输入数据	28
4.1.	导入数据	28
4.1.1.	导入 PMF 输入文件	29
4.1.2.	时间序列导入	31
4.1.3.	导入源谱文件	34
4.1.4.	合并矩阵	34
4.2.	定义 PMF 输入	34
4.2.1.	数据和误差矩阵	35
4.2.2.	时间 IGOR wave	35
4.2.3.	变量向量	35
4.2.4.	数据类型	36

5. SoFi – 预处理	39
5.1. 处理 PMF 输入中的缺失数据	39
5.2. 处理外部数据	40
5.2.1. 源谱文件	40
5.2.2. 时间序列	41
5.3. 探索 PMF 输入变量	43
5.3.1. 变量的选项	43
5.3.2. 根据类别/时间进行选择	47
5.3.3. 图形表示的选项	48
5.4. 选择或修改有效的 PMF 输入	49
5.4.1. 类别	49
5.4.2. 输入数据	52
5.4.3. 基于 MDL 黑名单单个点	54
6. SoFi – PMF runs	56
6.1. 定义路径、可执行文件数量和 HDF5 文件名称	56
6.1.1. 路径	56
6.1.1. 可执行文件数量	56
6.1.2. HDF5 文件名	57
6.2. 设置一般模型选项	57
6.2.1. 设置因子数量	57
6.2.2. 缺失数据	58
6.2.3. 降低权重	58
6.2.4. 误差矩阵（一般）	60
6.2.5. 收敛标准 Convergence criteria	62
6.2.6. 预 PMF 评估 Pre-PMF Evaluation	63
6.2.7. 其他各种选项	64
6.3. 为模型矩阵定义选项	67
6.3.1. 种子 seed	67

6.3.2.	峰值 fpeak.....	69
6.3.3.	a 值约束.....	71
6.3.4.	位移 Displacement.....	75
6.3.5.	拉取方程 Pulling equations.....	75
6.3.1.	因子名称自定义 Factor specifications.....	79
6.3.2.	考虑旋转工具的优先级.....	81
6.4.	统计不确定性的估计	82
6.4.1.	Bootstrap 重采样策略	82
6.4.2.	输入扰动 Input perturbation.....	83
6.5.	滚动 rolling 机制.....	84
6.6.	调用 ME-2 并执行 PMF	85
7.	SoFi – PMF 结果	86
7.1.	导入结果文件（HDF 文件）	86
7.1.1.	定义 ME2.exe 文件的路径.....	86
7.1.2.	选择包含 PMF runs 的 HDF5 文件。.....	86
7.1.3.	将选定的 HDF5 文件导入 IGOR.....	87
7.2.	手动选择 HDF 文件中的 PMF runs	87
7.2.1.	设置.....	87
7.2.2.	选择 PMF runs choose PMF run(s)	88
7.3.	基于条件的选择 Criteria-based selection.....	89
7.3.1.	条件类型.....	90
7.3.2.	条件列表.....	91
7.3.3.	主动（排序）与被动（监控）条件.....	92
7.3.4.	条件语法.....	93
7.3.5.	如何设置条件列表的示例.....	96
7.3.6.	评估条件.....	104
7.3.7.	检查条件分数.....	105
7.3.8.	检查选定的 PMF runs	109

7.3.9.	检查条件的稳定性	110
7.3.10.	一个 HDF 文件中的不同因子数量	110
7.4.	文件设置和日志文件	113
7.5.	关于如何导入所选 PMF runs 的选项.....	113
7.5.1.	平均	113
7.5.2.	平均（分辨率）和数据选项	114
7.5.3.	自定义因子名称	115
7.6.	结果文件夹的技术信息	116
7.6.1.	结果文件夹的结构	116
7.6.2.	参数文件	117
7.6.3.	主要信息	118
8.	SoFi – PMF 分析	120
8.1.	选择一个 PMF run	120
8.2.	一般信息	121
8.2.1.	选择 PMF runs 以进行探索.....	121
8.2.2.	相邻列表中所有 PMF runs 的信息.....	121
8.2.3.	从相邻列表中选择的 PMF runs 的设置.....	121
8.2.4.	导出 PMF 的最终解.....	121
8.3.	子面板图形 Graph 和数据 Data	122
8.4.	概述图	123
8.4.1.	时间序列特定选项	123
8.4.2.	因子源谱特有选项	124
8.5.	详细图	125
8.5.1.	占比 fraction	125
8.5.2.	标准 Standard	126
8.5.3.	散点图 Scatter	128
8.5.4.	残差 Residual	131
8.5.5.	图像 Image	134

8.5.6.	HR 图 plot.....	140
8.5.7.	类别图 Category plot.....	143
8.6.	平均解 Averaged solution.....	144
8.6.1.	重复 Repeats.....	144
8.6.1.	质量误差估计 Mass error estimation.....	145
8.6.2.	使用的 α - 值.....	146
8.6.1.	t-谱 t-profiles.....	147
9.	独立分析.....	150
9.1.	高亮有趣的时间段.....	150
9.2.	独立 bootstrap 分析.....	150
9.3.	选框 Marquee 工具.....	153
9.4.	聚类面板 Cluster panel.....	158
9.4.1.	聚类选项 Cluster options.....	158
9.4.2.	标准图 Standard plot.....	160
9.4.3.	散点图 Scatter plot.....	161
9.4.4.	轮廓图 Silhouette plot.....	162
9.4.5.	标准选项 Criteria option.....	163
10.	SoFi 中的实时模块 (SoFi RT)	164
10.1.	关于.....	164
10.1.1.	SoFi RT 是什么.....	164
10.1.2.	工作原理.....	165
10.2.	配置文件 Configuration file.....	165
10.3.	手动操作的第一步.....	168
10.3.1.	启动 SoFi 并准备输入数据.....	168
10.3.2.	定义主要模型设置.....	171
10.3.3.	调用 SoFi RT.....	176
10.4.	SoFi RT 模块.....	176
10.4.1.	定义输入 INPUT 选项。.....	176

10.4.2.	定义模型选项。	177
10.4.3.	定义 RESULT 选项。	181
10.4.4.	报告 Report	182
10.5.	启动 SoFi RT	184
10.5.1.	开始/停止 SoFi RT	184
10.5.2.	SoFi RT 概览结果	185
10.6.	使用 SoFi Pro 检查 SoFi RT 结果	187
10.7.	ACSM 数据的建议	187
10.8.	从 AE33 生成每日 BC 文件	187
11.	定义和缩略语列表	189
12.	参考文献	190

1. 关于

1.1. 什么是 SoFi

源查找器 (SoFi) 是一个功能强大且用户友好的软件包，用于源解析和环境数据分析。该软件最初在保罗·谢尔研究所 (PSI) 开发，现在由瑞士维利根的 Datalystica 有限公司积极维护和进一步开发。SoFi 是一个在 IGOR Pro (Wavemetrics Inc., 美国) 软件环境中运行的软件包。SoFi 广泛应用于大气科学和环境化学，用于启动和控制多线性引擎算法 (ME-2, (Paatero, 1999)) 来解决正矩阵因子分解算法 (PMF, (Paatero, 1997))，并分析 PMF 结果的所有细节，以提取数学上和环境意义均合理的 PMF 结果。

1.2. 适用对象

SoFi 旨在简化和加速源解析分析。虽然 SoFi 最初是为分析气溶胶化学成分监测器 (ACSM) 数据而开发的 (Ng et al., 2011b)，但它的功能远不止于此。它几乎可以应用于任何对降维或因子分析感兴趣的数据集。

凭借其高度面板化和人性化的结构，SoFi 确保从数据输入到数据输出的丝滑源分配分析，几乎不需要与软件环境 IGOR 进行任何交互，使 SoFi 适合完全没有编程或 IGOR 知识技能的数据分析师。只要某类数据适合进行降维分析，就可以尝试应用该软件。SoFi 是 Datalystica Ltd. 的知识产权，仅可在 Datalystica 的条款和条件下使用，Datalystica 是 SoFi 的官方分销商（请查阅 Datalystica 的主页 (<https://datalystica.com>) 以获取条款和条件）。当 SoFi 启动时，用户必须阅读并同意相同的条款和条件，才能使用 SoFi 软件。

1.3. 软件及个人电脑权限需求

确保您已经拥有 ME-2 的官方密钥（必须从 Datalystica 的主页 (<https://datalystica.com>) 购买），您对您的个人电脑具有完全的管理员权限（文件读取和写入权限），因为 SoFi 将在操作系统上创建、修改和执行脚本文件，以及 IGOR 9 或 10 的官方许可证（可以通过我们与 SoFi Pro/RT 购买组合购买，或者也可直接从 Wavemetrics 官网 (<https://www.wavemetrics.com>) 购买）。

1.4. 关于 SoFi 结构的更多细节

SoFi 有两个版本，分别是 **SoFi Pro**（包含所有 SoFi 高级功能的商业密钥）和 **SoFi RT**（实时模块，提供完全自动化和实时源解析，同时包括所有 SoFi Pro 功能）。SoFi Pro 和 SoFi RT 密钥可以从 Datalystica 购买，网址为 <https://datalystica.com>。有关更多细节，请咨询 Datalystica 的主页。本手册涵盖了 SoFi 软件的所有功能。

本用户手册并未详细描述 PMF 算法背后的理论或 ME-2 求解器及其脚本代码，而是介绍上述软件的具体操作流程。然而，下一章中提供了关于 PMF 的理论简单介绍。感兴趣的读者可参考 ME-2 算法的原始论文 (Paatero, 1999), 介绍 PMF 算法的论文 (Paatero and Tapper, 1994; Paatero, 1997; Paatero, 2004), 提供 ME-2 算法不同控制旋转多义性工具概述的论文 (Paatero 和 Hopke, 2009), 以及关于 ME-2 的最新方法论文 (Brown et al., 2015; Paatero et al., 2014)。此外，读者还可就 Robust 模式参考 Paatero (1997), 关于如何平衡 PMF 的误差矩阵则参考此论文 (Paatero 和 Hopke, 2003)。

读者还可以参考 Canonaco 等人 (2013),(2015) 和 (2021) 分别介绍了 SoFi 的论文及其在 ACSM 数据上的应用，以及 Crippa 等人 (2014) 则介绍了 α 值方法在许多欧洲气溶胶质谱仪 (AMS) 案例研究中得到了充分测试，建立了典型的一次有机气溶胶 (primary organic aerosol, POA) 的平均质谱，并提出和讨论了一种使用 α 值方法处理 AMS 数据的通用策略。关于使用滚动 PMF (rolling PMF) 技术的欧洲 ACSM 数据集的类似研究由 (Chen 等人, 2022) 提出。

1.5. SoFi 版本之间的差异

目前在 IGOR 中有两个可操作的 SoFi 版本，即 SoFi 9 适用于 IGOR 9，SoFi 10 适用于 IGOR 10。本手册中包含的 SoFi 功能描述指的是最新版本，即 SoFi 10。这里介绍的许多功能在之前的 SoFi 版本中也可用。请注意，我们自 2021 年中期正式停止支持 SoFi 6，从那时起 SoFi 6 不再更新。我们计划在 2026 年某个时候停止对 SoFi 8 的支持。因此，我们强烈建议升级到 SoFi 9 或 SoFi 10。如果您仍在使用 SoFi 6 或 SoFi 8，请查阅历史文件 (<https://datalystica.com/downloads/>) 并确保您没有遭受 SoFi 6 中严重未解决的程序错误！

一般来说，Datalystica 无法百分百保证 SoFi 9 在 IGOR 9 中、SoFi 10 在 IGOR 10 中，或者其他交叉组合（例如在 IGOR 10 中运行 SoFi 9）在所有情况下都能完全无错误地运行。SoFi 软件不可避免会存在一些在我们的测试阶段未测试或未遇到的情况和条件，这可能会导致 SoFi 或 IGOR 出现一些意想不到的问题。对此我们深表歉意，我们极度依赖所有 SoFi 用户的报错与反馈，以此来尽量减少 bug 从而尽可能得改善 SoFi 以满足您的需求！无论是什么反馈，我们都会秉承认真并及时的态度来尽可能地提高您的用户体验！我们再次提前感谢您对 SoFi 提供的宝贵的指正与建议！

Datalystica 团队敬上！

2. PMF 理论

2.1. 双线性模型 – 概述

2.1.1. 一般双线性方程

一组可测量的随时间变化的变量（例如，ACSM 的 UMR m/z 强度）可以通过数据矩阵 X 来表达，其中每列 j 是可测量到的变量，每行 i 代表特定单独的时间点上测到的变量的信号或强度。

在所谓的因子分析模型中，为了降低维度，变量被分组为几个因子用以进一步研究，从而避免分析错综复杂的测量变量的时间序列。一种常用的方法是将变量分组为两个矩阵使用所谓的双线性模型。双线性模型的几个例子有主成分分析，PCA (Wold 等人, 1987), 化学质量平衡方法，CMB (Miller 等, 1972) 或正矩阵分解 PMF (Paatero 和 Tapper, 1994)。双线性因子模型以矩阵表示法定义为：

$$X = GF + E \quad (1)$$

其中测量矩阵 X 由矩阵 G 和 F 的乘积近似。参数 p ，即模型矩阵 G 的列数和模型矩阵 F 的行数，被称为因子数。矩阵 G 的每一列 j 代表每个 p 因子的时间序列，而 F 的每一行 i 代表相应的因子特征（例如，污染源的质谱）。注意，根据公式 (1) 的定义，因子特征是静态的，不允许随时间变化。矩阵 E 代表模型残差。PCA 和 PMF 之间的差异主要由于这些模型的限制。PCA 对因子特征施加正交性，即 F 的任意两行的标量为零，并且不要求非负条目。相反，PMF 要求 G 和 F 中的条目均为非负。PMF2 和 ME-2 是能够解决 PMF 算法的求解器。这两个求解器使用最小二乘算法拟合 G 和 F 中的条目，迭代最小化数量 Q_{main} , Q^m ，即所有输入点 ij 的模型残差 e_{ij} 除以相应不确定度 σ_{ij} 平方后的总和(2)：

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

在这里，不确定性 σ_{ij} 通常包含输入点 ij 的测量不确定性。对于比率 $e_{ij} / \sigma_{ij} \gg 1$ 的数据点，这些点会对 Q^m 产生较大贡献，因此这些点在模型迭代过程中会产生较大影响。通常，这一条件确保信噪比高的数据比接近检测限的测量数据获得更高权重。然而，对于某些数据点，当比率 $e_{ij} / \sigma_{ij} \gg 1$ 时，也可能由于局部事件或测量设备内的电子噪声占主导地位，而不应被模型考虑。

2.1.2. Robust 模式

为了防止最终结果受到这些事件的影响，SoFi 通常默认以“Robust”模式运行，在这种模式下，异常值对解的影响被减少。在迭代过程的每一步中，异常值是根据残差与不确定性的比率来定义的：

$$\text{outlier} := \left| \frac{e_{ij}}{\sigma_{ij}} \right| > \alpha \quad (3)$$

α 是用户定义的阈值。该文献建议将 4 作为 Robust 模式下异常值的定义标准 (Paatero, 1997)，因此 SoFi 也将 4 作为默认值。为消除 Q^m 的变化与异常值残差变化之间的依赖关系，残差被动态重新加权。

$$\frac{dQ^m}{dE_{\text{outliers}}} = \text{const.} \quad (4)$$

这防止了模型最终结果受到少数强信号点的影响。建议在处理易受例如稀有瞬态源、高强度点、电子噪声或低信噪比的数据时，始终以“稳健”模式运行模型，这通常适用于外场观测的环境数据。

2.2. 通过 Q 的期望值 (Q_{exp}) 对 Q 进行归一化

通常，监测和比较总计 Q 在各种 PMF runs 之间是没有意义的，因为期望值取决于数据矩阵的大小和所选因子的数量。因此，通常 Q^m 除以 Q_{exp} 来归一。其中 Q_{exp} 是剩余的自由度同时取决于数据矩阵的大小和因子数，并且也是因子数量的函数。

$$Q_{\text{exp}} = n \cdot m - p \cdot (n + m) \quad (5)$$

在过去，AMS 研究报告了 Q/Q_{exp} 比率的典型值在 1 到 5 之间。有时报告的值低于 1。请注意，当以稳健模式运行 PMF 时，稳健模式值为 4 时， Q/Q_{exp} 的最高值约为 16。对于理想的 PMF runs，只有在所采用的不确定性同时考虑了测量的不确定性和模型的不确定性时， Q/Q_{exp} 才会降低到约 1。然而，模型不确定度很难评估，尤其是对于环境数据，通常仍然未知。此外，降权程序（见 6.2.3 可能会影响 Q 的值，这取决于弱/坏变量的数量和大小。因此， Q/Q_{exp} 绝对值不能严格用作判断模型结果的指标。相反，应考察不同模型运行中 Q/Q_{exp} 的相对变化，以此作为选择合理模型解的指导。

2.3. 模型解的旋转多义性

一般双线性模型具有大量解，因为 G 和 F 可以分别通过矩阵 T 和逆矩阵 T^{-1} 进行变换（见公式(6)）。采用正交性或非负性等约束条件可以减少旋转的程度（也称为旋转多义性），但其影响仍然很强 (Paatero et al., 2002)。其中会发生两种不同类型的旋转，分别为纯旋转和近似旋转。对于纯旋转，目标函数 Q^m 在空间中的旋转前后保持不变：

$$\bar{\mathbf{G}} = \mathbf{G}\mathbf{T} \text{ and } \bar{\mathbf{F}} = \mathbf{T}^{-1}\mathbf{F} \quad (6)$$

其中 $\mathbf{T}$ 是一个维度为 $p \times p$ 的非奇异矩阵， $\mathbf{T}^{-1}$ 是其逆矩阵， $\bar{\mathbf{G}}$ 和 $\bar{\mathbf{F}}$ 是旋转后的矩阵。矩阵乘法 $\bar{\mathbf{G}}$ 和 $\bar{\mathbf{F}}$ 的乘积与 $\mathbf{G}$ 和 $\mathbf{F}$ 的乘积相同，因此 Q^m 保持不变。如果变换矩阵 $\mathbf{T}$ 不满足公式(6)，这种旋转则被称为近似旋转，在这种情况下， Q^m 会发生变化。

2.4. ME-2 中的控制旋转多义性的技术

2.4.1. 初始种子运行

用随机或伪随机数据对 PMF 模型进行初始化的过程称为初始种子运行，因为模型矩阵在此过程中被数据“播种”或填充。这种操作通常在寻找目标函数 Q 空间的全局最小值时进行。然而，由于初始种子运行拥有不同的起始点，且在数据具有高度旋转多义性的情况下，全局最小值非常容易出现，因此不同的初始种子运行最终会落在全局最小值的不同位置。因此，初始种子运行也可以对旋转多义性进行部分探索。但是，该技术无法进行系统性的探索，为了实现系统性探索的目的，建议使用下文中所述的其他技术之一。

2.4.2. 全局 fpeak

对于 PMF2 和 ME-2 引擎，有一个用户特定的参数称为 **fpeak**，用符号 ϕ 表示，用于对这种旋转进行全局控制。对于正的 ϕ ，模型会执行一次或一系列基本旋转，使矩阵 $\mathbf{G}$ 的列增大，同时使矩阵 $\mathbf{F}$ 的行减小，并在这一过程中保持质量守恒。对于负的 ϕ ，则发生相反的变化。然而，**fpeak** 工具仅在高维空间的一个维度中探索旋转，如果 $\mathbf{G}$ 和 $\mathbf{F}$ 的值为正且选择了多个因子，则旋转空间是多维的，相应的多义性可能非常大（例如，对于三个因子的模型，旋转空间是九维的）。

对于 PMF2 用户：请注意，ME-2 中的 **fpeak** 值与 PMF2 的 **fpeak** 成正比。它们不一定相等 (Paatero 和 Hopke, 2009)。

2.4.3. 个体 fpeak

与 PMF 求解器相比，ME-2 求解器的一个优势是改进的旋转控制，例如，选择的因子可以相加/相减，而不是转换整个矩阵。因此，可以在更受控的环境中探索旋转多义性。

分析师必须利用这些探索旋转多义性的工具在众多 PMF 结果中寻找符合环境自然规律的 PMF 结果。

旋转多义性可能会非常高，特别是对于包含许多变量、大数据集、长测量时间、高因子数量、低信噪比等情况，因此，引入先验信息到 PMF 模型中以减少旋转多义性是有利的。这可以以已知的 $\mathbf{F}$ 行（因子质谱）和/或已知的 $\mathbf{G}$ 列（因子时间序列）的形式出现 (Paatero 和 Hopke, 2009)。这种先验信息防止模型矩阵旋转，从而得到近乎唯一的 PMF 结果。在 ME-2

中主要通过两种不同的方法来实现，即 **a** 值方法及拉动技术，以下小节将简要讨论这些方法。

2.4.4. **a**-值方法：部分约束矩阵 **F** 和/或 **G**

在 **a**-值方法中，**F** 矩阵（因子特征）或 **G** 矩阵（因子时间序列）的元素可以由用户自行定义和约束。用户输入一个或多个因子特征（**F** 的行）/ 因子时间序列（**G** 的列）以及由标量 **a** 定义的约束，该约束可以应用于整个特征/时间序列或仅应用于特征/时间序列其中的某个元素。**a**-值决定了输出结果的 **F_{solution}** 或 **G_{solution}** 可以在多大程度上偏离输入的 **F_{initial}** 或 **G_{initial}**，依据：

$$f_{j, solution} = f_j \pm a \cdot f_j \quad (7)$$

$$g_{i, solution} = g_i \pm a \cdot g_i \quad (8)$$

其中 f_j 和 g_i 分别表示矩阵 **F** 和 **G** 的行和列。索引 j 在 0 和变量数量之间变化， i 在 0 和测量点数量之间变化。

标量 **a** 的范围在 0 和 1 之间。将所有因子特征的 **a**-值设置为零，可以完全约束因子特征矩阵 **F**，类似于化学物质平衡（CMB）的方法。相反，将特征或时间序列的所有条目的 **a**-值设置为 1，则允许其类似于无约束的情况。

2.4.5. 拉动方法：部分约束矩阵 **F** 和/或 **G**

用户可以将拉动方程引入模型，以将某些因子中元素拉向目标值，即所谓的锚点（此处仅针对矩阵 **F** 的行 j 为例）：

$$\overline{a_j} = \overline{f_j} + r_j \quad (9)$$

在方程(9) 中， $\overline{a_j}$ 代表模型必须拉动的迭代值的锚点 $\overline{f_j}$ ， r_{ij} 代表残差。锚点是用户作为先验信息引入的某个已知值。拉动方程创建一个额外的辅助项 Q^{aux} ，该项与 Q^m 分开添加。因此，如果引入拉动方程，模型将针对矩阵 **G** 和 **F** 中的所有条目来最小化 Q 的参数：

$$\arg \min_{G,F} (Q) = \arg \min_{G,F} (Q^m + Q^{aux}) \quad (10)$$

Q^{aux} 概念上与 Q^m 有相似的方面：

$$Q^{aux} = \sum_{ki=1}^K \left(\frac{r_k}{s_k} \right)^2 \quad (11)$$

方程(9 中的索引 j 已被 k 替换，因为 k 表示添加到模型中的拉动方程的索引（跨越多个因子源谱图/时间序列）。拉动参数 s_k 指定了拉动的“柔软性”。 s_k 越小， Q^{aux} 在迭代过程中对第 k 次拉动的影响就越大。拉动方法相对来说比较灵敏，因为如果拉动方程与特定数据矩阵不兼容，即当 Q^{aux} 随着 $\overline{f_{ij}}$ 达到其锚点值而减少时，

Q^{aux} 的下降远小于 Q^m 的增加，那么模型会忽略这一拉力。通常，用户需要提供 Q^m 的可接受范围， dQ ，并通过改变 dQ 和拉力参数 s_k 来监测 Q^m 的变化从而判断其是否可以接受。

锚点可以是因子源谱图/时间序列的任何元素，一个或多个因子源谱图/时间序列元素的数学组合，也可以涉及外部数据。

3. 概述

3.1. HDF5 接口的安装（Igor 8）

模型结果的存储基于面向对象的 HDF5 文件系统。由于 IGOR 软件在 IGOR 9 之前默认不加载 HDF5 文件系统的程序，因此用户使用 IGOR 8 时需要在操作系统（OS）中手动链接所需的程序（对于 IGOR 9 或 IGOR 10 不需要）。

创建以下文件的快捷方式

- I. C:\Program Files\WaveMetrics\Igor Pro Folder\More Extensions\File Loaders\HDF5.xop
- II. C:\Program Files\WaveMetrics\Igor Pro Folder\More Extensions\File Loaders\HDF5 Loader.ipf
- III. C:\Program Files\WaveMetrics\Igor Pro Folder\WaveMetrics Procedures\File Input Output\HDF5 Browser.ipf

将快捷方式放置在

- I. C:\Program Files\WaveMetrics\Igor Pro Folder\Igor Extensions\HDF5.xop
- II. C:\Program Files\WaveMetrics\Igor Pro Folder\Igor Extensions\HDF5 Loader.ipf
- III. C:\Program Files\WaveMetrics\Igor Pro Folder\Igor Procedures\HDF5 Browser.ipf

3.2. SoFi 整体结构

SoFi 处理上一章讨论的所有 ME-2 实现方式。操作框架在图 1

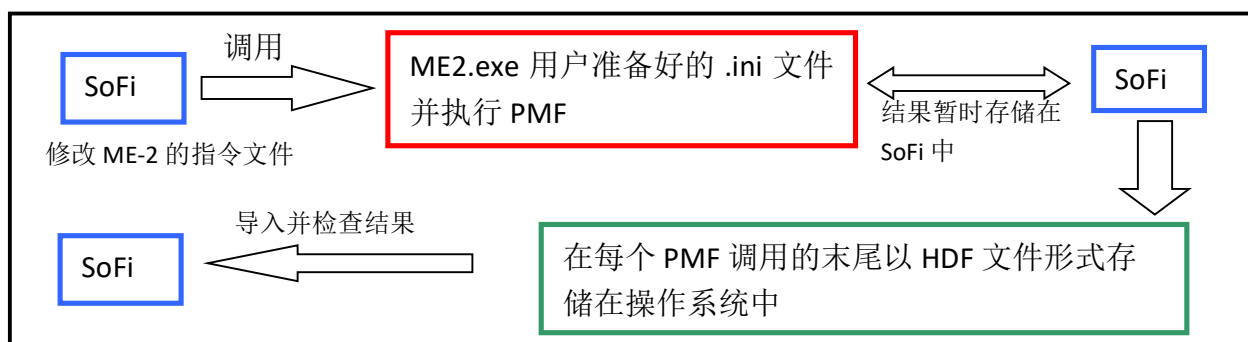


图 1

3.3. 第一步

初始化 SoFi 软件的第一步是打开一个新的 IGOR 实验文件。之后，用户需要导入 SoFi 程序文件（双击 SoFi 程序文件或将文件拖放到 IGOR 中），文件名为：

SoFi-10.ipf（主文件）和 **SoFi_key.ipf**（SoFi Pro/RT 许可证密钥，请勿更改名称）

SoFi procedure 编译完成后，IGOR 的主菜单中会出现一个新标签。

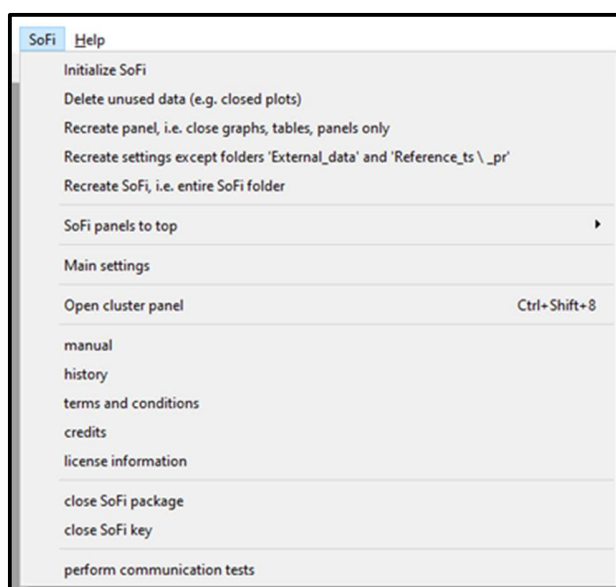


图 2

3.4. SoFi 菜单

“ SoFi” 标签页包含几个主要选项：

3.4.1. 初始化 SoFi

“initialize SoFi”（初始化 SoFi）启动 SoFi 代码并设置其所有变量和字符串

在新的 Igor 实验中第一次启动 SoFi 之前，用户必须阅读用户协议并同意条款和条件。只有在此之后，才能使用 SoFi。条款和条件可以作为 PDF 下载，或随时在“条款和条件”下重新查看。

3.4.2. 删除未使用的数据（例如，已关闭的图表）

“Delete unused data (e.g. closed plots)”, SoFi 会删除所有不再使用的先前图表中的数据（waves）。

3.4.3. 重新创建面板，即关闭图形、表格，仅面板

“recreate panel, i.e....” (重新创建面板，即……) 会关闭所有图形并重新启动主面板。此按钮不会影响 PMF 输入、外部数据和导入的 PMF 结果数据。

3.4.4. 重新创建设置，除了文件夹 “External_data” 和 “Reference_ts/pr”

“Recreate settings except folders “External_data” and “Reference_ts/pr”” （重新创建设置，除了文件夹 ‘External_data’ 和 ‘Reference_ts/pr’ ）除了会关闭所有图形并重新启动主面板外，还会重置与 SoFi 相关的所有变量和字符串。此按钮不会影响导入的 PMF 输入数据、外部数据和 PMF 结果。

3.4.5. 重新创建 SoFi，即整个 SoFi 文件夹

“Recreate SoFi, i.e. entire SoFi folder” （重新创建 SoFi，即整个 SoFi 文件夹）此选项类似于 “初始化 SoFi”。所有内容将被重新创建，所有数据将被删除。

3.4.6. 将 SoFi 面板置于顶部

“SoFi panels to top”，（将 SoFi 面板置于顶部）当打开许多图表、表格和面板时，这是一个有用的选项。可以通过此菜单或指定的快捷键轻松将打开的面板带回前面。面板必须处于在后台打开或被最小化状态，不能通过此菜单构建。

3.4.7. 打开集群面板

“Open cluster panel”，（打开聚类面板）聚类面板不能从主 SoFi 面板或其他任何面板打开，只能在这里打开。注意，Ctrl+Shift+8 可以打开和关闭面板，Ctrl+8 可以将已打开的面板带到前面。

3.4.8. 主设置

“Main settings”（主设置） 允许用户修改主设置。目前，可以从这里修改以下设置：

“GENERAL options” (常规选项) 启用/禁用进度条，启用/禁用关于窗口位置的弹出窗口，设置 PMF 因子位置的颜色，PMF 结果的默认单位，RT 设置

“**DATA options**” (数据选项) 用于设置 SoFi 的信噪比 (S/N)，结果的降权重和稳健值，(未)解释的变异图的 S/N 阈值，日变化图的时间分辨率，周期图的平均阈值

“**FAMILY wave**” (家庭 wave) 用户为 HR 数据定义的 family。详细信息可以在 3.4.9

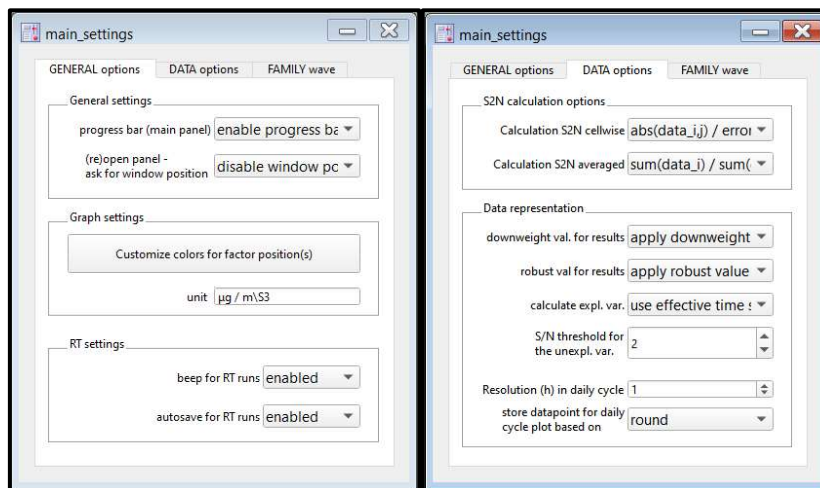


图 3 3

3.4.8.1. 常规设置 (General settings)

用户有两个主要设置可以选择启用或禁用。主 SoFi 面板中的进度条默认启用。我们已经发现启用进度条也会消耗一些计算能力，因此我们提供选项来禁用它。

默认情况下，SoFi 会在关闭面板的最后位置重新打开它们（只要它们没有连接到另一个面板）。如果在不同的屏幕上打开实验，这可能会搞乱面板的位置，从而在屏幕上看不到的地方打开。在这种情况下，应该启用（重新）打开面板选项，以便面板将在新的 IGOR 实验中以打开时的位置打开。

3.4.8.2. 自定义因子位置的颜色 (Customize colors for the factor position(s))

用户可以为最多 100 个因子位置定义颜色。按下 “SHIFT” 键可以选择相邻的因子位置。按下 “CTRL” 键允许用户添加不相邻的因子位置。选择因子后，按下 “**Start defining** 开始定义” 会弹出颜色调色板，在其中定义所选因子位置的颜色。该过程是累积式的。也就是说，用户可以分多次为不同因子设置颜色，也可以修改已经设置过的因子颜色。

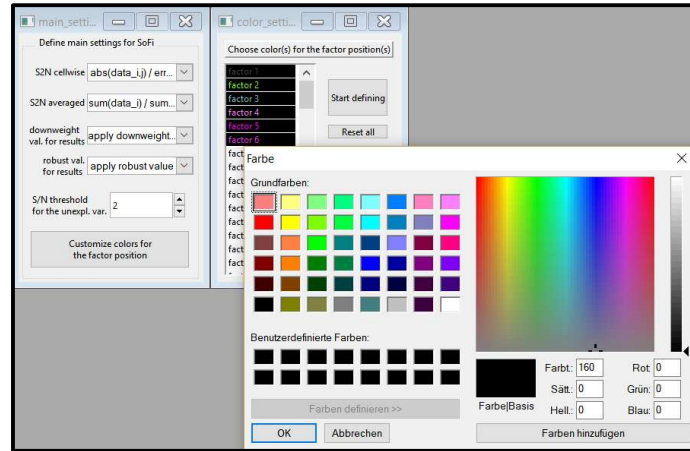


图 4: 4

3.4.8.3. RT 设置 (RT settings)

默认情况下，SoFi 在 RT 运行时发出提示音并自动保存实验。用户可以禁用这些选项。

3.4.8.4. 信噪比 (Signal to noise (S/N))

用户定义在 SoFi 中应采用哪种 S/N 计算方法，以用于逐格(cell-wise approach)或平均方法(averaged approach)。这会影响变量 S/N 图（见 5.3 或当在估计结果图中的 (un)explained variation（见 8.5。如果启用逐格计算，则为 PMF 输入的每个格/条目单独估计 S/N 比率，而平均处理提供每个变量的平均（时间维度）S/N 比率。

Cell-wise（逐单元格）

$\text{abs}(\text{data}_{i,j}) / \text{error}_{i,j}$

$\text{data}_{i,j} / \text{error}_{i,j}$

见补充材料 Brown 等人（2015）

Averaged（平均）

$\text{sum}(\text{data}_i) / \text{sum}(\text{error}_i)$

$\text{sum}(\text{abs}(\text{data}_i)) / \text{sum}(\text{error}_i)$

$\text{sqrt}(\text{sum}(\text{data}_i^2)) / \text{sqrt}(\text{sum}(\text{error}_i^2))$

$\text{data}_i / \text{error}_i$

$\text{sum}(\text{data}_i) / \text{sum}(\text{error}_i)$

见补充材料 Brown 等人（2015）

“i” 是时间上的索引，“j” 是变量的索引。

3.4.8.5. 对结果进行降权 (Downweight values for results)

PMF 结果的误差矩阵可以包含降权值（默认情况）或不显示降权值（见 6.2.3）。前者的优点在于用户可以看到 ME-2 求解器在迭代过程中尝试最小化的内容。后者显示残差，而没

有相关的降权重值（并且也给出了更现实的 Q 值），以防许多变量被分类为弱或坏并被降权重。

3.4.8.6. Robust 模式的结果 (Robust values for result)

Robust 模式通过重新缩放并因此修改相应的误差，消弱了迭代过程对缩放残差的依赖（见 2.1.2），防止例如异常值驱动 PMF 结果。用户可以选择查看包括此修改的结果（默认情况）或从误差矩阵中去除此效果。应用 Robust 模式的优点在于可以检查 ME-2 求解的最小化过程。去除此选项可以判断在所有达到并保持到 Robust 极限的点中，哪些变量或时间点的被解释得更好或更差。

最后这两点（3.4.8.5 Fehler! Verweisquelle konnte nicht gefunden werden.）可以在检查 PMF 结果时调整。这允许用户比较应用这些阈值限制前后的结果。

3.4.8.7. 计算 explained variation (Calculate explained variation)

默认情况下，explained variation 是基于有效时间序列计算的（排除黑名单内的时间点）。如果应使用整个时间序列来计算解释的变异（例如，用于滚动 PMF），可以在此处进行更改。

3.4.8.8. 未解释的变化程度的信噪比阈值 (S/N threshold for the unexplained variation)

Paatero (2010) 引入了一个指标，用于估算各因子能够解释的测量值的变化程度。已解释的变化程度 (explained variation, EV) 是一个无单位常量，通常用来表示每个因子解释了每个变量多少随时间的变化程度。例如，第 k 个因子在第 i 个时间点的已解释变化程度 EV 的方程如下：

$$EV_{ik} = \frac{\sum_{j=1}^m (g_{ik} \cdot f_{kj} / \sigma_{ij})}{\sum_{j=1}^m ((\sum_{h=1}^p |g_{ih} \cdot f_{hj}| + e_{ij}) / \sigma_{ij})} \quad (12)$$

通过将分子中的乘积 $(g_{ik} \cdot f_{kj} e_{ij})$ 替换为 e_{ij} ，公式 12 将成为未被解释的变化程度 (unexplained variation, UEV)。将变量 j 的解释和未解释变化程度表示为 EV_{jk} ，只需在表达式中将对 j 的求和替换为对 i 的求和。如果模型解释了所有变化程度，则 $EV = 1$ 。根据 Paatero (2010)，只有当该变量的 UEV 小于 25% 时，才应将其视为已被充分解释。

UEV 进一步分为具有高信噪比值的数据的真实 UEV 和噪声数据的 UEV。此区分是使用本节中定义的信噪比阈值进行的（默认值为 2）。

3.4.8.9. 日变化的时间分辨率 (Resolution in daily cycle)

在 SoFi 中支持的日变化图的分辨率可以在此处更改。默认值是以 1 小时的时间分辨率形成经典的 0 点到 24 点的日变化。当您的 PMF 原始数据或外部验证数据的分辨率超过一小时请务必根据您的数据进行相应的修改。

3.4.9. 自定义高分辨率 HR 数据的 family

对于 HR 数据，用户可以根据一系列标准定义自定义离子类别 **family**。这不限于特定的仪器类型，但需要输入化学公式信息。用户自定义的 **family** 可以在数据初始化后创建（5.1）。通过点击“添加新 **family** (Add new family)” ，会弹出一个交互对话框，用户需要输入 **family** 名称（后续图例中显示的内容，必须以字母开头，不得有空格或特殊字符）、判定条件 **criterion**（详细信息见下文）和颜色 **color**。SoFi 会自动为离子创建一个 **family**，其名称由数字组成，“未识别(not_identified),” 以及未分配到任何 **family** 的离子（“其他(other)”）。在创建与 **family** 分配相关的 IGOR wave (speciesfamilytext) 之前，“变量数量 (number of variables)” 的列将为空。

可以通过“更新 **family** (Update family)” 按钮更新 **family**（名称、判定条件和颜色）。系统会询问要更新哪个 **family**，并要求提供位置（行，从 0 开始），随后会弹出一个交互对话框。用户还可以删除表格中的单个条目或整个表格。此外，**family** 及其标准和颜色的列表可以保存和加载（通过.itx 文件）。

点击“创建 (CREATE) ‘speciesfamilytext’ ”按钮将创建名叫 speciesfamilytext 的 IGOR wave，并为每个离子分配 **family**。分配给特定 **family** 的离子可以通过在下拉菜单中选择该 **family**，然后点击“检查 ‘...’ (inspect ‘...’)” 按钮进行查看。

3.4.9.1. 根据化学式定义一个 family

定义 **family** 的一种方法是根据元素组成对化学式进行分类（类似于 HR-AMS 数据的默认设置）。该标准对于确定 **family** 分配至关重要。它遵循一个元素（区分大小写）后跟圆括号的格式。如果圆括号为空，则表示接受该特定元素的任意数量。如果在圆括号内指定了一个数字，则表示仅接受该元素的确切指定数量。如果该特定元素的数量应大于或小于后面的数字，SoFi 也接受在括号内的>和<符号。注意：只有包含指定元素的化学式才会被分配到该 **family**。同时，可以为一个 **family** 定义多个标准。如果在整个 **family** 分配过程中应忽略某个特定元素（例如 EESI 数据中的 Na），可以通过“忽略元素(ignore element)” 按钮将其添加到被忽略元素列表中。要移除被忽略的元素，请点击“检查被忽略元素 (check ignore element(s))” 按钮，这将弹出一个表格，在其中删除该元素。

如果用户在“定义 PMF 输入 (define PMF input)” 选择中将数据定义为 AMS 数据，系统会询问用户是否让 SoFi 检查 Aerodyne 默认分配给某个 **family** 的列表中的一些离子，即使这些离子并不完全符合 **family** 标准（例如，CN 被分配给 CHN，CO 被分配给 CHO1 或 CO2 被分配给 CHOgt1）。如果用户希望检查，将弹出一个窗口，列出需要注意的离子，并可以通过下拉菜单来重新分配 **family**。SoFi 会尝试建议 **family**，但如果 Aerodyne 默认的名称与用户定义的 **family** 不同，自动推荐则会失败。在这种情况下，用户必须自己重新选择 **family**。

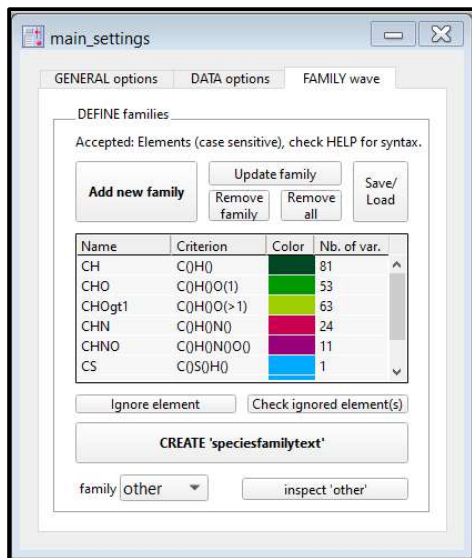


图 5: 5

3.4.9.2. 基于特定元素的数量定义一个 family

定义 family 的另一种方法是根据化学式中特定元素的原子数量进行分类。这种定义 family 的语法是井号 (#) 后跟特定元素，括号内是该元素的数量。例如，基于化学式中氮的数量分配 family 可能看起来像图 6 请注意，“其他 (other)”类别是由 SoFi 自动创建的。

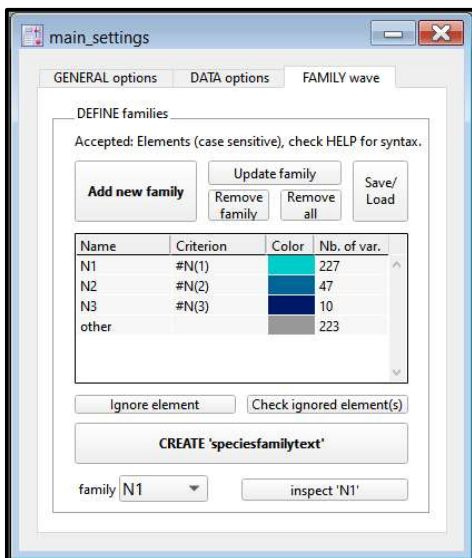


图 6

3.4.10. 用户手册 (Manual)

选项“用户手册 (Manual)”从 Datalystica 服务器下载最新的 SoFi 用户手册，格式为 pdf。需要互联网连接。

3.4.11. 历史记录 (History)

SoFi 从 Datalystica 服务器下载最新的历史记录文件（需要互联网连接）。历史记录文件包含自 SoFi 6.5（2018）以来的所有错误修复和改进。强烈建议始终使用最新的 SoFi 版本。如果使用旧版本，用户有责任确保结果不受新版本中已解决的错误的影响。

3.4.12. 条款和条件(Terms and conditions)

此选项从 Datalystica 服务器下载条款和条件，格式为 pdf。这是用户在首次启动 SoFi 之前必须阅读和接受的内容。

3.4.13. 致谢 (Credits)

“致谢 (Credits)”选项提供了一个电子邮件地址和一个电话号码以获取进一步技术支持。

3.4.14. 许可证信息 (License information)

获取有关许可证类型及其有效期的信息。此外，许可证持有者和许可证序列号也会显示。

3.4.15. 关闭 SoFi 包/密钥程序 (Close SoFi package/key procedure)

SoFi 程序文件是隐藏的，无法手动关闭（终止）。如果例如，SoFi 出现问题并希望重新加载程序文件，可以使用此选项关闭 SoFi 程序。

3.4.16. 执行通信测试 (Perform communication tests)

SoFi 检查 Igor 与操作系统之间以及 Igor 与 Datalystica 服务器之间的通信。如果 SoFi 许可证的激活存在问题，应执行这些测试并将结果发送给我们的支持（support@datalystica.com）。测试结果会显示在 Igor 历史记录中（Ctrl+J）。

3.5. SoFi 密钥激活步骤

在主 SoFi 面板上按下任何标签页或按钮将自动连接到 Datalystica 服务器并验证您的 SoFi Pro 密钥。请确保您已连接到互联网。否则，验证过程将失败，您将被要求在没有许可证的情况下使用 SoFi，即在许可证无法验证的情况下，选项非常有限。

密钥按钮在成功验证时变为绿色，并显示当前正在使用的 SoFi Pro 包的许可证数量。在左侧，SoFi Pro 许可证剩余天数以绿色显示。下面以灰色显示面板在下次服务器连接之前的剩余时间（图 7

要从服务器登出，请按密钥按钮（绿色按钮）。此选项在仅购买少量许可证的情况下非常有用，SoFi Pro 应该在多台 PC 上顺序使用。之后，必须手动重新编译 IGOR。最简单的方法是保存、关闭并重新打开 IGOR 实验。

SoFi 需要互联网进行密钥验证，并将在登录后每 24 小时连接一次服务器。在这 24 小时内，SoFi Pro 可以完全离线工作。因此，如果必须离线使用 SoFi Pro，请确保在连接到互联网时先验证您的密钥。

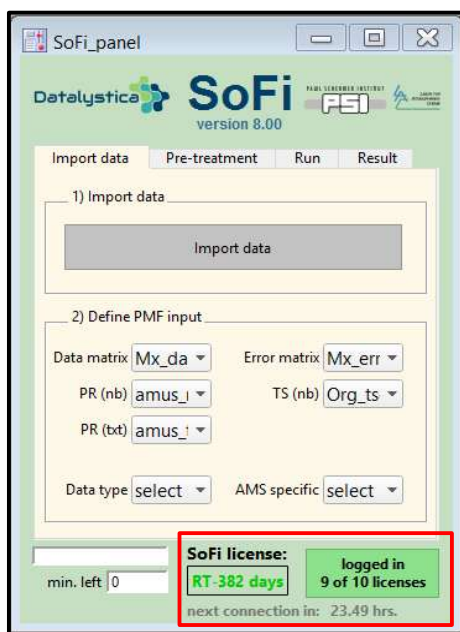


图 7

3.6. IGOR 中的数据层次结构

一旦启动 SoFi，名为 SoFi 的文件夹及其所有子文件夹将在子文件夹下创建（见图 8）

“External_data”	用户加载的原始数据
“Raw_input”	原始 PMF 输入文件
“Raw_pr”	外部参考配置文件及其相关变量基础（必须为 txt 格式，详细信息请参见 5.2.1
“Raw_ts”	外部数据的时间序列及其相关时间戳（详细信息请参见 5.2.2
“Reference_pr”	包含与 PMF 输入矩阵中相同变量的参考的源谱文件。
“Reference_ts”	与 PMF 输入时间（默认分辨率）匹配的外部信息，SoFi Pro：此外，还会计算每小时、每日、每周、每月和每年的分辨率

注意：用户不必担心外部数据的长度（时间序列）/变量的数量（源谱文件）。SoFi 有一个匹配算法（见 5.2，确保外部信息与 PMF 输入的协调与匹配。

“Eff_input”	当前 PMF runs 所使用的 PMF 输入文件。
“Results”	一旦用户导入模型结果，它们将存储在文件夹 “root:SoFi:Results:XXX”中，其中“XXX”是用户定义的 HDF5 文件的名称。
“Preanalysis”	与预处理部分相关的数据（主 SoFi 面板中的第 2 个标签页）。
“Variables”	所有变量对此软件都是必要的。 <u>用户不应在此处修改任何内容！！</u>
“Indep_Analysis”	独立分析面板所需的所有变量和波形 wave（bootstrap、选取框 marquee 和聚类面板 clustering panel）

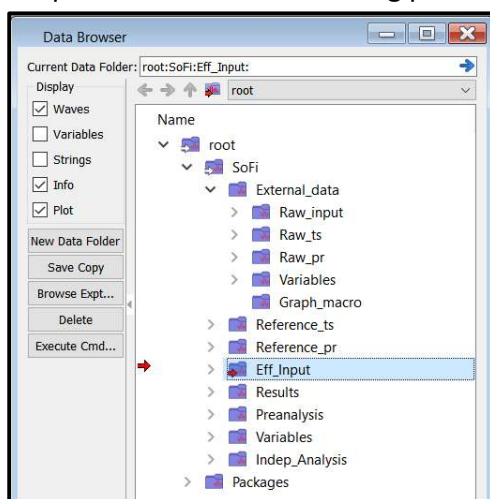


图 8

3.7. 弹窗提醒更新 SoFi 版本

每当 SoFi 软件包成功编译时，SoFi 会在我们的主页（<https://datalystica.com/downloads/>）上检查是否有更新的 SoFi 版本。这通常发生在例如第一次打开 Igor 实验时。每当在主页上找到更新的 SoFi 版本（SoFi 9 和 SoFi 10 独立）时，用户会收到一条专门的警告信息（见图 9

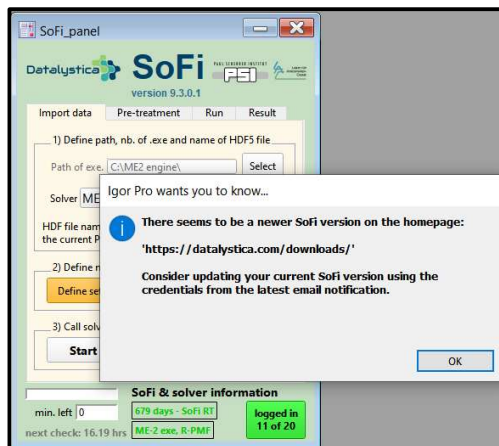


图 9

有几种不同的方法可以将您的实验更新到更新的 SoFi 版本。如果所有 SoFi 实验都已关闭，最简单的方法是直接替换原位置的 SoFi-X.ipf 文件。当实验再次打开时，它将自动读取最新版本。在使用 SoFi 时，可以通过主菜单关闭当前的 SoFi 软件包，然后将新的 SoFi-X.ipf 拖放到打开的 IGOR 实验中并编译。

4. 导入并定义输入数据

SoFi 采用基于具体任务的面板系统结构。主面板如图 10

子面板应通过点击各自的“打开 open”按钮来打开和关闭。使用右上角的 x 关闭面板和/或图表可能会导致某些功能出现意外行为。

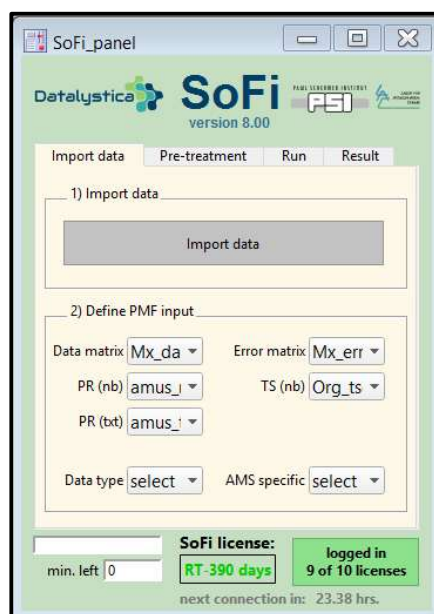


图 10

4.1. 导入数据

第一步是将数据导入当前 IGOR 实验的正确文件夹。这可以手动或自动完成。SoFi 目前支持从 IGOR 文本文件 (.itx) 和 Excel 文件 (.xls 和 .xlsx) 自动导入 PMF 输入，以及从 IGOR 文本文件 (.itx)、Excel 文件 (.xls 和 .xlsx)、逗号分隔值文件 (.csv)、数据文件 (.dat) 和文本文件 (.txt) 导入外部时间序列和外部配置文件。使用“导入数据 (Import data)”按钮会打开导入面板，在此可以指定自动导入的详细信息。有关手动导入，请查阅子章节以了解数据需要复制到的位置。

如果您在导入数据时遇到问题，请联系 support@datalystica.com。我们始终乐意扩展数据导入功能。

4.1.1. 导入 PMF 输入文件

导入 PMF 需要存储在以下路径：“root:SoFi:External_data:Raw_input”。PMF 输入文件应至少包含数据矩阵和误差矩阵，以及时间序列和变量向量。如果用户从 Excel 文件导入，SoFi 也可以帮助计算误差矩阵（图 11）

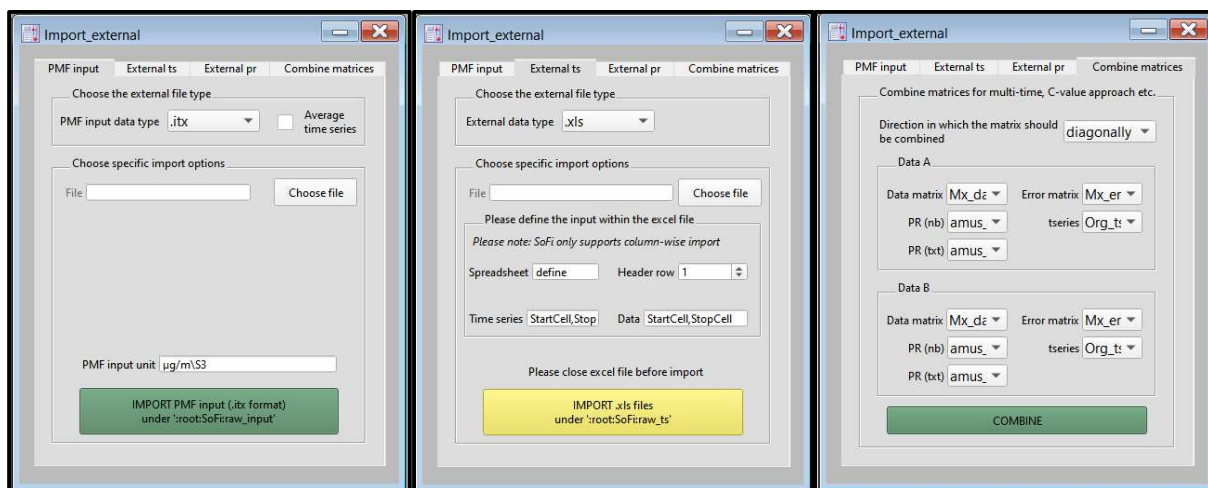


图 11

通过顶部的下拉菜单，用户可以选择 SoFi 应该从中导入 PMF 输入的数据文件类型。对于 .itx 文件，所需的矩阵应作为单独的一维或二维的 IGOR waves，以便在无需其他信息的情况下可以自动导入。“导入（IMPORT）”按钮会打开一个文件选择窗口让用户选择指定的 .itx 文件。

当导入 Excel 文件时，用户首先需要选择文件（通过“选择文件（Choose file）”按钮），然后指定数据矩阵、误差矩阵、时间序列和变量文件所在的单元格。此外，必须指定特定的电子表单名。数据矩阵和误差矩阵可以在不同的电子表单中。对于 Excel 文件，SoFi 根据行名称将变量文件向量导入为文本格式的 IGOR waves，并根据需要自动生成相应的数值 IGOR waves 来定义与 SoFi 兼容的 PMF 输入文件。如果误差矩阵的字段留空，SoFi 假定没有误差矩阵，并提供选项让误差矩阵（后称“Error_Mx_SoFi”）自动计算。对于此功能，SoFi 支持两种类型的误差矩阵计算：

“Polissar 等, 1998”

Polissar 等人 (1998) 的计算方法将数据矩阵和误差矩阵视为环境数据最低检测限 (method detection limi, MDL) 的固定比例：

浓度 > 最低检测限时：

$$\text{浓度} = \text{测量浓度, 误差} (\sigma) = \frac{1}{3} \text{MDL} \quad (13)$$

浓度 < 最低检测限时:

$$\text{浓度} = \frac{1}{2}MDL, \text{误差} (\sigma) = \frac{5}{6}MDL \quad (15)$$

“Norris 等, 2014”

在这种方法中, 对于高于最低检测限 (MDL) 的数据, 使用基于方程的不确定性和误差分数。对于低于 MDL 的数据, 该方法则继续遵循 Polissar 等人 (1998) 的方法。此错误矩阵计算是也用于美国环保局(EPA) PMF。

浓度 > 最低检测限时:

$$\text{误差} (\sigma) = \sqrt{(\text{error fraction} \cdot \text{浓度})^2 + (MDL)^2} \quad (16)$$

浓度 ≤ 最低检测限时:

$$\text{误差} (\sigma) = \frac{5}{6}MDL \quad (14)$$

两个误差矩阵计算方法都需要输入每个元素/物种/离子等的检测限 (MDL), 如图 12 所示。根据 Norris 等 (2014) 的基于方程的不确定性, 另外还需要输入误差分数。因此, SoFi 接受每个元素的误差分数或所有元素的一个误差分数。输入值和误差分数都需要用分号分隔。

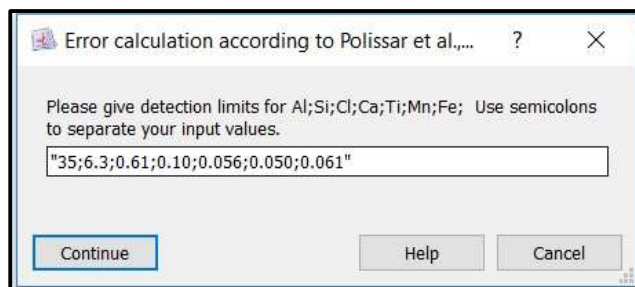


图 12

通过勾选“平均时间序列 (Average time series)”框, SoFi 将导入的 PMF 输入文件平均到所需的时间分辨率 (以分钟为单位), 并将原始输入数据和平均输入数据存储在“root:SoFi:External_data:Raw_input”下。在导入原始 PMF 输入后, 会弹出一个窗口, 用户需要定义所需的时间分辨率 (以分钟为单位)。对于从 .itx 文件的导入, 用户还需要额外定义数据矩阵、误差矩阵和时间序列。

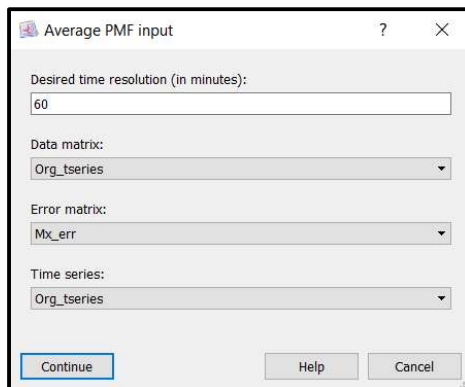


图 13: 用于平均 PMF 输入文件的界面。所需的时间分辨率需要以分钟为单位输入。从 .itx 文件导入 PMF 输入文件时，数据矩阵、误差矩阵和时间序列需要被定义。

如果 minerr1ion 这个 IGOR wave 存在，它也将被自动平均。然而对于 PMF，SoFi 将仅对名为 “minerr1ion” 存在与误差矩阵中的 IGOR wave 进行平均。因此，用户必须手动删除平均后的结果名称后缀，并确保原始的 minerr1ion IGOR wave 具有不同的名称或已经从原始输入文件夹中被删除。

4.1.2. 时间序列导入

独立于 PMF 输入文件的时间序列需要存储在 “root:SoFi:External_data:Raw_ts” 下，如果选择了 “导入外部 ts (Import external ts)” 选项，则会自动导入以上路径。所需的最少数据是一个时间序列和一个相应的数值序列。首先，用户必须选择要导入数据的文件类型。接下来将解释不同文件类型的规格：

itx 文件： 不需要进一步的信息，“导入 (Import)” 按钮将打开一个文件对话框以选择 .itx 文件，文件中的所有 IGOR waves 将被导入。

.xls/.xlsx 文件： 在第一步中，用户必须通过点击 “选择文件 (Choose file)” 来选择文件。用户需要定义时间序列的单元格位置（一次只能一列），通过给出起始和结束单元格，例如 A2, A40。数据（浓度）也是类似的。用户必须再次给出起始和结束单元格，但这里支持多列，例如 B2, H40。还必须定义电子表格名称。SoFi 将按列导入数据。目前 SoFi 还不支持行导入。

.csv, .dat, txt 文件： 用户必须通过点击 “选择文件 (Choose file)” 来选择文件。首先，用户必须定义时间序列可以在第几列找到。

请注意，列号从 0 开始。在接下来的三个选项中，用户应定义所使用的时间格式。“格式 (Format)” 可以定义日期格式，其中 “*” 是日期分隔符（例如 “.”、“/”、“-”），且必须在 “Date separator (*)” 中指定（不带 “”）。时间序列是否包含时间信息，如果包含，则需

在“Time”下进行选择。在下一步中，需要指定包含数据的列。“数据的第一列（1st column of data）”是数据应从中导入的列（注意，列号从 0 开始）。“列数（Nb of columns）”定义应导入多少列。列数可以设置为特定数字，或者要导入第一列数据之后的所有列，则设置为 0。要正确命名导入的 IGOR waves，请设置“标题（行）Header (row)”以找到名称（注意，行号从 0 开始）。如果没有标题，Igor 将把浓度 wave 命名为“waveX”，其中 X 是递增的数字。“跳过的行数（Nb of rows to skip）”定义应跳过多少行。标题不会自动被包含在内。在“列分隔符（Column separator）”中需要定义列分隔符。如果列是由制表符分隔的，请使用\t。“小数字符（Decimal character）”预定义为点，但可以更改为逗号。

如果在不同步骤中导入同名数据的 IGOR WAVES，它们不会被覆盖，而是会在新导入的 IGOR WAVE 名称末尾添加一个连续的数字。

4.1.2.1. AE33 数据导入

SoFi 提供自动读取多个 AE33 文件的选项，提取有效数据并进行源解析 (Sandradewi et al., 2008)。首先原始数据需要被导入。SoFi 提供以下三种选项：(a) 仅导入日期和总的 eBC 浓度数据 (BC6)，(b) 仅导入日期、所有波长和一些诊断参数，以及 (c) 导入所有可用数据。请注意，选项 (a) 不支持后处理（因为仪器状态未被导入），方式 (c) 是推荐选项，但如果导入文件较多，处理过程可能耗时较长。导入的 IGOR waves 将保存在 root:SoFi:External_data:Raw_ts:AE33_raw 下，这是一个包含所有导入的原始 AE33 IGOR waves 的单独文件夹。所有 IGOR waves 和文件夹的名称均应保持原样，否则数据的进一步处理可能会失败。

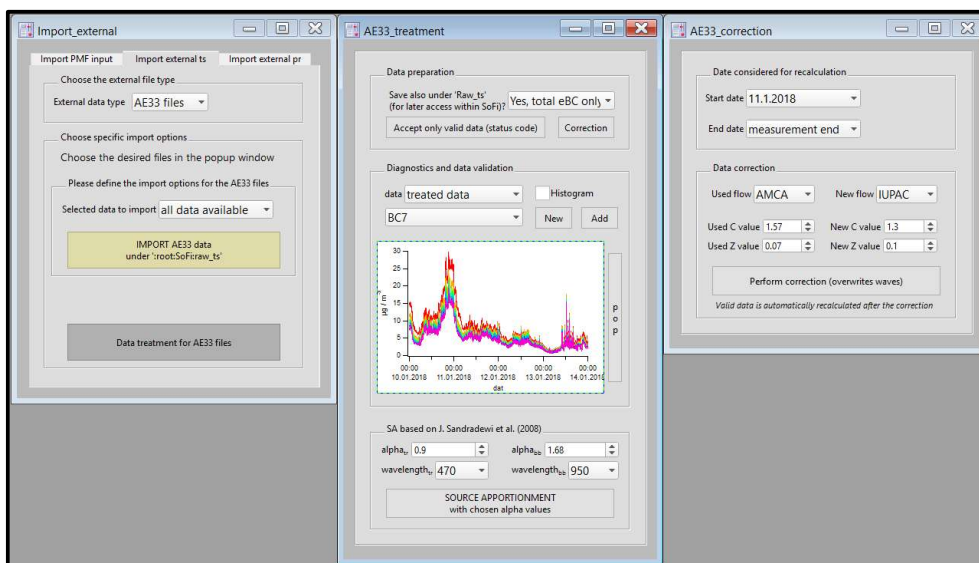


图 14

数据处理在通过“AE33 文件的数据处理 (Data treatment for AE33 files)”按钮打开的单独面板中进行的。通常，原始 AE33 数据包含完整的时间序列，还包括滤膜带移动等校准以及有效测量的数据点。在数据预处理阶段，SoFi 筛选所有有效测量（状态 = 0, 128, 256），并将单位更改为常用单位。有效数据保存在 `root:SoFi:External_data:Raw_ts:AE33_treated` 下。由于在处理外部数据时（黄色面板）SoFi 不会访问此文件夹，用户可以选择将 (a) 总 eBC (BC6)，(b) 所有导入的 IGOR waves 或 (c) 不保存任何 IGOR waves 复制在 `root:SoFi:External_data:Raw_ts`。

在诊断和数据验证中，一些关键参数（来自所有波长的 eBC、总气流量、来自所有波长的 k 参数以及 BCx 与 ATNx_1 的关系）可以绘制以便快速视觉验证（通常或作为直方图）。用户可以访问原始和处理过的 AE33 IGOR waves。BCx 与 ATNx_1 图的衰减经过 ATN0 校正。在完成 BC 的源解析之后，同时 SoFi 也可以绘制出吸收指数 AAE。

SoFi 还提供对 eBC 进行源解析的选项。有关 BC 的源解析方法的更多信息，请参考由 (Sandradewi et al., 2008) 或 (Zotter et al., 2017) 发表的工作。SoFi 使用用户自定义特定波长相对应的 α 值进行源解析。源解析结果直接保存在 `root:SoFi:External_data:Raw_ts` 下，以便后续在 SoFi 中轻松使用。

可以校正滤膜带设置（多重散射参数 (C) 和滤光器泄漏 (Z)），以及重新计算气流量。这些计算可以应用在整个导入数据集上，或从特定日期进行。日期指的是滤膜更换日期（基于带更换程序的状态代码）。对于气流量，用户可以从 AE33 中的五种气流量报告标准中进行选择（表 1 1 对于使用的气流量报告标准，多重散射参数 (C) 和滤膜漏气 (Z) 值，用户应检查设置文件（“AE_SETUP_AE33-[...]”）。

表 1 1

标准	压力	温度
AMCA	101325 Pa	21.11 °C
EPA	101325 Pa	25 °C
ISO	101325 Pa	20 °C
NIST	101325 Pa	0 °C
IUPAC	100000 Pa	0 °C

通过按下“执行修正（覆盖 IGOR waves）perform correction (overwrites waves)”，重新计算将覆盖 AE33_raw 和 AE33_treated 数据文件夹中受修正影响的所有导入 IGOR waves。原始 .dat 文件不会被更改。

4.1.3. 导入源谱文件

外部源谱文件系列需要存储在“root:SoFi:External_data:Raw_pr”下，如果选择了“导入外部源谱（**Import external pr**）”选项，则会自动完成。所需的最小数据是一个源谱文件系列（变量向量）和相应的强度序列。在第一步中，用户必须选择要导入数据的文件类型。导入所需的规范与时间序列导入类似，只是不需要关于所需时间格式的信息。请参考“导入时间序列（**time series import**）”下的信息，了解如何设置导入并用源谱文件系列替换时间序列。SoFi 将配置文件变量向量作为文本波形导入，以便可以直接在“处理外部数据（**treat external data**）”选项中使用。

4.1.4. 合并矩阵

SoFi 可以对矩阵进行对角合并（例如，包含不同时间分辨率的 **PMF** 输入文件）、水平合并（例如，合并来自同一仪器的有机气溶胶和无机气溶胶矩阵）和垂直合并（例如，合并来自同一仪器的不同季节的数据）。首先需要导入待合并的矩阵，之后用户需要提供两个矩阵的名称以及应该合并的相应源谱文件/时间序列。如果需要合并超过两个矩阵，则必须重复该过程多次。

SoFi 将自动去除空行/列，就像在初始化过程中一样，只要缺失值是 **nan** 或 **0**。这对于对角合并的矩阵尤为重要，因为这些行/列在初始化过程中不会被识别。如果 SoFi 找到名为“**minErr1ion**”的最小误差的 **IGOR wave**，SoFi 将询问是否在合并矩阵时也应用于此 **IGOR wave**。

4.2. 定义 **PMF** 输入

SoFi 中的内部字符串变量必须与刚导入的 **PMF** 数据链接。这是通过下拉菜单“数据（**Data**）”（用于数据矩阵）、“误差（**Error**）”（用于误差矩阵）、“**PR (nb)**”（用于以数字格式表示的质谱文件（变量））、“**PR (txt)**”（用于以文本格式表示的质谱文件（变量））和“**TS (nb)**”（用于以数字格式表示的时间序列）来完成的（图 15

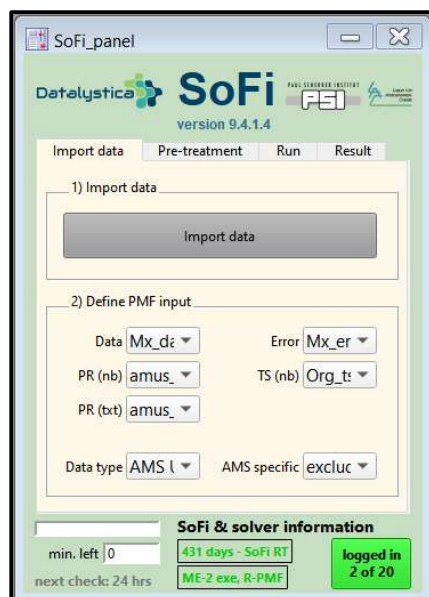


图 15: SoFi 主面板。在“定义 PMF 输入文件（Define PMF input）”下，用户必须定义 PMF 输入文件。

4.2.1. 数据和误差矩阵

包含测量数据的主要矩阵和相应的误差矩阵。SoFi 可以在数据导入期间计算简单的误差矩阵（见 4.1.1）。误差矩阵的名称中必须包含“_err”这个词。这允许 SoFi 在数据和误差矩阵之间进行内部区分。

应避免使用较长的名称，例如超过 15-20 个字符。这是 ME-2 求解器给出的限制。

4.2.2. 时间 IGOR wave

“TS(nb)”包含时间序列的数值信息。日期不必连续，但为了后续绘图的方便，选择线条样式时应避免来回跳跃，建议按照时间顺序排列日期。在热图数据的情况下，设置定义温度的 IGOR wave（无论是连续的还是热图周期）为时间 wave，因为 PMF 将在这个维度上来运行（即，将其视为传统 PMF 中的时间 wave）。

4.2.3. 变量向量

“PR (nb)”包含变量的数值信息。“PR (txt)”必须为文本格式，通常包含变量的文本信息，例如元素分析的元素名称或 HR-AMS 分析的离子名称。然而，即使对于单位质量分辨率（UMR）变量，也需要这两个向量。在这种情况下，两个向量中的信息是相同的，但表示方式不同，即“PR (nb)”为数值，“PR (txt)”为文本信息。在为 PR (nb)选择 IGOR waves 后，SoFi 会询问是否应生成具有相同信息但为文本格式的文本 wave。这对于 UMR 数据非常有用。

4.2.4. 数据类型

下拉菜单允许指定 PMF 输入数据的详细信息，例如数据类型：Q-ACSM/Q-AMS（UMR）、ToF-ACSM/ToF-AMS（UMR）、HR-AMS、Xact、Figaero TD 等。这一点很重要，以确定是否应排除与 CO₂ 相关的变量（ACSM/AMS）（见 4.2.4.14.2.4.2.4.2），以便正确访问离子碎片数据库（如有），并指定 PMF 输入文件的单位。SoFi 会自动为不同的仪器的数据设置不同的单位，对于 ACSM 和 AMS，单位为“ $\mu\text{g}/\text{m}^3$ ”；对于 Xact，单位为“ ng/m^3 ”；对于 Figaero，单位为“counts/s”和“°C”。对于其他情况，用户需要手动定义单位。单位信息在 SoFi 中每次可视化源贡献时都会被使用。

目前，SoFi 为 AMS/ACSM 数据提供以下选项：

表 2: SoFi 中可选的数据类型选项。

选项类型	备注
Q-ACSM / Q-AMS (UMR)	m/z 16、17、18、19、20 基于 m/z 44。
ToF-ACSM / ToF-AMS (UMR)	m/z 16, 17, 18, 19, 20 和 28 基于 m/z 44。
AMS HR / ToF-ACSM X (CO=CO ₂)	m/z 16, 17, 18 和 28 基于 CO ₂ 信号。
AMS HR (CO 拟合)	m/z 16, 17, 18 基于 CO ₂ 信号，CO 应该被拟合！
Xact	-
Figaero TD	PMF 在 Figaero-CIMS 热图上的表现
Other	-

4.2.4.1. UMR ACSM/AMS 的碎片化计算表

随着时间的变化，人们在使用不同的碎片计算表，导致数据结构不一致。为了与现有文献保持一致，特别是涉及排除与 CO₂ 相关变量的修正，输入数据需要与正确的碎片计算表链接起来。根据 Aerodyne 的说法，表 3 表 3 中显示的两个碎片计算表中为最新，任何其他计算方法都已经过时。

表 3

m/z	Q-ACSM/AMS	ToF-ACSM/AMS
16 (O)	0.04*frag_organic[18]	0.04*frag_Org[18]
17 (HO)	0.25*frag_organic[18]	0.25*frag_Org[18]
18 (H ₂ O)	1*frag_organic[44]	0.225*frag_Org[44]
19	0.000691*frag_organic[18],0.002*frag_organic[17]	0.000691*frag_Org[18],0.002*frag_Org[17]
20	0.002*frag_organic[18]	0.002*frag_Org[18]
28 (CO)		frag_Org[44]
44 (CO ₂)	44	44

用户可以选择两个碎片计算表之间的选项。SoFi 会自动检查输入数据中的碎片计算表是否与用户定义的仪器类型匹配。如果不匹配，SoFi 会提醒用户并调整输入数据，以使其与所选的仪器类型对齐。这也可以用于更好地比较来自不同 ACSM 的数据，以便它们都使用相同的碎片计算表。

4.2.4.2. 排除与 CO₂ 相关的变量（适用于 ACSM/AMS）

根据碎片计算表（见表 3，有很多个离子是直接根据 CO₂ 测量计算出的，这些离子并不是实际测量值。排除这些与 CO₂ 相关的变量对于获得物理上有意义且稳定的 PMF 解是至关重要的。在 PMF 输入中包含它们会引入重复信息，这违背了 PMF 的假设条件，即每个变量提供独立的信息。

当在 Robust 模式下运行 PMF 时，这个问题变得尤为关键，这也是标准做法。在迭代优化过程中，m/z 44 的个别副本可以被不同对待：当其残差与不确定性比率超过 Robust 模式的阈值时，一个副本可能会被降低权重，而其他副本则不会。因此，即使相同的测量值代表相同的物理量，它们被处理的方法也无法保持一致。

因此，SoFi 在运行 PMF 之前移除所有与 CO₂ 相关的变量，并在之后使用适当的碎片计算表将其添加回来。这确保了 CO₂ 信号的一致处理，并以受控且保持其物理意义的方式解决了此问题。

如果仅仅降低 CO₂ 相关变量的权重是不够的，因为即使他们的权重被减少，这些变量仍然存在于 PMF 输入文件中。因此，重复信息和对相同信号的不一致处理的问题依然存在。此外，这可能会使 PMF 结果更偏向于解释 CO₂ 的变化程度，这将对含氧的且氧化程度较高的因子造成极高的影响，从而使化学上重要但强度较低的离子被 PMF 算法所“忽视”。

为了确保正确地将 CO₂ 相关变量排除，数据中的碎片化表必须与所选仪器类型匹配。因此，SoFi 根据用户选择的仪器强制执行一致的碎片计算表，以确保可靠和可重复的分析。

4.2.4.3. ACSM 和 AMS 数据的最小误差

对于 ACSM 和 AMS 数据，用户可以进一步将最小误差应用于误差矩阵（如果存在的话）。重要的是在“root:SoFi:External_data:Raw_input:”下将最小误差命名为“minErr1ion”的 IGOR wave（由之前的数据分析软件 SQUIRREL/PIKA 或 Tofware (ACSM) 生成）。注：请勿更改此 wave 的名称！

4.2.4.4. Figaero TD

使用此选项分析来自 FIGAERO-CIMS 的热解气溶胶质谱数据，以根据解吸温度曲线（挥发性）获得对排放源和气溶胶形成过程的洞察。在 SoFi 中进行 PMF 分析之前，数据必须先使用 MaS-PET（由 Dr. Angela Buchholz 提供）进行处理，以确保数据结构、校准和不确定性处理的正确性。从 MaS-PET 导出以下数据。



将所有与 PMF 相关的 IGOR wave（数据(data)和误差(error)矩阵、温度、带有变量信息的数值和文本 wave 以及 idxSample）导入到以下路径：“root:SoFi:External_data:Raw_input:”。选择此选项时，用户还必须向 SoFi 提供原始滤膜的时间戳，这将用于平均后续的热图。它需要是一个与 idxSample 长度相同的数字 wave。

5. SoFi – 预处理

5.1. 处理 PMF 输入中的缺失数据

PMF 数据矩阵的每个单元格代表在 ME-2 最小化过程中使用的一个方程。PMF 输入中的缺失信息必须相应处理，并从方程组中排除。可以区分两种情况：a) 整行/列没有信息，b) 少量缺失数据点。

这两种情况应按此顺序处理：a) 然后 b)，第一步是在“定义缺失数据的符号（**Define symbol for missing data**）”下定义表示缺失信息的符号。

a) 整行/列没有信息

请注意，一旦按下“初始化 PMF 输入并检查缺失数据（**initialize PMF input & check for missing data**）”按钮，整个行/列将直接从“root:SoFi:External_data:Raw:input”中移除。该过程是可以重复的，如果缺失信息存在各种符号，例如 AMS 数据的 NaN 和 0，可以使用相应的按钮更改符号，并可以重复初始化过程。此操作还会在“root:SoFi:Eff_input”下创建当前 PMF 输入文件的副本。

如果存在多个表示缺失数据的符号，则顺序很重要，用户必须从在整个行/列中连续存在的缺失符号开始。这允许正确移除整个缺失的行/列。

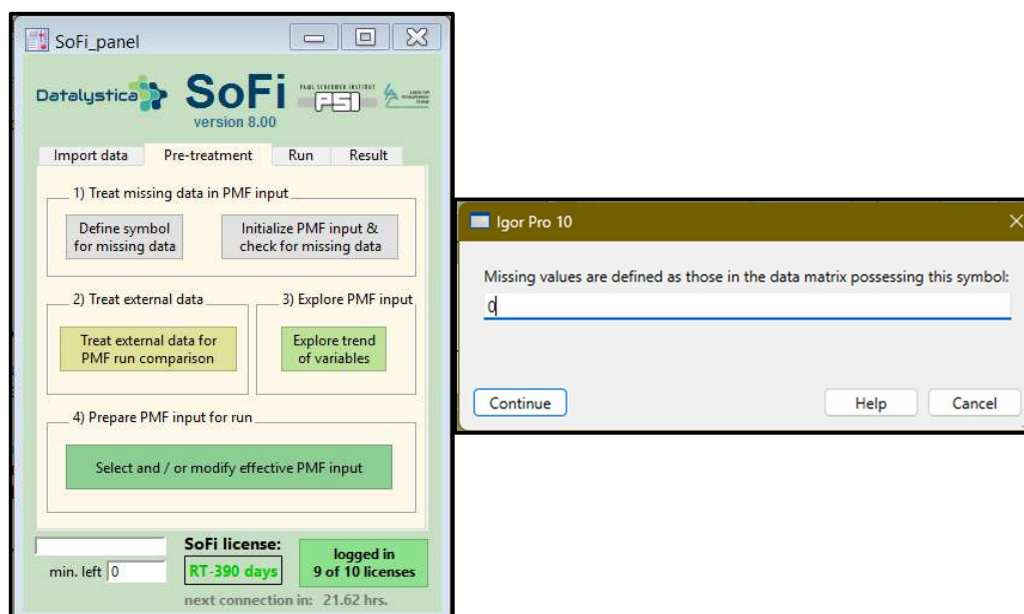


图 16

b) 少量缺失数据点

在此阶段，仅移除完全空的行和列。有关标记和排除单个缺失数据点的信息，请参阅 6.2.2 “II. 缺失数据（II. Missing data）”。

5.2. 处理外部数据

外部信息通过点击“处理 PMF runs/结果比较的外部数据（Treat external data for the PMF run/result comparison）”按钮，使用黄色面板来匹配 PMF 输入数据的时间戳（图 16）。

5.2.1. 源谱文件

该算法提取外部向量中存在的 PMF 变量，并将外部向量的副本存储在“root:SoFi:Reference_pr”下，长度与 PMF 矩阵中的变量数量相同。外部向量中缺失的变量被定义为“NaN”。提取过程通过比较文本信息进行，例如 PMF 输入的离子名称与外部数据的名称。因此，外部数据的 x 轴必须与文本信息相关联（图 17）。

参考源谱文件可能不包括 PMF 输入中存在的所有变量，特别是在 HR 数据的情况下。由于 SoFi 执行精确的文本匹配，这些“缺失的离子”在重新匹配的源谱文件中将留空。如果检测到此类缺失变量（即 NaNs），SoFi 会询问用户是否应该创建缺失离子 wave。该缺失离子波将为所有 NaN 变量赋值。该值基于具有最高强度的变量的比例计算。用户可以自行定义该比例。

5.2.1.1. 在线源谱数据库

Datalystica 提供一个可以直接从 SoFi 访问的参考源谱平台。这需要互联网连接在根据“导入数据 Import data”标签页下“2) 定义 PMF 输入 Define PMF input”中定义的数据类型，如果该数据类型在源谱数据库中存在参考源谱文件，则可以选择“从数据库 from database”选项。一旦选择了“从数据库”选项，可用的参考源谱列表将显示在表格中。用户可以选择他想要导入的参考源谱文件，并按下“提取外部信息 [...]Extract external information [...]”，所选的源谱文件将自动下载并存储在“root:SoFi:External_data:raw_pr”（原始参考源谱文件）下，同时，具有与 PMF 输入相同长度和数量的参考源谱文件（在 AMS/ACSM 的情况下具有相同的碎片化表）将存储在“root:SoFi:Reference_pr”下。

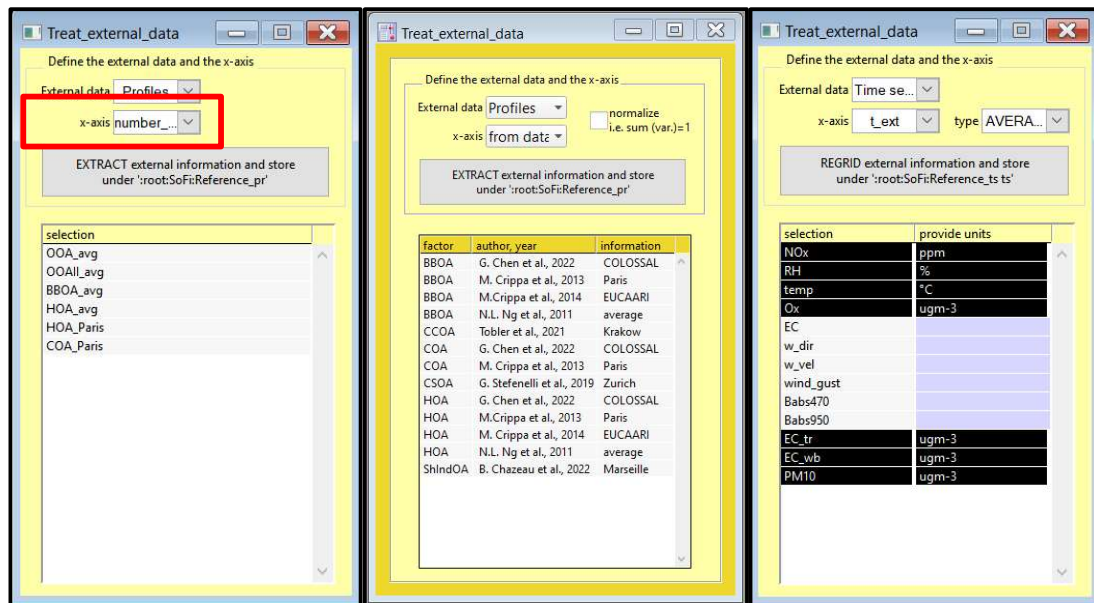


图 17: 左侧面板：参考源谱文件的提取。中间面板：来自数据库的参考源谱文件。右侧面板：用于同步外部时间序列的重新匹配面板。

5.2.2. 时间序列

算法对外部数据的时间序列重新匹配至 PMF 输入数据的时间轴。以下选项可用：

“提取（同时时间戳） **EXTRACT (same timep.)**” 在外部数据的时间戳与 PMF 输入的时间戳相同，但某些时间点的存在或缺失。

“插值 **INTERPOLATE**” 在外部波的时间分辨率与 PMF 输入相似但不相等的情况下。

“平均 **AVERAGE**” 如果外部波的时间分辨率更高，例如 PMF 输入分辨率为 30 分钟，外部分辨率为 5 分钟。

(开始/中间/结束保存时间 **start/middle/end sav. t.**) 定义了数据应该存储的位置，可以是开始/中间/结束保存时间。根据您的 PMF 输入数据的保存时间方式选择选项。PMF 输入的时间序列作为平均的基础，因此应选择与 PMF 输入数据实际测量方式相匹配的平均选项。

每小时、每天、每周、每月和每年的平均值（如图 18: 会自动计算。默认情况下，不会绘制平均值图。如果出于合理性考虑，用户希望绘制所有平均值的图，则应勾选“绘制合理性检查 plot sanity check”复选框。这个选项非常有用，当

- 将 PMF 结果在每小时、每日、每周、每月和每年的分辨率下进行比较
- 对 PMF 结果的快速检查，因为高时间分辨率的分析将需要过长的等待时间。

- 将 PMF 结果与时间分辨率较低的外部数据进行比较，例如将 PMF 应用于 AMS 和 24 小时滤膜数据作为外部数据。然后可以使用每小时或每日选项进行比较。

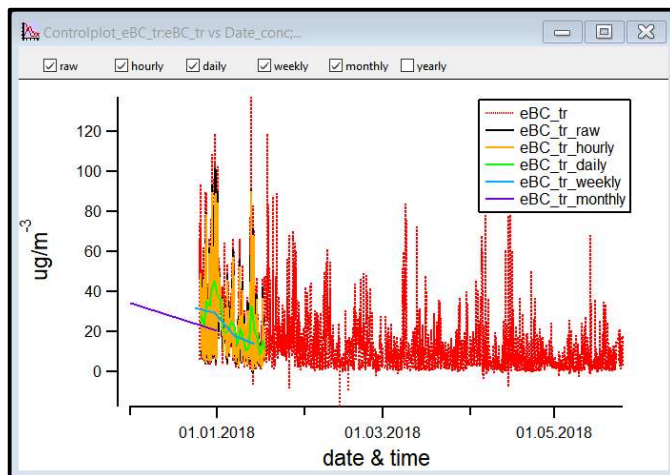


图 18: 默认分辨率下的温度随时间变化的合理性图（黑色，仪器分辨率），使用 PMF 输入的分辨率（红色，原始分辨率）、每小时分辨率（绿色）和每日分辨率（蓝色）进行平均。可以从顶部的控制栏启用/禁用各种分辨率。

该算法会存储外部 wave 的副本，其时间点数与 PMF 输入矩阵中的点数相同，存储在“root:SoFi:External_data:Reference_ts:raw”下，以及相应的每小时、每日、每周、每月和每年的副本，存储在“root:…:hourly”、“root:…:daily”等。缺失的变量被定义为“NaN”。x 轴是以数字表示的常规时间序列信息。

添加外部时间序列的单位是方便的（图 17 右侧面板）。该信息在结果部分绘制相应的外部时间序列时会用到。按下“SHIFT”键可以选择相邻条目，而按下“CTRL”键则允许用户添加非相邻条目。保存外部 wave 后，会对每个选定的外部时间序列弹出合理性图并比较原始时间序列和重新匹配的时间序列（图 18）。

5.2.2.1. 风向和风速数据

风向和风速数据在 SoFi 中进行平均时会用特别方式处理。对于风向，算法通过分别平均角度在 x 和 y 的分量，然后在平均后将它们合并在一起（方向相关平均）。对于风速信息，应用正常平均以及上述风向的方向相关平均。

注意：只有当包含风向信息的波的名称为“wind_dir”时，风向才会进行方向相关平均。当包含风速信息的波的名称为“wind_vel”时，风速会进行方向相关平均。此外，风速的基于方向平均不会直接显示在“Sanity plot”图上（该图显示正常平均的风速），但该波会生成并存储在相应的时间序列文件夹中，名称为“wind_vel degdep”。

5.3. 探索 PMF 输入变量

强烈建议在进行 PMF 分析之前进行变量探索。用户应该知道和了解，特别是关键变量，它们随时间的变化情况。因此，Datalystica 在 SoFi 上开发了一个额外的面板，在点击“探索变量趋势 **Explore trend of the variables**”后弹出（图 16 图 19

5.3.1. 变量的选项

“plot”	点图 point plot、图像 image 和 HR 系列 HR-family（仅适用于 HR-AMS 数据）
“y-var.”	PMF 输入矩阵中所有变量的列表，包括外部数据
“x-var.”	时间，PMF 输入矩阵中所有变量的列表，包括外部数据
“color-var.”	正常 normal（无）、时间、小时、PMF 输入矩阵中任何变量，包括外部数据
“data”	所有数据 all data，有效数据 effective data（排除黑名单数据 = 有效 PMF 输入数据）
“t-plot”	普通时间序列、日变化、周周期、月周期、年周期
“type”	绝对值 abs、比例 fraction（由所有输入变量的总和归一化的变量）、信噪比 signal to noise (S/N)（请参阅 3.3
“根据类别/时间选择 select based on class/time ”	见 5.3.2
“移动窗口 moving window ”	仅在 col-t 与正常值不同的情况下工作。用户可以使用滑块移动，以更好地理解特定日期、时间或浓度下的关系。

5.3.1.1. 点图

点图 (Point plots) 允许用户探索数据，可以作为标准图，即变量或外部与日期/时间的关系，或作为散点图。如果为散点图选择了颜色方案，则可以进行动态图表分析，用户可以通过移动左侧的滑块滚动日期。默认情况下，将绘制有效输入数据（不包括黑名单数据），但用户可以选择显示所有数据。

绘制变量时 SoFi 在后台使用字符串搜索算法，因此变量的名称（在变量文本波中）必须不同。例如，如果存在来自不同粒径范围的元素，则元素名称必须包含粒径范围的信息，例如“Na”对于两个粒径范围的元素是不够的，应该是“Na_fine”和“Na_coarse”或类似的规格。

如果绘制分数 ACSM m/z 44 与 m/z 43 的图，“Sally’s triangle”(Ng et al., 2011a)可以自动添加到图中（图 19

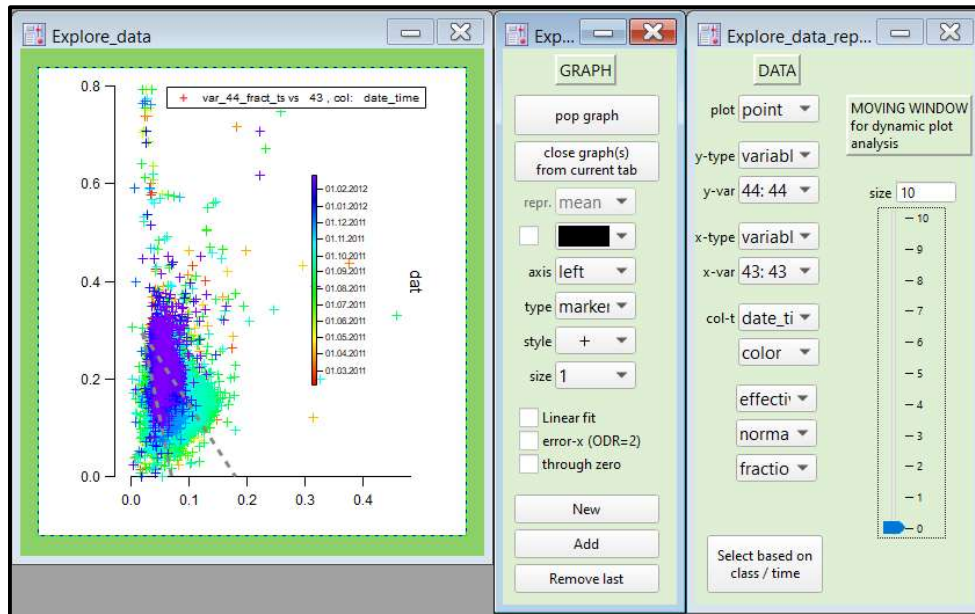


图 19: 输入变量探索面板。左侧部分显示图表，右侧的子面板控制图表和数据的选项。

默认情况下，SoFi 中的周期图在“开始存储时间 start saving time”中报告信息。例如，存储在下午 2 点的每日变化数据代表下午 2 点到 3 点之间的数据。这可以在主设置中更改。

5.3.1.2. 图像图

用户还可以使用第一个下拉菜单中的“图像 image”选项绘制整个数据矩阵（图 20）。如果数据包含多个异常值，用户可以使用子面板“图形 GRAPH”上的滑块（红色矩形）修改颜色比例。此外，用户可以通过直接绘制选框 Marquee 或使用变量选项中的选框限制（蓝

色矩形) 从图像图中选择感兴趣的区域, 然后通过右键选项弹出此信息 (图 21)。

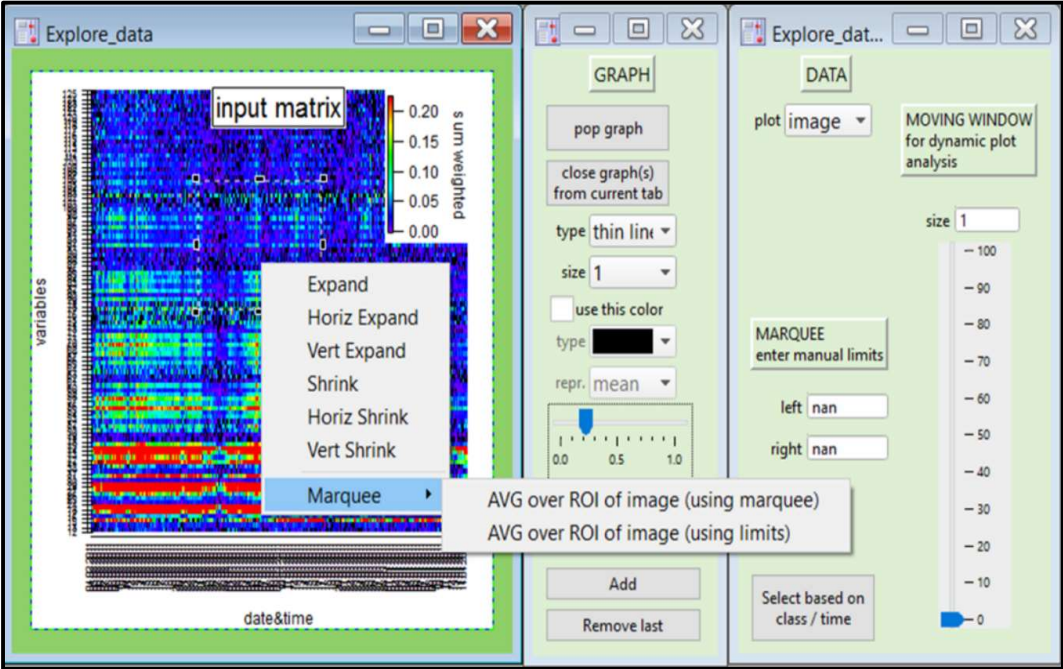


图 21

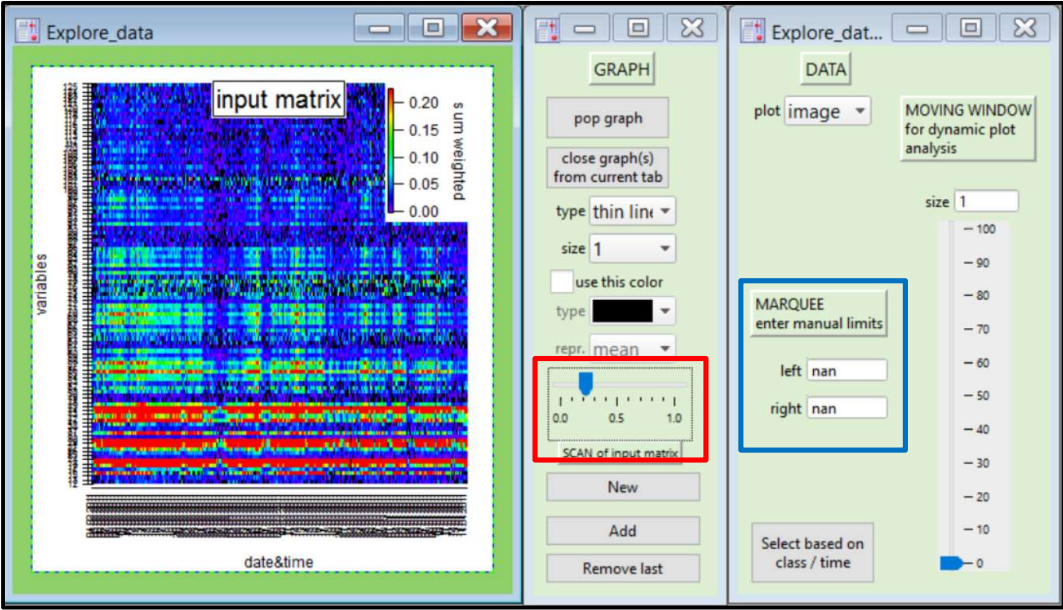


图 20

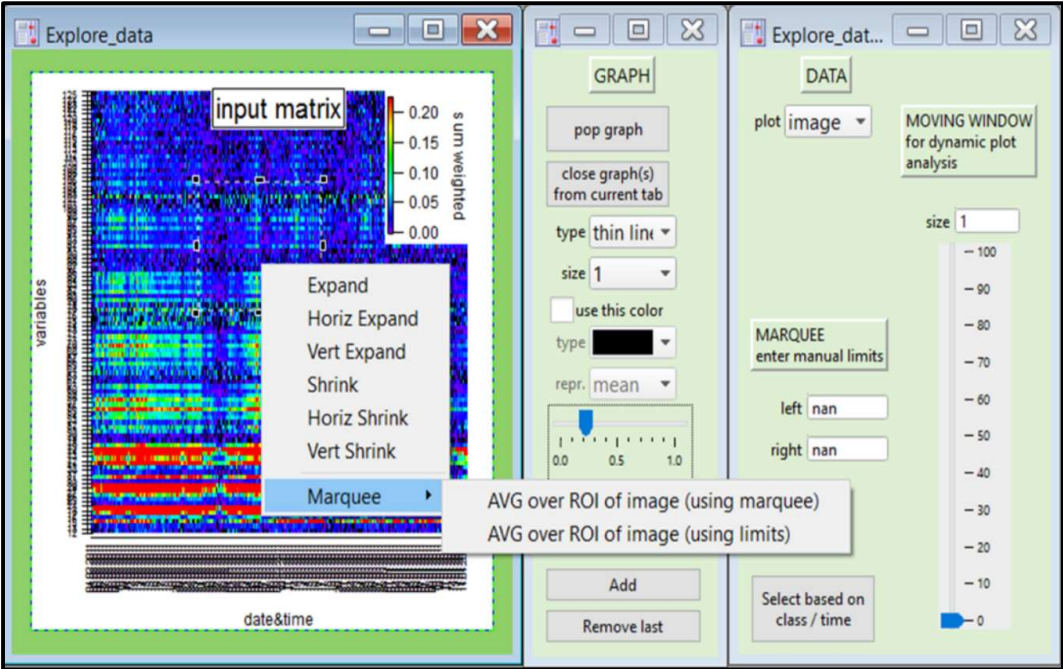


图 21

5.3.1.3. HR-family 图

对于 HR-AMS 数据，用户可以检查每个测量点的 OM:OC、O:C、H:C、N:C 和 S:C 的元素比值（图 22

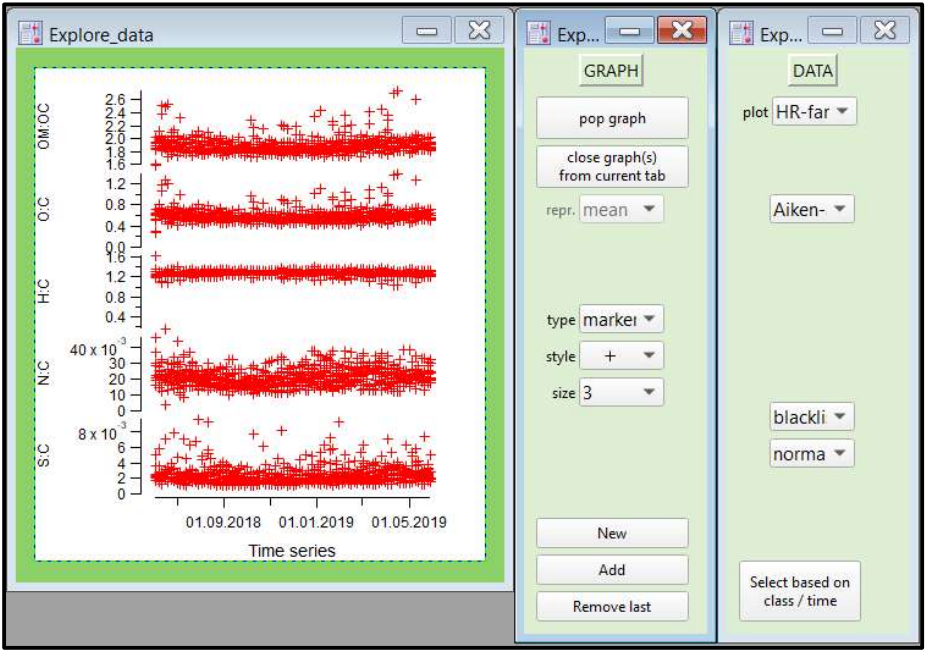


图 22

此外，可以检查质量缺陷图 **mass defect plot**。有关质量缺陷图的更多细节，请参见第 8.5.6.2。请注意，如果用户选择排除与 CO_2 相关的变量，则这些变量不会被重新添加到进行计算的矩阵中，因此这里的图和结果分析面板中的图（变量被重新添加到矩阵中）可能会有所不同。

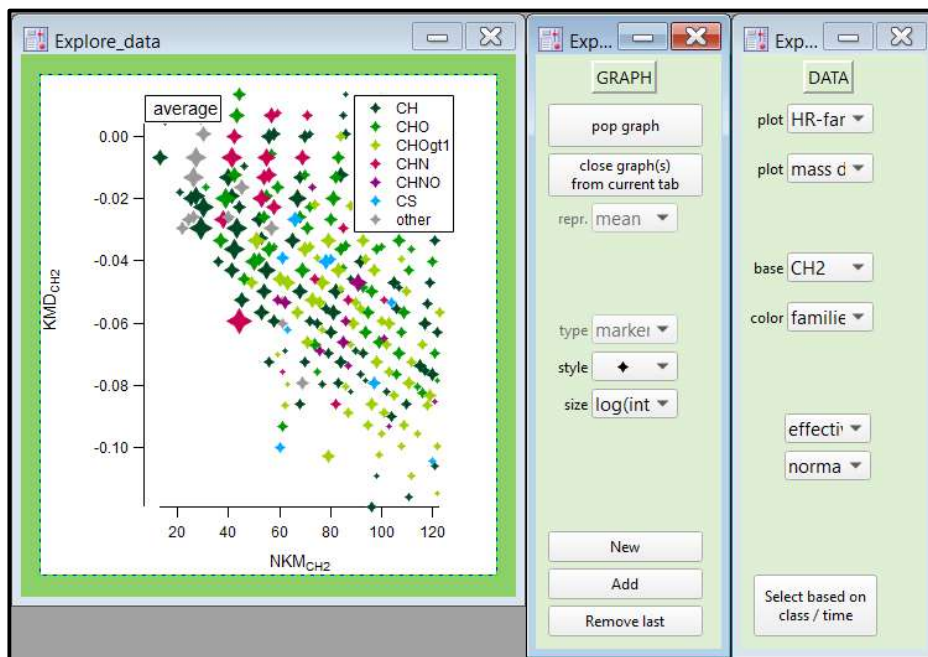


图 23

5.3.2. 根据类别/时间进行选择

此选择面板允许用户轻松高效地根据总结在图 24

“时间选择 **time-selection**” 选择小时、天、月和相关年份

“时间概览 **overview over time**” 用户可以使用选框选择多个感兴趣的区域（在感兴趣的时段周围绘制选框（只有 x 和 y 坐标都落在选框内的点才会被考虑），右键单击并在“选框 **Marquee**”菜单下确认。注意：此选项是累加的，即用户可以多次使用选框工具选择和确认不同的区域。用户在“时间序列波 **time series wave**”下也可以选择外部 **IGOR wave**。这允许根据 **PMF** 输入选择时间点，也可以基于外部 **IGOR wave**，例如在示踪物浓度高的时期调查变量。

“类别 **class**” 所有条目默认被选中，即使用所有点。用户可以指定某些点属于特定类别，并在此标签页中选择这些类别（见 5.4.1

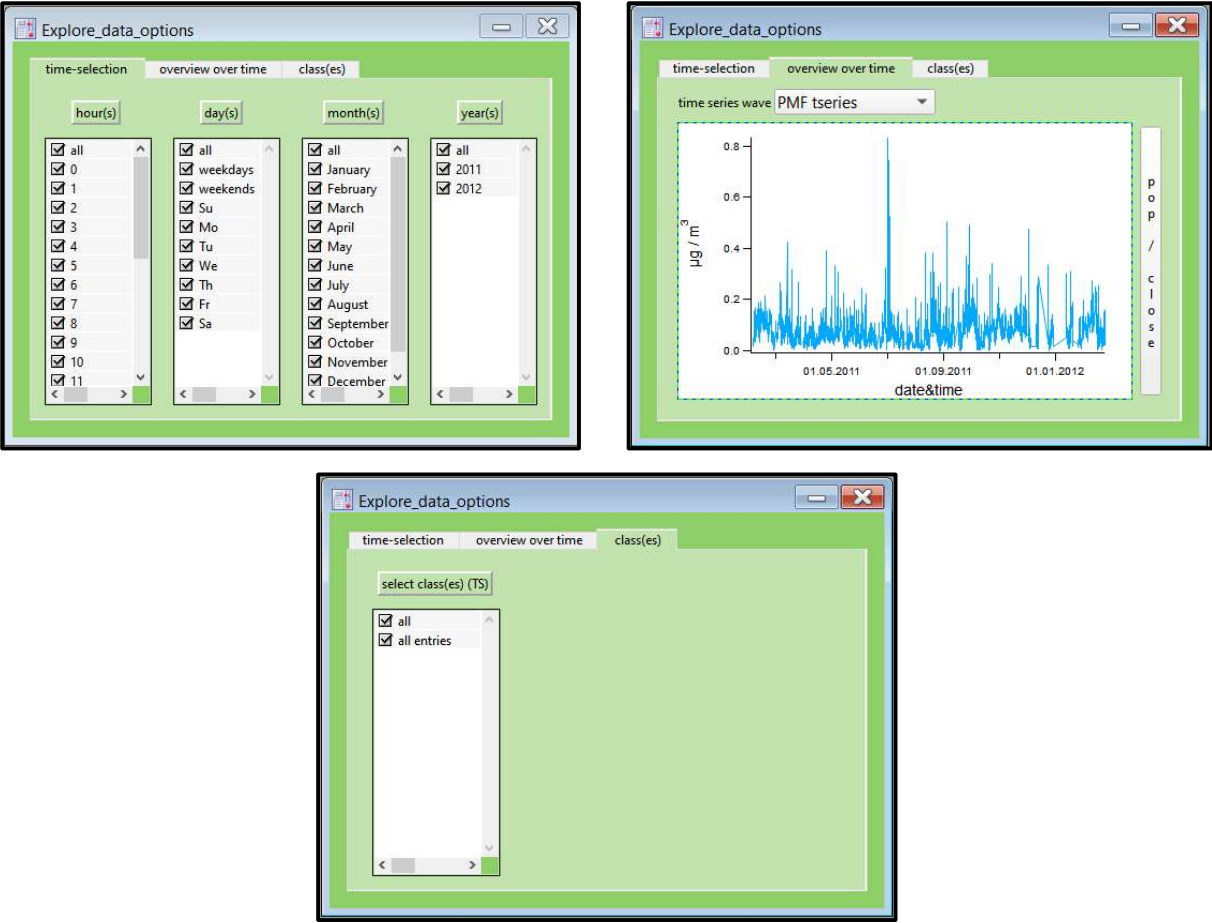


图 24: 基于时间/类别的选择子面板的三个标签页。

5.3.3. 图形表示的选项

- | | |
|----------------------------------|--|
| “弹出图形 Pop graph ” | 从面板中弹出当前图形 |
| “关闭图形 Close graph(s)... ” | 关闭与探索面板相关的所有图形 |
| “类型 type ” | 用户可以在各种默认样式表示之间进行选择，例如实线、虚线等。 |
| “size” | 轨迹或者图案的大小 |
| “repr.” | 仅适用于周期图的选项，可以将均值或中位数及相应的离散度（标准差、四分位距（IQR）、带有第 10 和第 90 百分位数的 IQR、带有百分位数的 IQR 以及最小/最大值）添加到图示中 |

“use this color/define color”	用户可以为当前数据系列定义颜色。如果没有特别指定，SoFi 将从颜色调色板中使用以红色开始的颜色，按行移动并使用每第二种颜色（避免重复使用相似颜色）。
“linear fit”	执行线性拟合，仅在执行散点图时有意义
“error-x..”	假设 x 波的误差，最小二乘法不是相对于拟合线的垂直距离，而是与拟合正交（ODR = 2 在 IGOR 中）
“through zero”	拟合被强制通过零点，即截距 = 0
“New” / “Add”	选择的数据可以使用“New”按钮绘制新图，或使用“Add”按钮添加到现有图表中。
“Remove last”	最后一条数据系列从图表中移除

此探索面板的两个图形示例显示在图 25

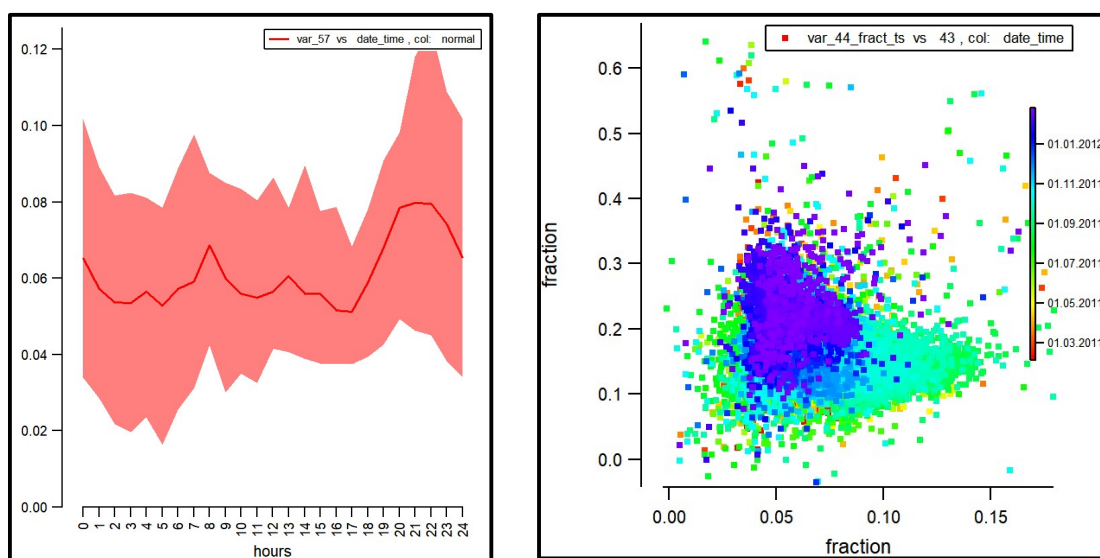


图 25: 每日变化的 m/z 57 质量浓度（左）， f_{44} 对 f_{43} 散点图，按照时间着色（右）。

5.4. 选择或修改有效的 PMF 输入文件

5.4.1. 类别

变量或时间点可以分组为类别。类别是一个强大的工具，用于：

“PMF 输入文件”

PMF 只能在某些变量或时间点的特定类别上进行，例如，PMF 在受大陆空气影响的气团上（回溯轨迹数据将是一个类别波），分别是烟花期间与非烟花期间。

“C 值 C-value”

类别在执行 PMF 时可能会收到不同的相对误差（见 6.2.4.4

“结果检查 Result inspection” 仅显示特定类别中的 PMF 因子源谱或时间序列结果。这在例如 PMF 针对 PM₁₀ 和 PM_{2.5} 滤膜数据运行时可能会很有趣，因为某些结果需要针对相应的粒径范围进行详细检查。

在 SoFi 中，类可以通过两种不同的方式定义。

从现有文本文件导入

如果类已经提前定义，需要在 SoFi 中读取和链接。类波可以存在于变量或时间点上，用户根据“PMF_input_sel”面板（上方红色矩形）中的“源谱文件，时间序列 Profile, Time series”下拉菜单来选择。图 26 图 26）。从 class waves 必须为文本格式，并且长度与 PMF 输入文件（变量/时间点）相同。如果读取了类别的 IGRO wave，则所有变量/时间点必须属于指定的类。如果没有，用户必须在 SoFi 中直接完成对未标记变量/时间点的定义。

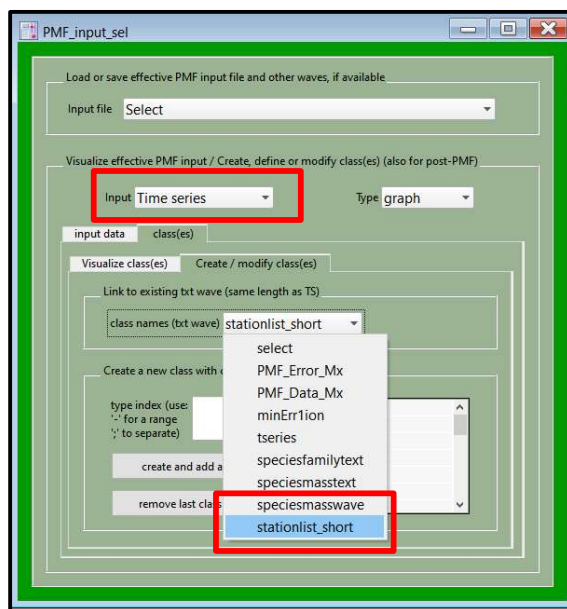


图 26: 基于文本 IGOR wave 创建类别。文本 wave 的维度与时间序列相同，并将每个时间点分配给一个类。

直接在 SoFi 中定义类

如果没有从外部类波定义类，则所有变量/时间点属于默认类“所有条目 all entries”。用户可以通过进入“输入索引 type index”在 SoFi 中为变量定义类：

- 基于索引的区段：使用“-”表示范围，使用“;”分隔单个索引或范围，然后按“创建新块 **create new block**”以创建区段并在弹出窗口中命名它们（红色矩形在 图 27
- 按月区段：使用关键字“**monthly**”按月创建区段。区段将自动以相应时间点的月份和年份命名。
- 按小时区段：使用关键字“**hours**”，后跟应用于区段的小时。例如，“**hours17-20**”将创建一个包含下午 5 点到 8 点之间所有点的区段。
- 分钟区段：使用关键字“**minutes**”后跟应用于区段的小时数。例如，“**minutes10-20**”将创建一个包含所有在 10 分钟到 20 分钟之间的点的区段。
- 固定长度区段：使用关键字“**block**”后跟您希望分组在一起的点的数量。例如，“**block10**”将每 10 个点创建一个新区段。请注意，此格式仅支持低时间分辨率的数据。
- 按仪器分区段：使用关键字“**instruments**”后跟您数据中仪器的数量。仪器的数据必须沿对角线相加。例如，“**instrumets2**”将在时间序列和配置文件中自动创建两个区段。对于仪器名称，会弹出一个窗口，询问仪器的名称，即组名。
- 按热图分区段：对于热图数据，其中 PMF 是在一系列热图上执行而不是时间序列，使用关键字“**idxSample**”。根据提供的滤膜索引，为每个热图创建一个新组。

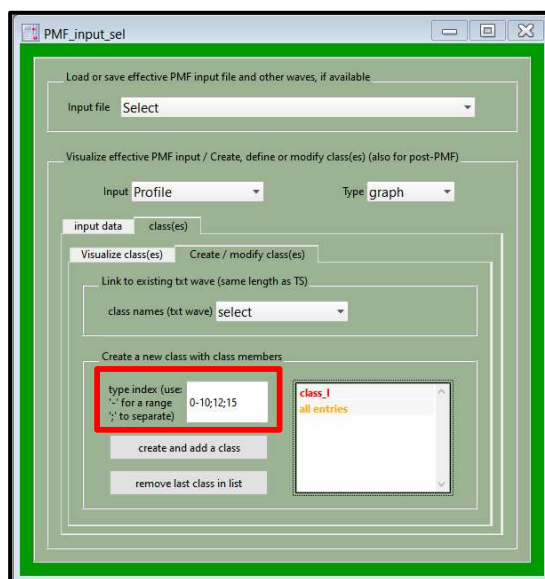


图 27: 定义位置 0-10 的变量类, 和位置 12 和 15 的变量。该类被命名为 “class_l”, 并使用红色。颜色可以通过右键单击要更改颜色的条目来更改。

点击 “创建并添加一个类 **create and add a class**” 会提示用户定义新类的名称和颜色。在示例 (图 27) 中, “所有条目 **all entries**” 类仍然存在, 因为目前只有一些变量被分类。一旦所有变量或时间点都成为用户定义的类的成员, 默认类 “所有条目 **all entries**” 将消失。用户可以在 “可视化类 **Visualize class(es)** (见图 28)” 标签下检查分组的变量或时间点。

颜色由 SoFi 根据类的数量分配。用户可以通过右键单击要更改颜色的类, 从颜色调色板中选择新颜色来轻松更改颜色。如果有拼写错误或错误选择的名称, 双击类名可以更改名称。

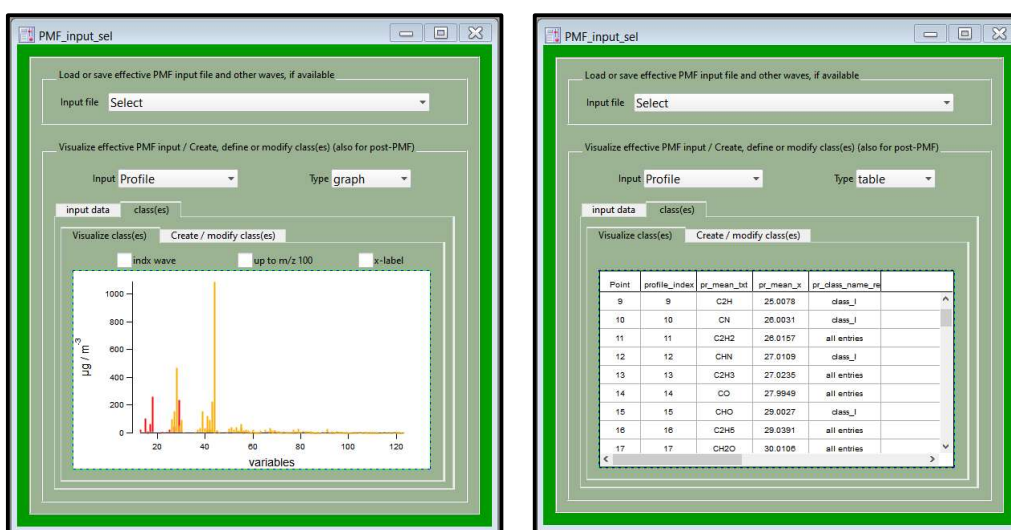


图 28

定义类应在黑名单 **blacklist** 处理之前进行, 因为类 **txt** 波形始终指向整个 **PMF** 输入文件。先进行黑名单处理, 然后再定义变量或时间点的组, 可能会在稍后结果面板中显示结果时导致一些意想不到的问题。

5.4.2. 输入数据

用于 **PMF runs** 的输入数据可以通过黑名单 **blacklist** 功能进行修改。当某些时间点 (由于瞬态事件、电子噪声或仅针对特定点的 **PMF**) 或某些变量 (有问题和/或不必要的变量, 例如非常嘈杂的变量或变化非常小的变量, 或仅针对特定变量的 **PMF**) 不应被考虑用于 **PMF runs** 时, 这一点是有利的。**PMF input** (绿色面板) 显示当前的 **PMF** 输入文件, 而黑名单操作在相邻的子面板上进行 (见图 295.3.2)

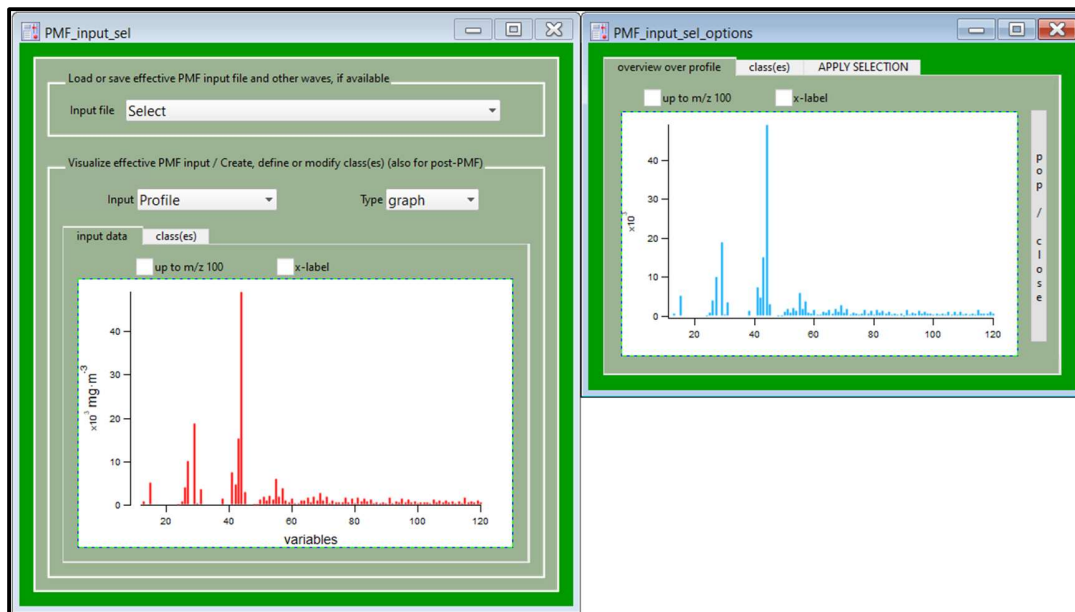


图 29

默认情况下，考虑所有时间点和所有变量作为 PMF 输入文件。移除某些点或变量需要用户在子面板上使用复选框、选框功能或光标（仅针对变量）选择选项。一旦选择完成（这是一个累加过程，必须分别对配置文件和时间序列进行操作）（见图 30 展示了一个随时间变化的示例，用户可以通过在相邻子面板的第三个标签页上按下“黑名单/移除选择...**BLACKLIST / REMOVE SELECTION...**”按钮来将选择加入黑名单（见图 31 图 32

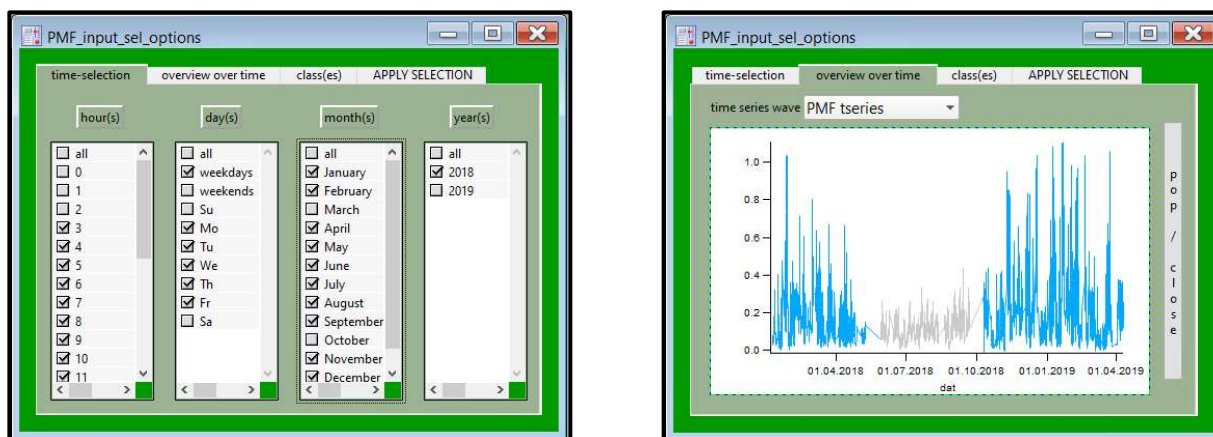


图 30: 左侧：2019 年 3 月和 10 月周末凌晨 0 点到 2 点之间的点被加入黑名单/移除的示例。右侧：通过框选移除中间时间段的示例。

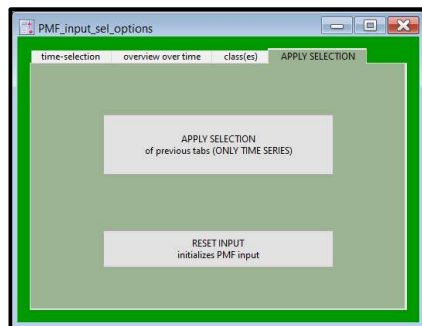


图 31

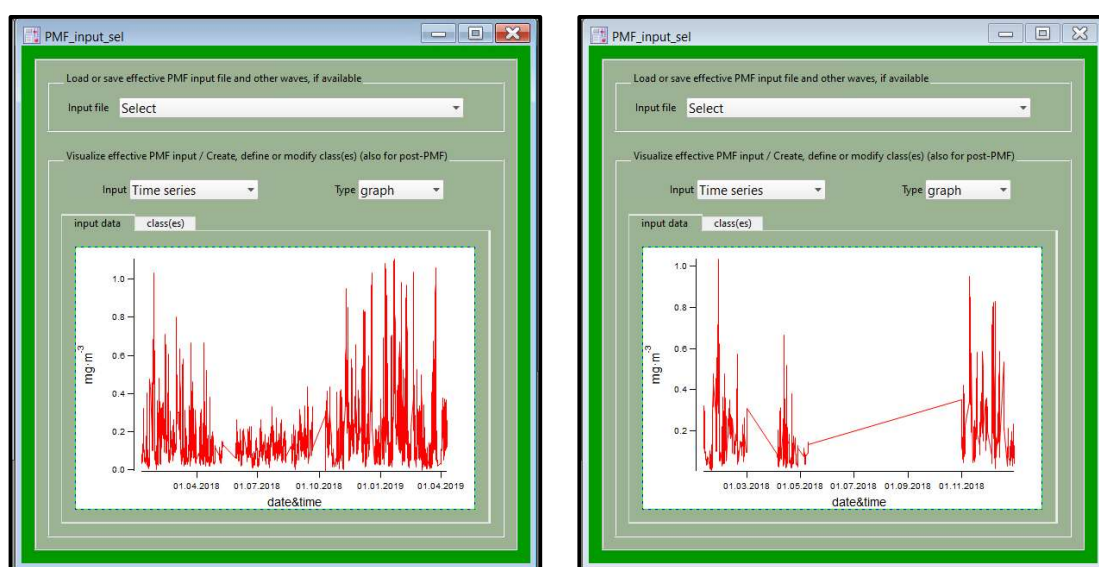


图 32

Blacklist 会移除所选数据。因此，用户应选择不希望继续保留在 PMF 输入数据中的内容。如果某些点被错误地移除，恢复 PMF 输入的唯一方法是点击第三个标签页上的“重置输入 RESET INPUT”按钮。该按钮调用“初始化 PMF 输入并检查缺失数据 Initialize PMF input & check for missing data”的选项（图 165.1

用户特定的 PMF 输入数据可以直接从 SoFi 的“输入文件 input file”中存储和加载 IGOR txt 文件（见图 296.3.16.3.3

5.4.3. 基于 MDL 黑名单单个点

SoFi 提供了根据用户定义的阈值对选定变量的单个测量点进行黑名单处理的选项。为此，必须选择变量并定义阈值。通过点击“黑名单所有低于阈值的数据 Blacklist all data below threshold”按钮，SoFi 将检查所选变量的所有测量值，所有低于阈值的条目将被设置为缺失数据阈值减一。缺失数据阈值在橙色面板中定义（见 6.2.2 此值以及误差矩阵替换值在此处黑名单处理后不应更改，否则数据将无法正确排除。

变量也可以通过点击恢复所选变量的所有数据 **restore all data for chosen variables** 来“取消黑名单”，所有数据点将重置为原始数据和误差矩阵。

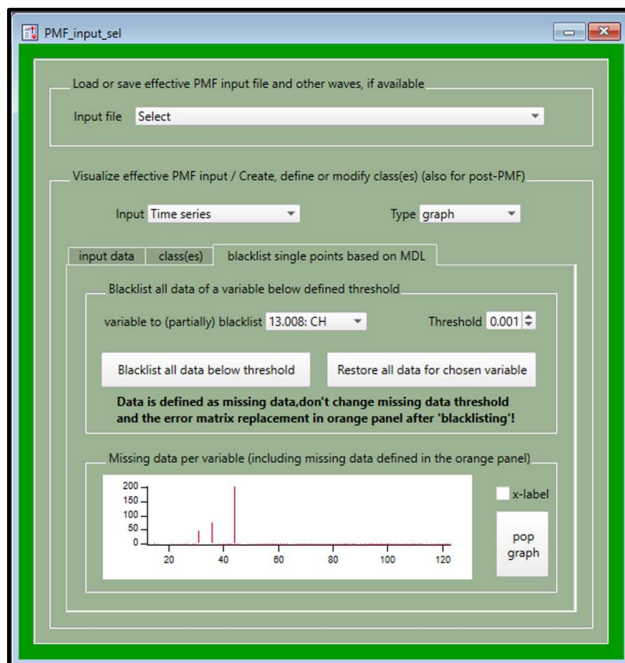


图 33: 基于用户定义的阈值（例如，检测限）对变量的单点测量点进行黑名单处理。

该图显示了被排除在 **PMF** 分析之外的数据点数量。要么是因为它们在这里基于阈值被列入黑名单，要么是因为存在被处理为空的单元格。6.2.2

6. SoFi – PMF runs

6.1. 定义路径、可执行文件数量和 HDF5 文件名称

“Run” 标签页专门用于设置特定的 PMF 调用，使用在“导入数据 Import data”和“预处理 Pre-treatment”标签页下定义的 PMF input 和误差矩阵，最后调用 ME-2（见图 34）

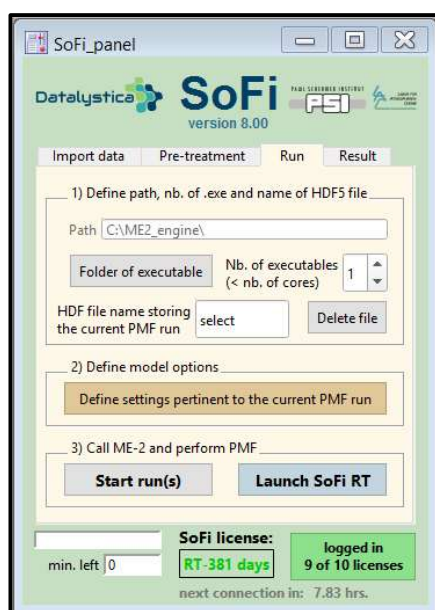


图 34

6.1.1. 路径

用户通过按下“可执行文件文件夹 folder of executable”按钮来提供 Me2.exe 文件所在的位置。

用户需要在 ME2 engine 文件夹的路径拥有写入权限，因为 hdf 文件将被生成并写入。

6.1.1. 可执行文件数量

用户可设置将并行运行的 ME-2 实例数量。此数量必须小于或等于您电脑中的 CPU 核心数量。

示例：如果您选择了 4 个核心，SoFi 将在解决过程中为四个实例中的任何空闲 ME-2 实例分配一个新任务。

可能会发生某些任务的计算速度比其他任务快（或慢）。因此，DOS 窗口不一定会同时弹出。此外，日志文件/命令历史中的 runs 编号不一定是递增的。特别是在 rolling PMF 的情

况下，ME2 引擎的初始化时间可能比运行 PMF 的时间更长，因此并不是所有核心都能得到利用（例如，使用 rolling PMF 时）。

6.1.2. HDF5 文件名

用户在“存储当前 PMF runs 的 HDF5 文件名 HDF5 file name storing the current PMF run”下定义 PMF runs 的适当名称。如果该名称已经存在，SoFi 会提示用户更换名称。

!!! “删除文件 Delete file”按钮会从操作系统中删除指定的 HDF5。尽管有最后的警告提示，用户应注意不要删除重要信息。SoFi 不管理 HDF5 文件的隐藏备份 !!!

6.2. 设置一般模型选项

橙色按钮“设置与当前 PMF runs 相关的设置 Define settings pertinent to the current PMF run”（图 34 图 43: 43

6.2.1. 设置因子数量

“起始因子数 **starting factor**”定义了模型所研究的最小因子数量（p）。

“结束因子数 **ending factor**”定义了模型所调查的最大因子数量（p）。

“PMF 调用次数 **nb. of PMF calls**”是指使用相同设置的调用次数。SoFi 将自动调整此值以进行 fpeak 分析（见 6.3.26.3.36.2.4.4

示例：“起始因子”= 1 和“结束因子”= 5，模型将返回 1、2、3、4 和 5 个因子的解。

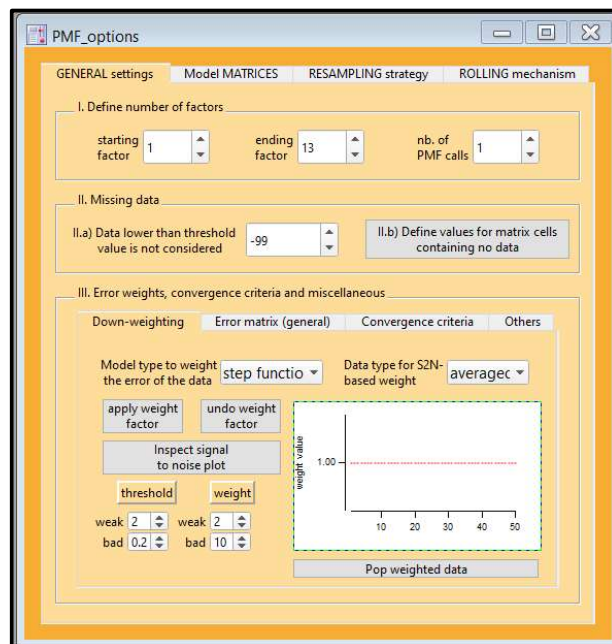


图 35

6.2.2. 缺失数据

ME-2 求解器需要明确的指示来定义什么是缺失值。在第一步中，在初始化期间，完全空的行和列被移除，如 5.1 在此阶段，PMF 输入必须仅包含一个缺失数据的符号，例如“NaN”、“inf”、“0”或“-999”，但不能超过一个。

点击“II.b) 定义不包含数据的矩阵单元的值 Define values for matrix cells containing no data”下方会提示用户定义缺失数据的符号以及其误差单元的不确定性值（见图 36）

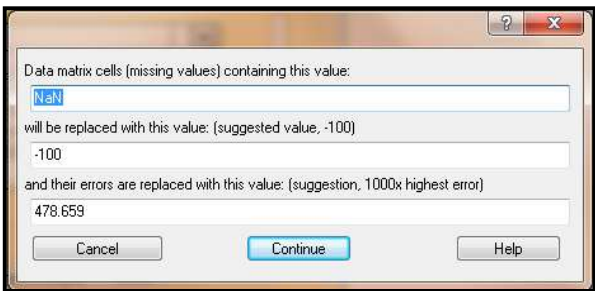


图 36

6.2.3. 降低权重

简而言之，如果估计的测量误差不确定，通常在检测限附近，尤其是当系统误差部分（背景或空白）未被考虑或估计不准确时，可以考虑增加误差/降低误差权重。权重策略基于信噪比（见 3.3 图 37）

模型类型 (Model type)

“选择 **select**”

不对误差矩阵进行修改

“阶跃函数 **step function**”

是 Paatero 和 Hopke (2003) 提出的两步函数，应用于误差矩阵。用户定义步骤及其对应的权重。默认情况下，如果信噪比（S/N）小于 2，则变量被视为弱，并按 2 的因子进行降权；如果信噪比（S/N）小于 0.2，则变量被视为坏，并按 10 的因子进行降权。

“1/S2N 函数 **function**”

当 S2N 小于 1 时，误差会随着惩罚函数 $1/S2N$ 持续降低，其他情况下则不受影响。这个想法最早在 Visser 等人的研究中讨论过， (2015).

数据类型 (Data type)

“平均 **averaged**”

加权数据是基于平均信噪比进行的

“单元级 **cell-wise**”

加权数据是在每个单元格上单独进行的。

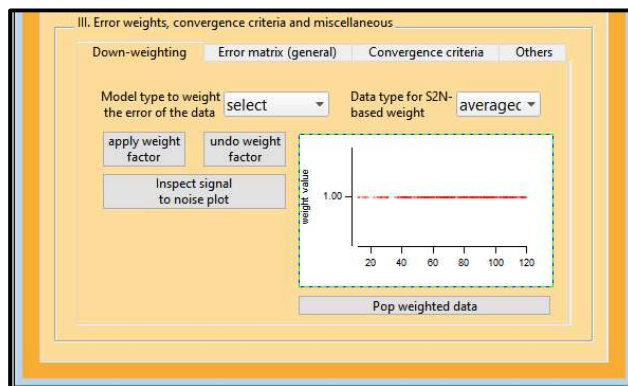


图 37

加权数据显示在小图中。有关更多详细信息，可以使用“检查信噪比图 inspect signal to noise plot”和“弹出加权数据 pop weighted data”按钮分别检查 S/N 图以及受影响的误差（图 38 可以使用“应用权重因子 apply weight factor”和“撤销权重因子 undo weight factor”按钮分别应用和移除权重。

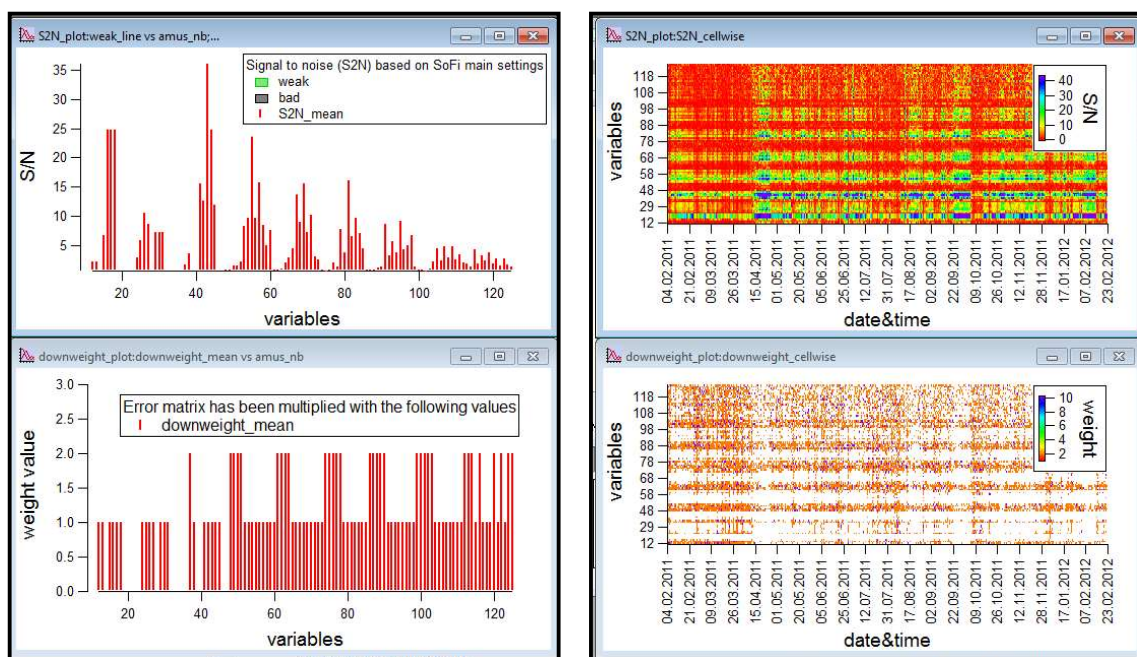


图 38

如果用户处理 AMS / ACSM 数据并选择“考虑所有变量 consider all variables”选项（不排除与 CO₂ 相关的变量），则可以使用降权重选项（图 39

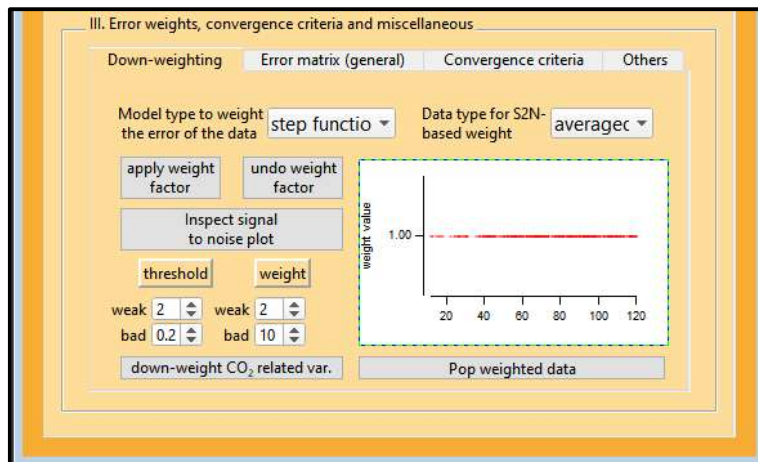


图 39

许多刚提到的自由参数尚未系统测试，因此在此阶段无法给出推荐值。但据我们所知有研究团队计划进行这方面的工作，以评估 PMF 结果对这些设置的敏感性。

6.2.4. 误差矩阵（一般）

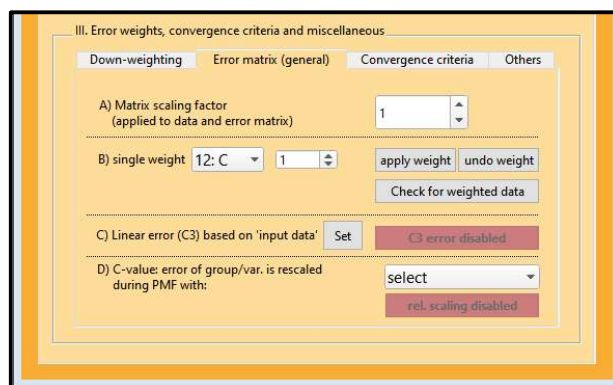


图 40

6.2.4.1. 矩阵缩放因子 Matrix scaling factor

如果传递给 ME-2 求解器的最小误差太小，求解器将无法运行。这对于质量浓度非常低的数据集可能是关键。在这种情况下，可以应用矩阵缩放因子，这意味着在将数据发送给求解器时，数据和误差矩阵将乘以该因子。SoFi 中的输入数据不会受到影响。在导入 PMF runs 时，相同的矩阵缩放因子应用于时间序列。因此，总体而言，PMF 结果将是相同的（因为缩放应用于误差和数据矩阵，结果分别被归一化（因子源谱）和重新缩放（时间序列））。

6.2.4.2. 单一权重 Single weight

可以使用变量的下拉菜单和权重值对单个变量进行降权（图 40 图 41

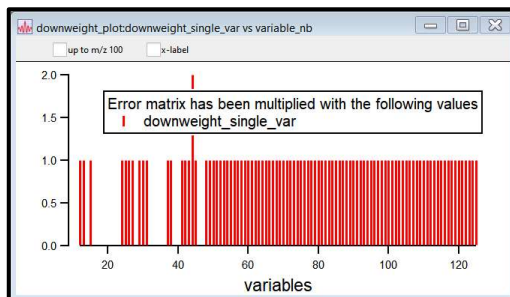


图 41

6.2.4.3. 线性误差 Linear error

在 SoFi 中，ME-2 中存在的线性误差表达式（C3 表达式）可以添加并启用用于 PMF 计算（图 40）

当前误差矩阵将相应更新：

$$error\ matrix[i][j] = error\ matrix[i][j] + C3 \cdot data\ matrix[i][j] \quad (17)$$

点击“A) 基于‘输入数据’的线性误差（C3）Linear error (C3) based on ‘input data’”允许用户定义 C3 值。一旦 IGOR 中的 C3 变量不同于 0（默认值），C3 误差按钮将变为绿色，并且 C3 值将被考虑。

6.2.4.4. 相对缩放（C 值方法）

相对缩放在分析师有理由相信某些变量的误差需要相对于其他变量增加/减少的情况下是有用的。几种原因可能使这种方法合理。简要讨论了两个可能相关的案例。

“各种相对误差 **various relative errors**” 来自不同仪器的数据具有不同的相对误差，例如一个仪器为 0.1%，另一个为 10% 的误差。PMF 将更好地拟合第一个仪器的数据，因为其 Q 贡献要大得多。

“不同数量的变量 **different amount of var**” 不同仪器提供的变量数量不同。比如一个仪器有 100 个变量，另一个有 10 个。两个仪器的变量都有适当的误差估计 5% 的误差。然而，PMF 会更好拟合第一个仪器的数据，因为它的 Q 贡献要大得多。如果同一仪器的变量之间的协方差很高，这一点就变得更加明显。对某些变量的误差进行重新缩放是通过 SoFi 中类的定义来实

“选择 select”

“用户特定的 C 值 User-specific C-values”

现的。需要重新缩放的变量必须属于同一类（见 5.4.1

误差矩阵不被修改

用户定义应在选择此选项后在弹出的表格中定义需要测试的 C 值（见图 42

$$\sigma_{new} = \sigma_{old} \cdot \frac{1}{C_{value}}$$

(18)

其中 σ 代表需要重新缩放的误差。默认 C 值为 1，即新误差保持不变。

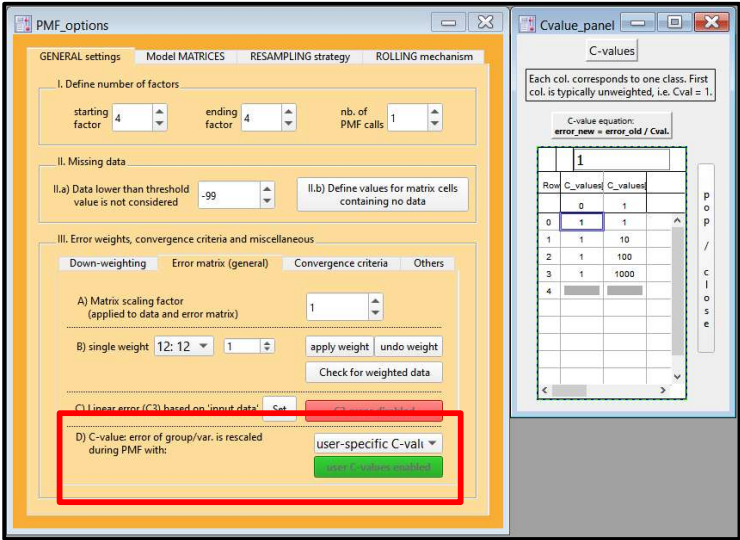


图 42

可以为每个定义类指定一个 C 值。例如，如果定义了三个类，则相邻表将弹出三个列以供填写。条目的数量也定义了 PMF runs 的数量。在当前示例中（见图 42

另见 **Fehler! Verweisquelle konnte nicht gefunden werden.**（组合标准的示例）用于使用 criteria 面板监测 C 值运行。

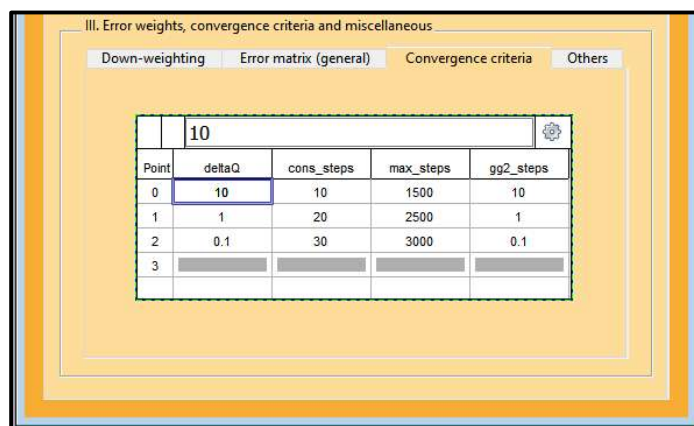
有关 C 值手动技术的更多详细信息，用户可以参考讨论 AMS 与气相数据之间 C 值加权策略的论文（Slowik et al., 2010; Crippa et al., 2013）。

6.2.5. 收敛标准 Convergence criteria

在 PMF runs 期间控制拟合优度的参数可以直接从第一个标签页上的表格中配置（图 43: 43

有两个收敛标准。第一个标准检查迭代过程步骤之间 Q 的稳定性。如果最后一步和当前迭代步骤之间 Q 的差异小于“deltaQ”，例如 <0.1 ，对于“cons_steps”列中提供的连续步骤数量，算法收敛。另一个收敛标准是“gg2_steps”，其中 Jacobian 平方矩阵的范数（包含 Q 对模型变量的所有偏导数的矩阵）与提供的阈值进行比较，例如 0.1 。“max_steps”列只是给定阶段允许的最大迭代步骤数量。如果在阶段 1 或 2 期间达到此最大步骤数，模型将进入下一个阶段并继续寻找收敛解。然而，如果在最后阶段发生这种情况，模型将返回一个非收敛解。非收敛解在导入后保持为空，这也会在日志文件中报告。用户应该能够轻松识别非收敛解，并可以随后使用不那么严格的收敛标准重新运行模型，例如将三个阶段的“deltaQ”和/或“gg2_steps”的值增加 10 或 100（它们应该一起增加/减少）。

注意：通常情况下，误差矩阵或 PMF 设置存在问题，因此无法实现收敛。因此，在更改收敛标准之前，用户应仔细检查误差矩阵和 PMF 设置。



Point	deltaQ	cons_steps	max_steps	gg2_steps
0	10	10	1500	10
1	1	20	2500	1
2	0.1	30	3000	0.1
3				

图 43: 43

6.2.6. 预 PMF 评估 Pre-PMF Evaluation

在运行 PMF 之前，可以进行快速的 SVD (singular value decomposition) 分析。SoFi 将显示奇异值与因子数量的关系以及奇异值变化的斜率。此分析可以提供有关 PMF 输入矩阵秩的一些见解，这可能有助于确定理想的 PMF 因子数量。

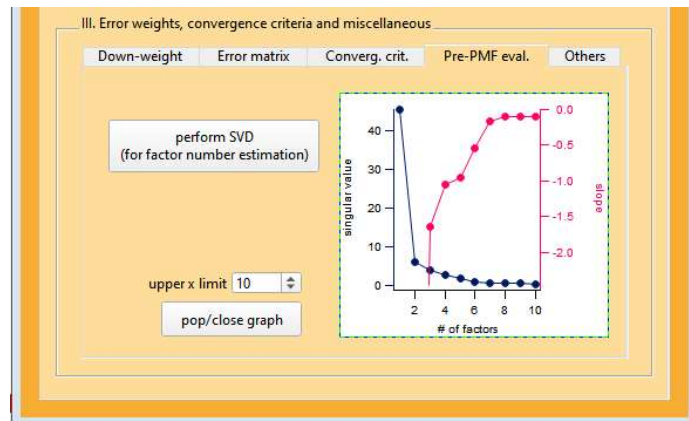


图 44: 44

6.2.7. 其他各种选项

6.2.7.1. Robust 模式阈值 (A)

PMF 在 Robust 模式下运行，以保护 PMF runs 不因异常值而偏离（见 2 图 45

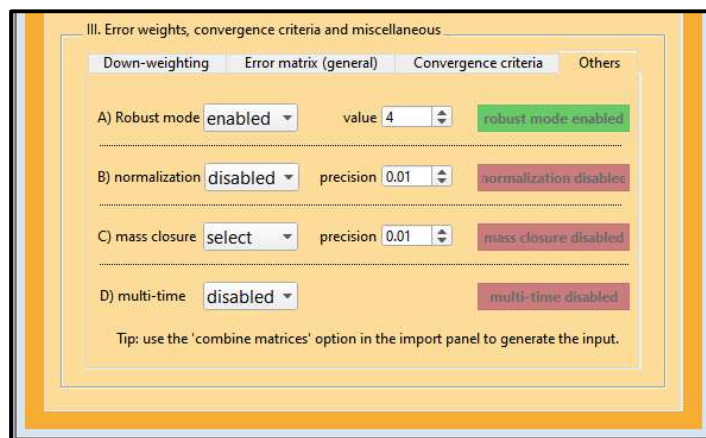


图 45

6.2.7.2. 模型归一化 Model normalization (B)

用户可以选择在迭代过程中每个因子配置文件的总和是否归一化为一（启用）或因子源谱文件保持未归一化（默认情况，禁用）。

6.2.7.1. 质量闭合 Mass closure (C)

可以选择将所有因子贡献的总和拟合到某个粒径质量，例如 PM_1 、 $PM_{2.5}$ 、 PM_{10} ，这可以通过弹出菜单进行选择。质量波必须存在于“root:SoFi:External_data:Reference_ts:raw”下（请参阅部分 5.2.2（图 45

仅在质量差异非常小的情况下才应使用此选项。用户不应定期拟合“不可见”的质量。还建议在不拟合缺失质量的情况下运行模型，并比较质量差异。

6.2.7.1. 多种时间分辨率 Multi-time (D)

多时间特征提供了在包含两个及以上类别的数据集上运行 PMF 算法的可能性，这些数据在时间分辨率上有所不同。(Ogulei 等, 2005)。目标是利用每类仪器的原始时间分辨率，而不是事先强制平均，相反 PMF 会考虑并在执行 PMF 时注意不同的分辨率。解决的 PMF 方程是(1)

$$x_{sj} = \frac{1}{(t_{s2} - t_{s1} + 1)} \sum_{k=1}^p \left(f_{jk} \sum_{i=t_{s1}}^{i=t_{s2}} g_{ik} \right) + e_{sj} \quad (19)$$

其中 x_{sj} 表示仪器样本 s 的模型矩阵 X 的单个点，以及变量 j ； t_{s1} 和 t_{s2} 分别是样本 s 的开始和结束时间； f_{jk} 表示因子 k 中变量 j 的质量分数； g_{ik} 表示因子 k 在时间点 i 的质量贡献； e_{sj} 表示残差矩阵 E 中样品 s 和变量 j 对应的未被模型解释部分。

具有最高时间分辨率的样本表示 X 中的主要方程。对于主要方程， t_{s1} 和 t_{s2} 是相同的，内部求和消失，开始时的归一化分数也消失。对于其他样本， x_{sj} 点表示为在采样期间存在的主要方程中所有点的平均值，即在 t_{s1} 和 t_{s2} 之间，如方程所述。(19)

如果低时间分辨率类别的测量存在一些测量缺口，那么传递这些类别的开始和结束时间信息是极其重要的。为了传递此信息，用户必须将 PMF 输入时间序列作为二维矩阵传递，其中第一列包含开始时间，第二列包含结束时间。请注意，主要/默认类别的开始时间不会使用，因此对于该类别，时间序列输入波可以（如果对用户更方便）包含两次相同的信息。

为了执行多时间分辨率的 PMF 时，用户应准备并读取 SoFi 中的输入文件和误差矩阵，如下图所示：

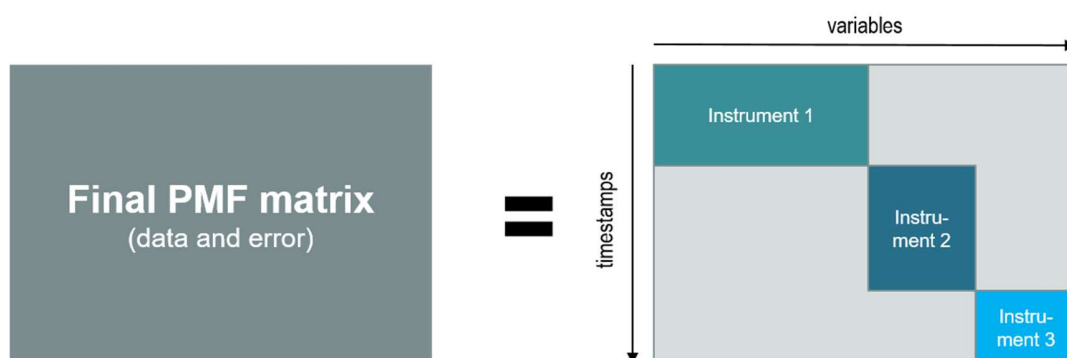


图 46: 多时间分辨率的 PMF 的输入数据和误差矩阵

其中样本 1 具有最高的时间分辨率，而样本 2、3 等始终具有较低的时间分辨率。最终矩阵中的空白部分（右图）都是“nans”。所有样本的时间序列以其本地时间分辨率传递。这意味着时间序列波将来回移动，因为样本 2、3 等应覆盖与样本 1 相同的时间范围。因此，SoFi 将找到适当的索引值并将其传递给 ME-2，以便对样本 2、3、4 等的 x_{sj} 方程进行额外的平均，如示例中所给。图 466.4.1

为了准备对角矩阵，用户可以在导入面板中使用“合并矩阵 **combine matrices**”选项（[4.1.4 用户应将数据和误差矩阵中的非对角元素定义为缺失值](#)（请参阅 [6.2.25.4.1](#)

之前描述的调整因子 η_j 作为波“调整因子”报告在相应的结果文件夹中：“root:SoFi:Results:HDF filename:Solutions:run_X”，以便进一步手动检查。请注意，接近 1 的调整因子表明不同测量之间的良好一致性。

6.3. 为模型矩阵定义选项

6.3.1. 种子 seed

种子选项用于填充模型矩阵 **G** 和 **F**。随机初始化通常探索局部最小值的数量，其中最终的 **Q** 值仍然很高，相比于全局最小值或 **Q** 空间中几乎相同的最小 **Q** 值区域（旋转多义性 rotational ambiguity，见 2.3

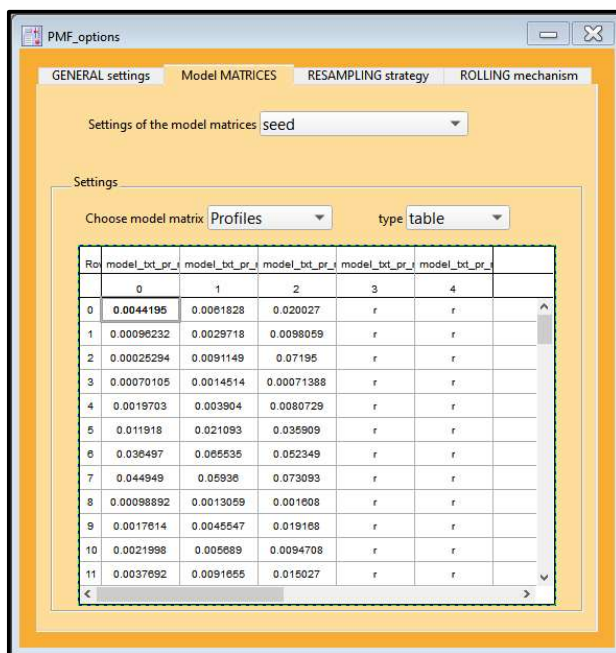


图 47

在“模型矩阵 **Model MATRICES**”标签中的表格对于选择“种子 **seed**”显示了所选模型条目的概述，适用于矩阵 **F** “源谱 **Profiles**”或 **G** “时间序列 **time series**”（图 47

6.3.1.1. 设置

“因子 **Factor**”

应设置种子值的位置 - 可以选择所有因子或单个因子位置，例如因子 **1**。

“初始模型值 **Initial model values**”

设置种子值的选项，可以是以下选项之一：

随机 **random** (**r**)：始终完全随机的条目（默认）

伪随机 **pseudo-random** (**pr**)：在 **PMF** 调用开始时随机初始化，并在属于同一 **PMF** 调用的所有其他 **runs** 中使用相同的随机值

零 **zero (0)** : 将整个因子/选定变量的种子值设置为零（在没有外部参考数据的情况下工作）

来自向量 **from vector (fv)** : 解矩阵 F / G 使用来自文件夹 “root:SoFi:Reference_ts:raw 或 root:SoFi:Reference_pr” 的外部信息进行填充。请参阅 5.2

“变量 **Variables**”

是否应为:所有变量（所有条目 **all entries**）、外部向量的所有变量或单个变量设置种子值

“归一化源谱变量 **Normalize factor var**” 用于决定作为初始值的向量是否进行归一化（设置下面的总和因子，默认值为 1）或不被归一化。

“写入模型矩阵 **Write into model matrix**” 更新模型矩阵，主 PMF 设置窗口中的表格将相应更新。

“重置 (...) 矩阵 **Reset (...) matrix**”

重置为随机到整个矩阵或当前因子位置

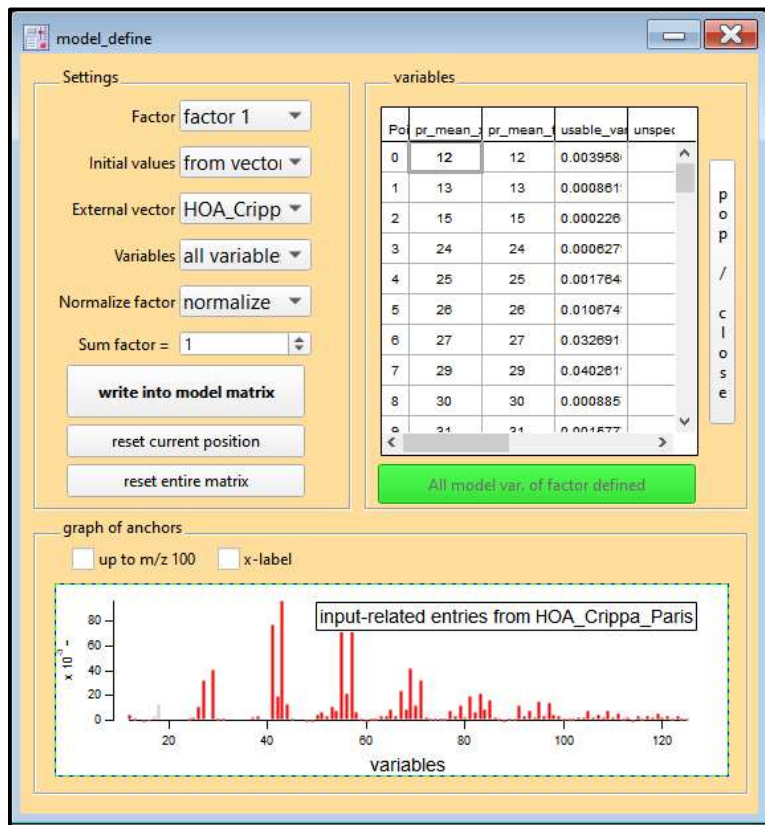


图 48: 从向量获取初始模型值的面板。HOA_Paris 源谱文件归一化为 1，将写入配置文件矩阵 F 的因子位置 1。

6.3.1.2. “从向量 (fv) from vector (fv)” 的选项

选择“从向量 (fv)”时，面板“model_define”（图 48 向模型矩阵写入信息是累加的。如果矩阵 F / G 的向量包含一些未定义的条目，则按钮“某些因子的变量仍未指定 **Some var. of factor still unspecified**”保持为黄色。一旦所有条目都被定义，该按钮变为绿色“所有因子的模型变量已定义 **All model var. of factor defined**”。

用户可以选择将条目归一化为“因子变量总和 **Sum factor var.**”下提供的值。此选项与写入的信息相关，而不是与模型矩阵中存在的最终向量相关。

示例

用户希望将一个源谱向量（矩阵 F）的总和归一化为 1，并且该向量的信息存储在两个独立的外部向量 A 和 B 中。然后应进行以下加权计算：

$$w \cdot (\text{sum}(\mathbf{A}) + \text{sum}(\mathbf{B})) = 1 \quad (20)$$

w 代表权重值。

$$\text{nor.value}(\mathbf{A}) = w \cdot \text{sum}(\mathbf{A}) \quad (21)$$

$$\text{nor.value}(\mathbf{B}) = w \cdot \text{sum}(\mathbf{B}) \quad (22)$$

其中和为在“model_define”面板下“Normalize factor var.”项中输入的值（*nor.value(A)* *nor.value(B)*）图 48

从“seed”选项将一个或多个外部向量的信息写入模型矩阵 G 或 F 并不意味着这些信息受到约束。这些信息仅表示 PMF runs 的起始条件。约束这些信息或基于这些信息派生的数据是通过以下两种技术进行的：“a-value”或“pulling equations”，见 6.3.36.3.4

6.3.2. 峰值 fpeak

以下 fpeak 选项可用：

“禁用 **disabled**” fpeak 旋转被禁用。

“全局 **global fpeak**” 启用影响整个模型矩阵的旋转。

“个体 **individual fpeak**” 启用仅影响选定贡献、因子源谱的旋转。

对于这些 fpeak 选项，支持以下类型的分析：

“精确 **exact fpeak**” 使用精确的 **fpeak** 值。

“灵敏度 **sensitivity fpeak**” 在用户定义的 **fpeak** 分辨率下，系统地测试初始和最终 **fpeak** 之间的 **fpeak** 值。

“随机 **random fpeak**” **fpeak** 值在用户定义的 **fpeak** 分辨率下，在初始和最终 **fpeak** 之间随机测试。

6.3.2.1. 全局 **Global fpeak**

全局 **fpeak** 和个体 **fpeak** 是之前在 2.4 图 49

示例

可以探索导致标准化缩放残差 (Q/Q_{exp}) 增加 10% 的 **fpeak** 范围。一些过去的研究仅调查了预定义的 **fpeak** 范围，例如 **fpeak** 值在 -1 和 1 之间，而不考虑其对 Q 的影响。这种方法应避免，因为 **fpeak** 旋转是模型的一个函数，因此， Q/Q_{exp} 将根据输入数据和模型结果而变化。

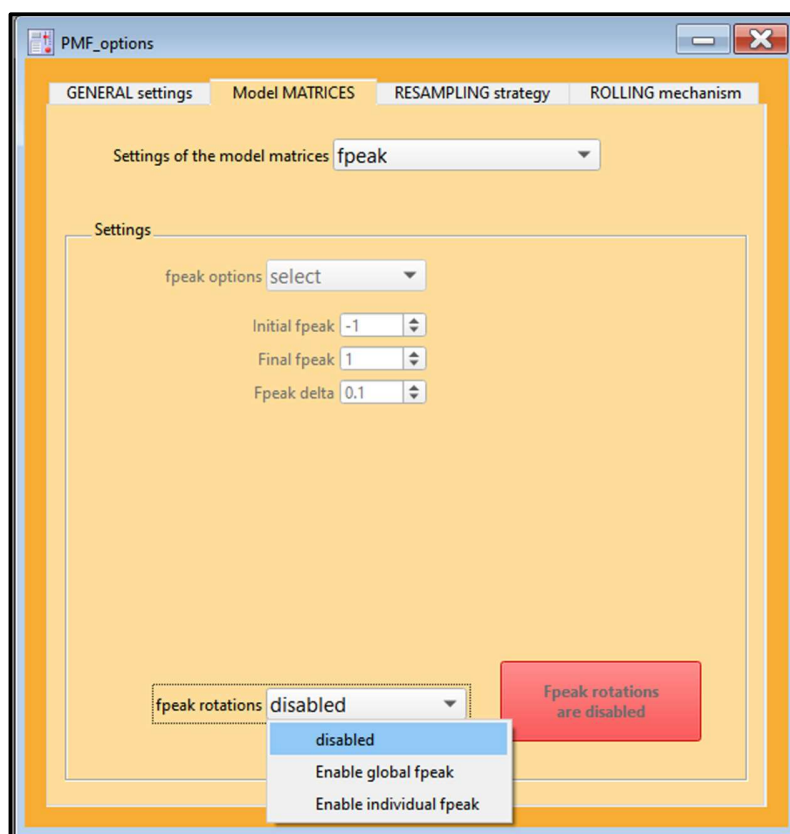


图 49

选择“全局 **fpeak**”或“个体 **fpeak**”中的一个选项，可以启用 PMF runs 的 **fpeak** 旋转。

6.3.2.2. 个体 fpeak

用户定义初始、最终 fpeak 和 fpeak 步长间隔，与全局 fpeak 相同。在矩阵的单元格中 (图 50 从概念上讲，它可以被视为一个标量，用于将一个因子的时间序列（基于 fpeak 的符号）加到或减去另一个因子的时间序列。该符号遵循线性代数的规则。

示例

例如，如果条目 $T[0,1] = -5$ ，这意味着因子 1 的时间序列从时间序列 2 中减去大小为-5 的值。反之，因子 2 的因子源谱以相同的大小和相反的符号 5 加到因子 1 的因子源谱上。

然而，T 中的 0 条目并不妨碍加法/减法的发生，而是表明如果发生这种旋转，模型既不偏好也不惩罚。

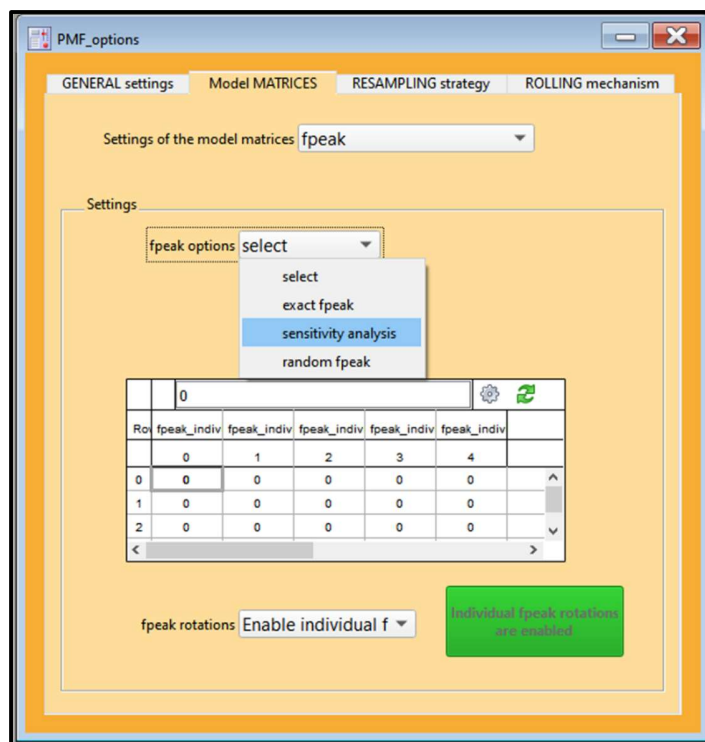


图 50

6.3.3. a 值约束

a 值只能在种子值定义后选择。

a 值选项允许约束一个或多个 F 源谱矩阵或 G 时间序列矩阵的条目（见 2.4.4 图 51 图 52 要启用 **a** 值约束，请使用表格下方的下拉菜单。只有当按钮“**A** 值约束已启用 A-value constraints are enabled”变为绿色时，ME-2 求解器才会使用 **a** 值（图 50

与种子的位移结合的 **a** 值选项



“因子 Factor”

所有因子或单个因子位置，例如因子 1

“约束类型 **Type of constraint**” “精确 a 值 **exact a-value**”

“第一维灵敏度 1st dim. sensitivity”

“第二至第五维灵敏度 2nd - 5th dim. sensitivity”

“随机 random”

“直接将限制传递给 ME-2 pass limits directly to ME-2”

“变量 Variables”

所有模型变量 all model variables, 所有外部变量 all variables of external variables 的所有条目 all entries, 单个变量 single variables, 例如 m/z44

用户可以选择使用哪个变量基础来传递约束的信息。约束变量是累加的, 正如定义变量时所示 (见 6.3.1)

关于选项的更多细节如下。

6.3.3.2. 精确的 a 值 (精确)

使用精确的 a 值, 例如 0.1 来约束定义的信息 (源谱文件/时间序列)

6.3.3.1. 第一维敏感性 (1 维敏感性)

a 值从“初始 a 值 initial a-value”系统性地变化, 使用“a delta”给出的步长间隔, 直到“最终 a 值”。这允许研究 PMF 结果作为一维中 a 值的函数。

如果您在一维中错误地定义了两个不同的敏感性分析设置, SoFi 将仅使用首次遇到的设置。

示例

第一个因子的最终 a 值为 1, 第二个因子的最终 a 值为 0.5, 并且两个因子都有约束: 一维敏感性分析。对于一维敏感性分析, 传递给 ME-2 的最终 a 值将是 1, 因为第一个因子在第二个因子之前被内部评估。

6.3.3.2. 第二维灵敏度 (2 维灵敏度) 对第五维灵敏度 (5 维灵敏度)

如果需要单独测试 a 值组合, 则应使用此选项。a 值从“初始 a 值”开始, 按照“a 增量”给定的步长系统性地变化, 直到“最终 a 值”。这允许研究 PMF 结果作为 a 值在第二维到第五维的函数。

如果只需在一个维度中进行灵敏度分析, 则必须选择第一个维度 (见上文)。使用更高维度并保持先前维度为空或未定义会在 SoFi 中生成错误消息。

示例

第一个因子的最终 a-值为 1, a-增量为 0.2, 并执行一维灵敏度分析。第二个因子的最终 a-值为 0.5, a-增量为 0.1, 并与二维灵敏度分析相关联。所有可能的 a-值组合将被运行, SoFi 将执行

$$\left(\frac{1}{0.2} + 1\right) \cdot \left(\frac{0.5}{0.1} + 1\right) = 36 \text{ runs} \quad (23)$$

6.3.3.3. 随机 a 值 (随机)

随机 a 值是从零到“最终 a 值”之间的均匀分布中传递的, 精度由“a-delta”给出。

6.3.3.4. 直接将限制传递给 ME-2（限制）

如果对于特定来源，例如交通排放污染源，已知源谱文件的范围，则可以使用此选项将该范围（限制）直接传递给 ME-2。提供了表示源谱文件/时间序列的下限和上限的两个向量，ME-2 在 PMF runs 期间在这些限制内优化源谱文件。请注意，种子向量也必须被定义。在定义相应的种子信息时，使用较低或较高的向量作为起始条件。

6.3.3.5. 所有模型变量

此选项以模型矩阵 F / G 中定义的变量为基础。这导致所有变量具有相同的 a 值约束。

6.3.3.6. 外部向量的所有模型变量

此选项以外部向量的变量为基础。当在一个受限向量中需要使用多个 a 值时，这很有用（来自不同的外部配置文件），例如，如果受限向量中的某些变量比其他变量更准确，或者需要受限的配置文件不包含 PMF 输入中存在的所有离子，而缺失的离子也应该以较松的 a 值进行约束。在这种情况下，较不准确的变量可以获得更高的 a 值。

类似于“model_define”面板（图 48

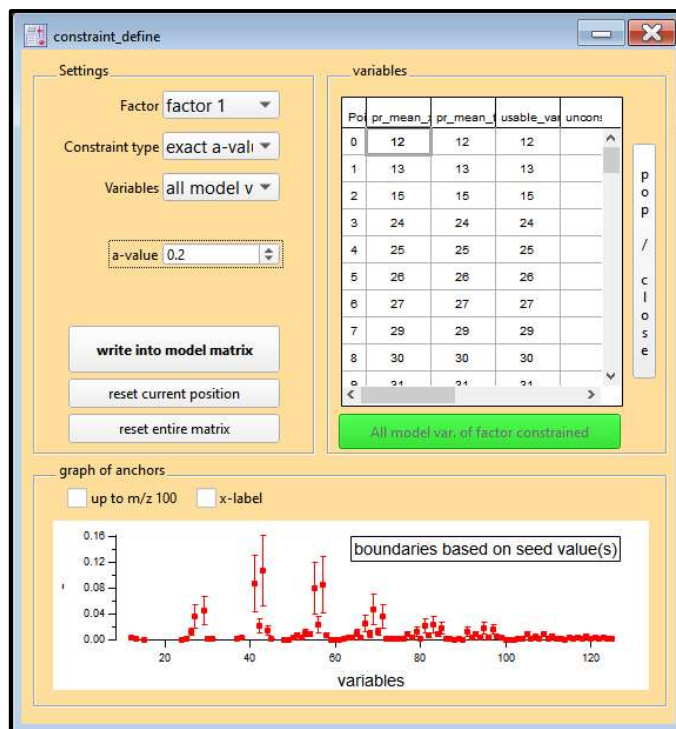


图 52 图 48

按钮“写入模型矩阵 write into model matrix”将写入所选设置，用户可以在主面板上监控更改（图 51 图 52

建议约束整个源谱文件/时间序列。如果一个因子仅部分被约束，其未约束部分可能在 PMF runs 期间增长，并在 SoFi 的归一化过程中变得主导。

6.3.4. 位移 Displacement

SoFi 中的位移方法允许用户在多达五个独立维度中引入选定锚元素的随机和/或系统性扰动。最初，锚值根据所选的 a 值约束进行修改。然而，一旦应用了位移，这些锚点将保持固定（“冻结”），确保模型探索围绕定义配置的变化，而不是不断重新调整约束。位移有助于探测解对关键锚点位置的小幅、可控变化的敏感性。由于锚点在位移后被冻结，存在过度约束模型的风险，或仅探索有限的旋转空间。如果所选的位移不具代表性，相关的替代解可能仍会被遗漏。

6.3.5. 拉取方程 Pulling equations

拉取是 ME-2 中用于将先验信息引入因式分解的另一种方法。除了拟合测量数据外，模型还可以朝着预定义的值或变量之间的关系进行引导，这些变量存在于配置矩阵 F 或时间序列矩阵 G 中（见 2.4.5

这通过在优化过程中添加额外的方程来实现，这些方程会惩罚与指定目标的偏差。因此，拉取允许用户在仍然允许模型根据测量数据调整解的同时，结合外部知识，因此被视为一种软约束技术。

拉取方程可以手动编写或借助 SoFi 进行编写。ME-2 求解器对拉取方程的形式及其限制非常敏感，因此应仔细定义这些方程。要启用拉取方程，请使用表格下方的下拉菜单。只有当“拉取方程已启用 Pulling equations are enabled”按钮变为绿色时，ME-2 求解器才会使用拉取方程（图 53

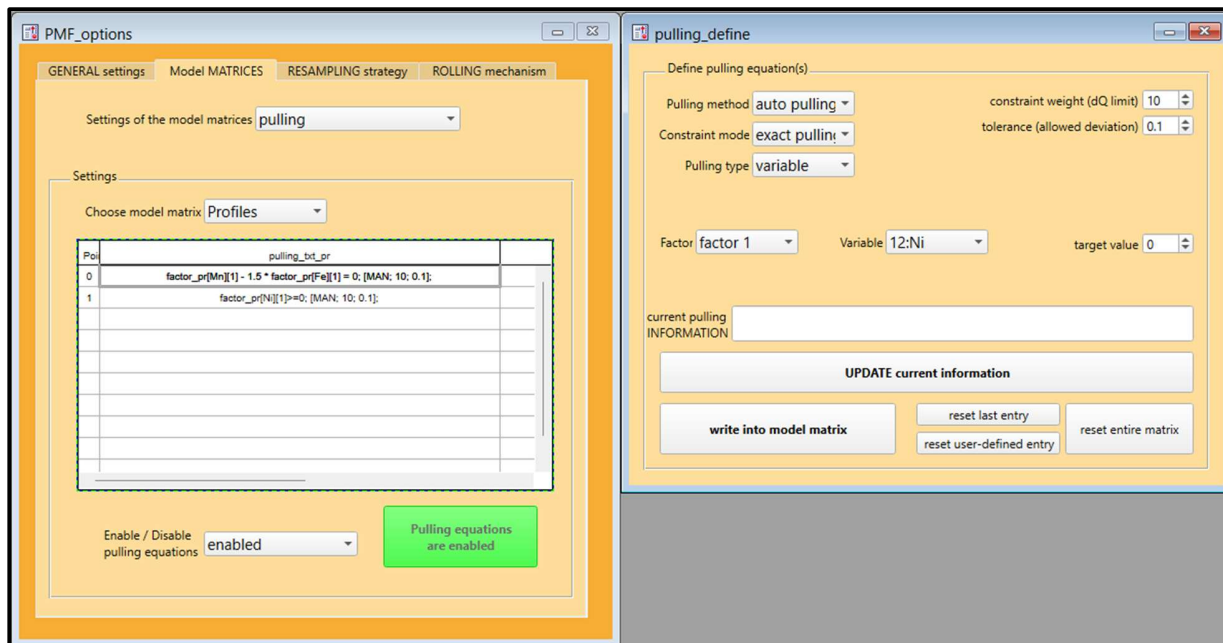


图 53

SoFi 提供了多种不同的拉取方程选项。它们在下面进行了描述。并非所有的数学选项组合都是相互兼容的。此外，某些信息必须手动添加。

6.3.5.1. 拉取方程 Pulling equation

SoFi（以及 ME-2 求解器）区分两种类型的拉取：自动拉取 **automatic pulling** 和手动拉取 **manual pulling**。在自动拉取中，用户指定一个目标值以及约束强度和容差。强度决定了偏离目标的程度对优化的影响，而容差定义了围绕目标值的可接受范围。这种方法提供了一种灵活的方式来引导解，而不严格执行约束。

在手动拉取中，约束被视为具有定义不确定性的额外方程。用户指定目标值和预期偏差，模型在优化过程中最小化计算值与目标之间的差异。因此，手动拉取通常在有可靠的先前信息可用时使用，并且希望有更明确的约束。

6.3.5.2. 拉取类型 Pulling type 和后续拉取方程 pulling equation

拉取信息的第一部分专门用于实际的拉取方程和目标值。根据要拉取的内容类型，拉取方程可能会略有不同，但主要格式是：

$$\text{要提取的模型变量} = [\text{提取值}];[\text{提取强度}];[\text{提取不确定性}] \quad (24)$$

为了方便用户使用，SoFi 提供了提取单个变量、总和、差异和比率的模板。特别是对于比率，用户应该利用这些模板，因为 ME-2 求解器的方程语法相当严格，比率必须写成两个项之差等于零的差异，提取方程中不允许/不被识别除法，ME-2 求解器只接受项的总和或差异，其中项可以是单个项或多个项乘积（图 54

对于手动拉取，仅允许等式，对于自动提取，求解器也接受不等式的提取方程（ $\geq$ 和 $\leq$ ），因为 ME-2 求解器可以访问不同的所谓误差模型，允许单侧拉取，而手动拉取仅允许对称拉取。

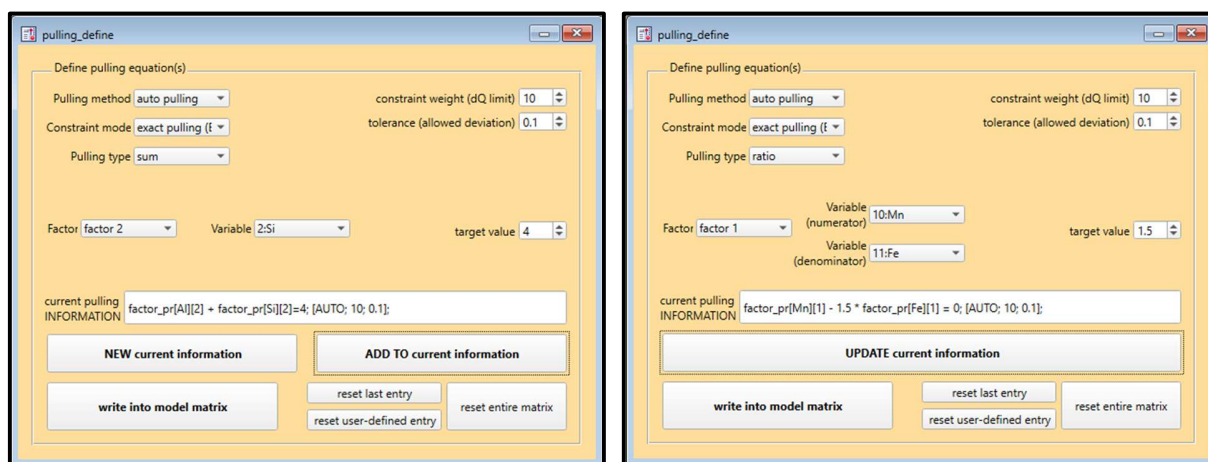


图 54

6.3.5.3. 拉取强度 Pulling strength

拉取约束对解的影响由其相关参数控制，因此拉取信息的第二部分专门用于设置这些拉取细节，即拉取方法（自动或手动 AUTO/MAN）、权重/强度 weight/strength 和容差/不确定性 tolerance/uncertainty。

约束权重/强度决定了模型满足指定约束的强度。增加该值会增加约束对目标函数 (Q) 的贡献，从而迫使解更接近目标值。低值会导致约束较弱，对解影响较小，而非常高的值可能会主导优化并减少测量数据对结果的影响。

对于自动拉取，容差定义了与目标值的可接受偏差。较小的容差会更严格地执行约束，而较大的容差则允许更大的灵活性。容差通常应根据目标值的预期变化程度或不确定性进行选择。对于手动拉取，不确定性起着类似的作用。它表示约束的预期误差，并决定偏差受到惩罚的强度。较小的值对应于更严格的约束，而较大的值则减少约束的影响。

6.3.5.4. 约束模式 Constraint mode

与 a 值方法类似，SoFi 提供了多个维度的精确、随机和自动灵敏度分析。这可以应用于目标值 target value 和/或约束权重/强度 constraint weight/strength 以及约束容差/不确定性 constraint tolerance/uncertainty。对于任何灵敏度分析或随机拉取值，符号如下：

[“类型”，“初始值”，“最终值”，“分辨率”] (25)

这三个参数，即目标、权重和容忍度被视为独立参数。因此，用户可以根据需要对这三个参数进行独立的灵敏度分析。这导致了三个独立的灵敏度分析维度，目标作为第一维；权重作为第二维；容差作为第三维。

6.3.5.4.1. 精确拉动 *Exact pulling (EXT)*

在迭代过程中，使用精确的目标值来拉动当前信息（配置文件/时间序列）朝向这个锚值。无需定义方括号，为了便于阅读，精确值直接传递给拉动信息。

6.3.5.4.2. 第一维灵敏度分析 *1st dimension sensitivity analysis (1D)*

用户定义的参数（目标或权重或容忍度）从初始值系统地变化，使用用户定义的步长间隔到达最终值。这允许研究 PMF 结果作为拉动信息在一个维度上的函数。

对于目标值，用户可以使用菜单设置最小值 min.和最大值 max.以及步长 step size。对于权重/强度和容忍度/不确定性，用户可以选择将模板添加到拉动信息中，并需要在其中手动更改信息（图 55

如果您在同一个维度中错误地定义了两个不同的灵敏度分析设置，SoFi 将只使用第一个遇到的设置。

图 55

6.3.5.4.3. 第二维灵敏度 *2nd dim. sensitivity (2D)* 到第五维灵敏度 *5th dim. sensitivity (5D)*

用户定义的信息（目标或权重或容差）从初始值系统地变化，使用用户定义的步长间隔到达最终值。这允许研究 PMF 结果作为第二维到第五维拉动信息的函数。

如果只需在一个维度中执行灵敏度分析，则必须选择第一个维度。使用更高维度并保持先前维度为空或未定义会在 SoFi 中生成错误消息。

示例

第一个方程的最小目标值为 0，最终目标值为 1，步长为 0.2，并执行一维灵敏度分析。第二个方程具有最小值和最大值分别为 10 和 50 的权重，步长为 10，并与二维灵敏度分析相关联。SoFi 将执行

$$\left(\frac{1}{0.2} + 1\right) \cdot \left(\frac{40}{10} + 1\right) = 30 \text{ runs} \quad (26)$$

6.3.5.4.4. 随机抽取 (RND) Random pulling (RND)

随机抽取值来自于初始值和最终值之间的均匀分布，分辨率为增量值 delta-value。

6.3.5.5. 写入拉取信息

一旦所有拉取参数被定义，拉取信息需要通过“写入模型矩阵 **write into the model matrix**”按钮写入模型矩阵。用户还可以选择从该矩阵中删除最后一条或特定的抽取信息，以及重置整个矩阵。

在运行 PMF 之前，请不要忘记启用抽取方程。

6.3.1. 因子名称自定义 Factor specifications

每个因子接收默认名称“因子[位置] factor [position]”。用户可以在“因子规格 factor specs”下更改这些名称（图 56

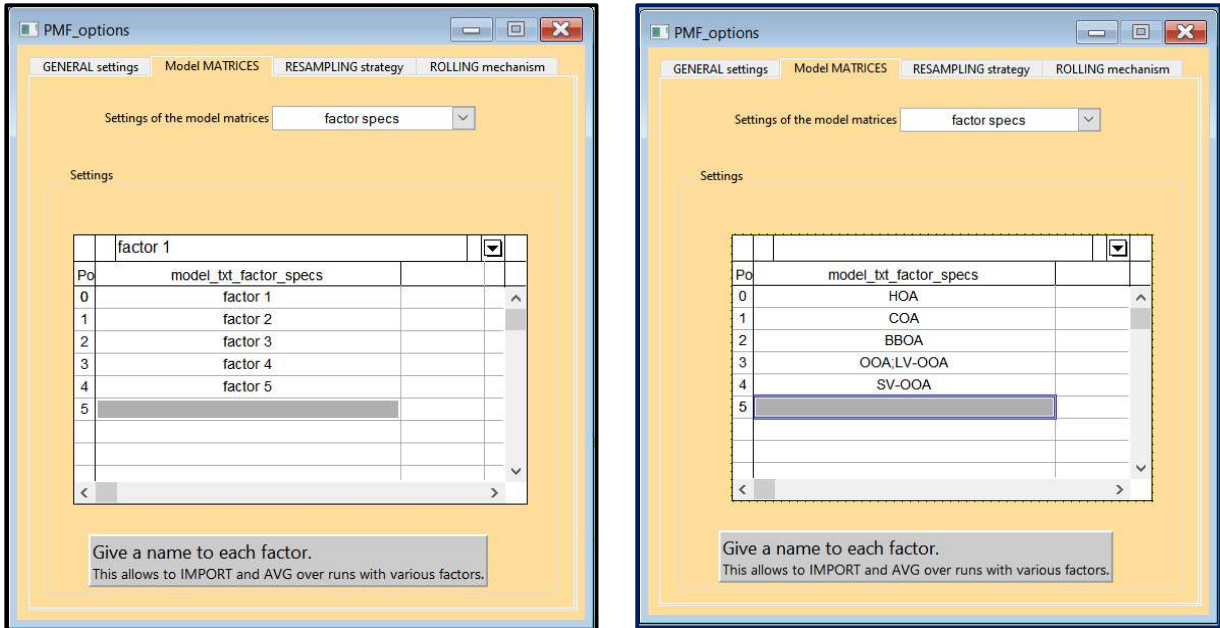


图 56

在 SoFi 中平均 PMF 结果时（见 7.5.1

如果一个 HDF 文件包含多因子数的 PMF 结果，并且同一位置的各种因子解不一定需要平均在一起，那么用户可以为一个因子位置指定不同的名称。因子名称不应包含“+”，否则 SoFi 会错误地将其视为两个应相加的因子。

示例

将三个不同的 HDF 文件进行平均，其中只有 HOA 和 COA 是它们之间的共同因子。

表 4: 4

HDF 1	HDF2	HDF3	SoFi 中的平均时的解
HOA	HOA	HOA	所有文件的 HOA
COA	COA	COA	来自所有文件的 COA
BBOA	SV-OOA	BBOA	来自 HDF1 和 2 的 BBOA
OOA	LV-	SV-OOA	来自 HDF 1 的 OOA
		LV-OOA	来自 HDF 2 和 3 的 SV-OOA
			来自 HDF2 和 3 的 LV-OOA

示例

一个包含 4 因子和 5 因子 PMF runs 的 HDF 文件，其中 4 因子运行的第 4 个位置代表的内容与 5 因子运行的第 4 个位置不同。在对 4 因子和 5 因子解进行平均时，SoFi 将为 OOA 和 LV-OOA 使用专用因子位置。这导致平均解为六因子解，因为 OOA 和 SV-OOA 彼此分开。

表 5: 5

HDF	SoFi 中的平均时的解
HOA	HOA
COA	COA
BBOA	BBOA
OOA;LV-OOA	OOA
SV-OOA	SV-OOA
	LV-OOA

如果在运行 PMF 之前没有执行此操作，因此用户定义的因子名称信息未存储在 HDF 文件中，用户也可以在平均解之前手动更改因子名称。因子规格列表可以在这里找到：

root:SoFi:Results:"NAME_OF_HDF_FILE":Parameter:model_txt_factor_specs

6.3.2. 考虑旋转工具的优先级

旋转多义性的多个探索性技术是可叠加的。除了其他技术（如 a 值或拉动方程或 C 值方法等），还可以启用 fpeak 旋转。例如：如果结果值超出了 a 值约束给定的限制，则 fpeak 旋转的优先级最低，即不允许旋转。

6.4. 统计不确定性的估计

最终的 PMF 结果/解必须也包含统计不确定性的定量度量。对于环境气溶胶数据，统计不确定性的主要贡献者主要是排放源的随机变化。过去的研究（(Paatero et al., 2014) 和 (Ulbrich 等, 2009) 通常采用“bootstrap”的重采样策略(Davison 和 Hinkley, 1997) 旨在解决和量化 AMS 数据中的这种错误。

6.4.1. Bootstrap 重采样策略

Bootstrap 通过从原始数据集中随机重采样生成一组新的输入矩阵进行分析 (Efron, 1979)。这种重采样策略涉及复制一些数据点，同时省略其他数据点，这有效地扰动了输入数据（见图 57

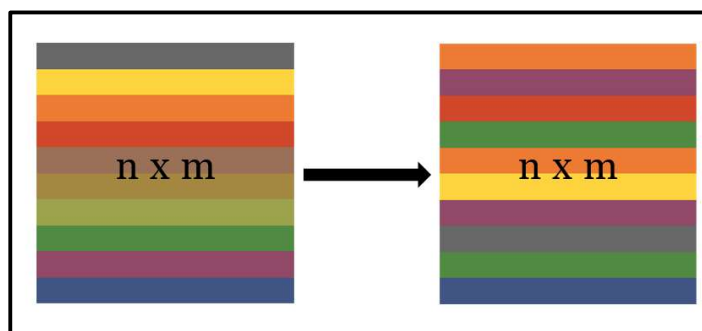


图 57: 默认 PMF 输入矩阵的示意图（左）和重采样的矩阵（右）。颜色表示重采样的行。请注意，深绿色出现了两次，而默认 PMF 矩阵中的橄榄绿色没有出现。

主要设置

“未锁定 **unblocked**”的单个时间点被随机采样

“使用定义类别 **using defined classes**”所有属于同一类别的点被分组（锁定）在一起。

请查看 5.4.1

在未锁定的 **bootstrap** 中，个别数据点是从原始数据集中随机抽样并替换，以创建新的重抽样数据集。每个新数据集包含与原始数据集相同数量的数据点，但某些数据点可能出现多次，而其他数据点可能根本不出现。在对 PMF 模型应用 **bootstrap** 策略时，一个关键点是必须有足够的变化性来识别提取预期的源。如果例如，数据覆盖了一个特殊时期，如烟花事件，可能会发生在重抽样后，反映这一特定时期的点最终缺失，且可用于区分特定源的变化性很少。锁定的 **bootstrap** 解决了标准 **bootstrap** 在时间依赖数据上的局限性。它不是单独重抽样数据点，而是将数据块一起重抽样。这些数据块可以由用户定义（5.4.1

bootstrap 重复次数在确保不确定性估计的稳健性和可靠性方面起着关键作用。较多的重复次数通常能提高对识别因子不确定性评估的准确性。这是因为每次重复都会引入数据的不同样本，有助于捕捉数据集中可能存在的更广泛的变化程度。**bootstrap** 重复次数不足可能

导致误导性的不确定性估计，因为重采样的数据集可能无法捕捉所有相关的数据结构或变化程度，尤其是在复杂或噪声较大的数据集中。**bootstrap** 法重复次数由“GENERAL settings”下的 **PMF** 调用次数控制。对于标准 **PMF** 分析（静态 **PMF** 输入），通常建议执行至少 500-1000 次 **bootstrap** 重复，但最终选择适当的重复次数往往是在计算的可行性与准确不确定性估计需求之间的权衡。在滚动 **PMF** 中，由于数据窗口较小且在数据集中滚动，因此每个窗口的 **bootstrap** 重复次数可以低于标准 **PMF**，这也导致每天有额外的重复次数。（Chen et al., 2022）。

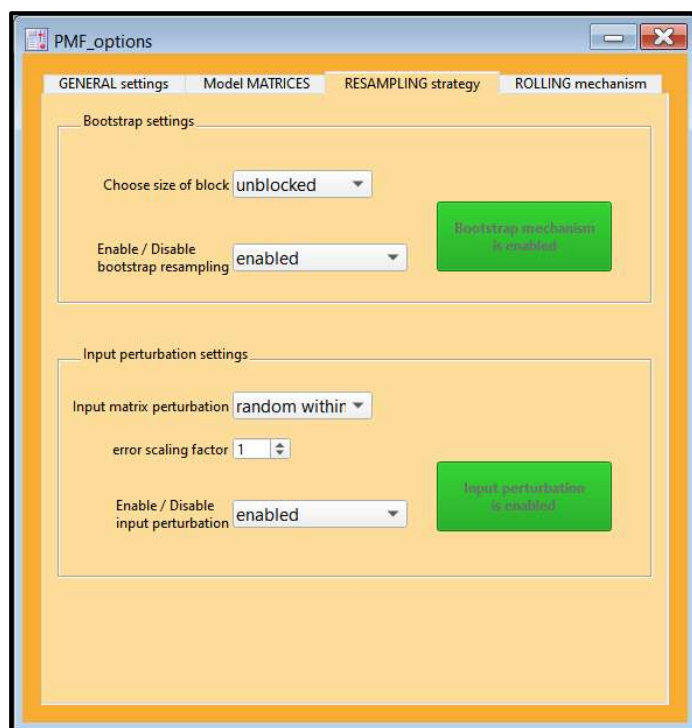


图 58

6.4.2. 输入扰动 Input perturbation

输入扰动将随机扰动输入矩阵，扰动值在 $\pm x * \text{误差矩阵范围}$ 内。缩放因子 x 可以由用户定义，默认设置为 1，以便每个值在 $\pm$ 其误差范围内随机变化。该技术用于评估 **PMF** 解对输入数据变化的敏感性，并估计分解结果的稳健性和不确定性。通过多次扰动输入数据，可以生成一系列 **PMF** 解。这些解之间的变化程度有助于估计因子特征和贡献的统计不确定性。这提供了对识别源的稳健性更现实的看法，考虑到输入数据固有的不确定性。

6.5. 滚动 rolling 机制

PMF 的主要限制是建模静态因子特征，这与动态系统相矛盾，后者的源特征应该随时间变化。为了考虑因子特征可能的时间变化，建议采用滚动方法（Parworth et al.(2015), Canonaco et al. (2021)）用于环境 AMS/ACSM 数据。

与传统的 PMF 不同，传统 PMF 是一次性分析整个数据集，而 rolling PMF 将 PMF 模型应用于在时间轴上移动的小的重叠数据窗口。这种方法允许跟踪因子贡献如何随时间演变（见图 59）

- PMF 窗口必须比整个数据集小，例如，如果总的 PMF 输入包含数年数据，则窗口可以是几天/几周。更重要的是，窗口长度应选择为使静态因子特征的假设是合理的。
- 为了避免连续窗口之间的剧烈过渡，重要的是要重叠窗口，这确保了相邻窗口中识别的因子随着时间的推移平稳演变，而不是从一个窗口到下一个窗口突然变化。

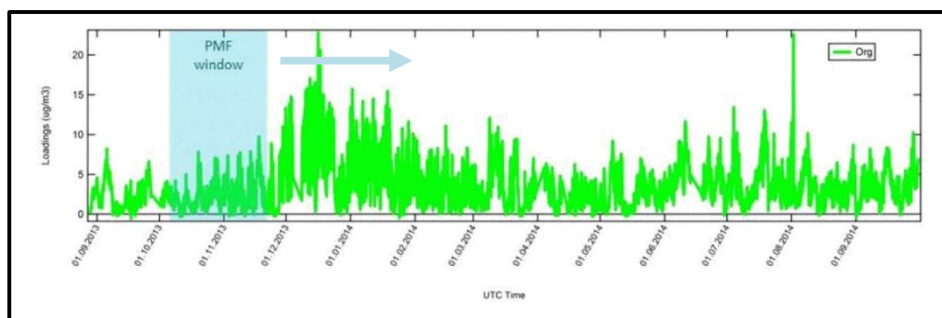


图 59: Rolling PMF 窗口的示意图。

主要设置

“窗口宽度 **window width**” 以天为单位表示的 PMF 窗口大小

“窗口移动 **window shift**” 以天为单位表示的 PMF 窗口移动距离

“重复次数 **nb. of repeats**” 根据窗口重叠而估计的重复分析次数

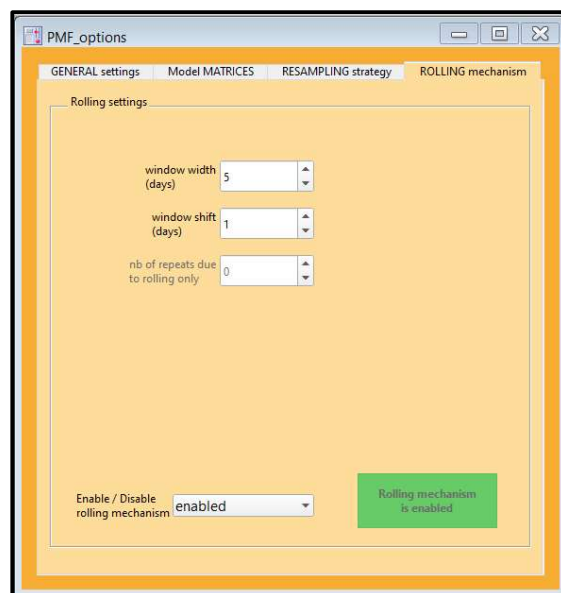


图 60

理想情况下，滚动 PMF 方法与 bootstrap 重采样策略相结合，以估计滚动 PMF 中的不确定性。由于每个窗口包含的数据点少于完整数据集，因此通常需要较少的 bootstrap runs。此外，如果任何因子受到限制，滚动 PMF 方法还应与随机 α 值策略相结合。有关更多详细信息，请参阅 Canonaco 等人 (2021) 和 Chen 等人 (2022) 及其中的分析。

6.6. 调用 ME-2 并执行 PMF

SoFi 调用 ME-2，并将 PMF 输入与当前指令文件以及定义和约束的信息（如果存在）一起传递给 ME-2。

在调用 ME-2 求解器之前，SoFi 会简要显示将执行多少次 PMF 调用，以及用户是否希望继续。进度可以通过主面板末尾的进度条“ME-2 调用 ME-2 call”进行监控（图 34

7. SoFi – PMF 结果

7.1. 导入结果文件（HDF 文件）

7.1.1. 定义 ME2.exe 文件的路径

用户通过按下“可执行文件的文件夹 folder of executable”按钮来提供 ME2.exe 文件所在的位置（图 61 如果“PMF result”标签页与之前的标签页完全独立，即使这些结果是从另一个 IGOR 实验或另一台 PC 生成的，也可以进行查询。唯一的限制是由操作系统下必须存在结构正确的 ME-2 文件夹。如果用户想要检查在另一台 PC 上运行的一些 PMF 结果，则包含 PMF 结果的 HDF5 文件必须存储在“ME2_engine”文件夹的“ME2_Results”中：... \ME2_Engine\ME2_Results。

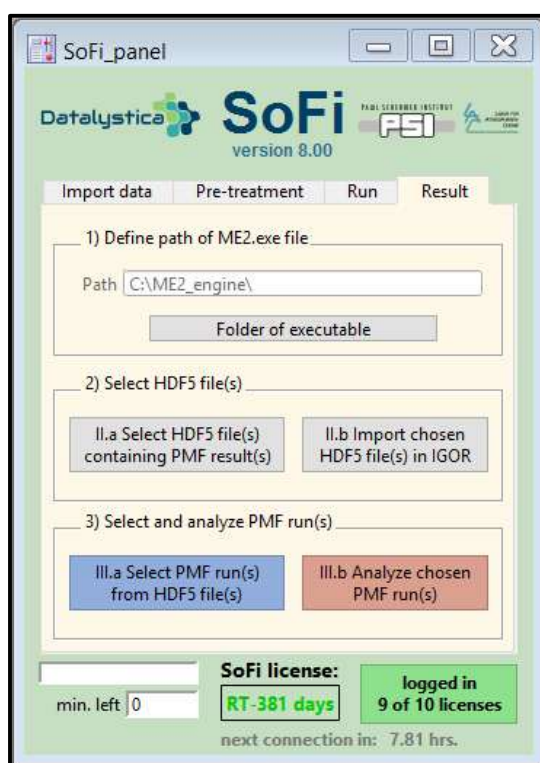


图 61

7.1.2. 选择包含 PMF runs 的 HDF5 文件。

当前的 SoFi 版本支持对单个 PMF 调用（一个 HDF5 文件）中的多个 PMF runs 的分析，也支持对多个 PMF 调用（多个 HDF5 文件）之间的分析。

按钮“II.a 选择包含 PMF 结果的 HDF5 文件 **II.a Select HDF5 file(s) containing the PMF result(s)**”会打开一个子面板，用户可以在其中决定将考虑哪些 HDF5 文件（图 62 此窗口必须通过再次按“II.a [...]”按钮关闭。）

可以通过按“SHIFT”键选择相邻的 PMF runs，或通过按“CTRL”键添加非相邻的 PMF runs 来选择多个 PMF runs。未选择的 PMF runs 将不会出现在 IGOR 中，或者如果之前存在，它们将在询问用户后通过提示消息被移除。这是为了保持 IGOR 实验的小型化，以避免可能的“内存不足”错误。从 IGOR 中移除 PMF runs 显然不会影响它们在 HDF 文件中的存储位置。

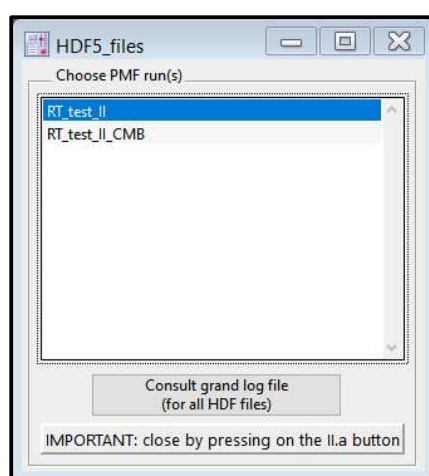


图 62

选项“咨询主日志文件（针对所有 HDF 文件） **Consult grand log file (for all HDF files)**”将弹出列出每个文件中的主要参数，例如，因子的数量、时间点的数量、约束类型等。这使用户能够保持对进行的 PMF runs 的概览，而无需强制导入每个文件进行视觉检查。

7.1.3. 将选定的 HDF5 文件导入 IGOR

按下“II.b 将选定的 HDF5 文件导入 IGOR **II.b Import the chosen HDF5 file(s) into IGOR**”按钮将导入 HDF5 文件的主要信息到相应的文件夹中。这样，仅导入了用于 7.2.2

7.2. 手动选择 HDF 文件中的 PMF runs

在按下“III.a 从 HDF5 文件中选择 PMF runs **III.a Select PMF run(s) from HDF5 file(s)**”后弹出的蓝色面板上（图 61 图 63

7.2.1. 设置

“Choice based on” 手动选择 Manual choice（默认情况），自动选择 automated selection（见 7.3

“Select HDF5 file” 导入的 HDF5 文件列表以选择 PMF runs

“avg. PMF run(s)” 是否应对 PMF runs 进行平均。有关更多信息，请参见 7.5

“x-axis type” 用户定义图表的 x 标签（见图 63C-value

7.2.2. 选择 PMF runs

作为一个小概述，用户有四种不同类型的图表来决定哪些 PMF runs 可能对进一步调查感兴趣：Q/Qexp、绝对值 absolute values、相对值 relative values、解释变异图 explained variation plot。Q/Qexp 图显示每个 run 的 Q/Qexp，而其他图则显示因子分布的更多细节：绝对值、相对值，并考虑（未）解释的变化程度。

要选择一个或多个 PMF run，用户最好在 Q/Qexp 标签页中进行选择，该标签页显示当前（选择的）HDF 文件中每个 PMF runs 的 Q/Qexp 概览图。用户通过在条形图周围拖动选框来选择 PMF runs，并通过右键菜单中的“**SoFi 选框 SoFi marquee**”>“**选择选项 Selection options**”>“**选择当前 SElect current**”来确认选择。“取消选择当前 UNSElect current”和“取消选择所有 UNSElect all”选项分别重置当前 HDF 文件或所有 HDF 文件的选项。表示需要进一步调查的 PMF runs 的灰色条的最顶端（仅 x 值）必须在选框内！

所选的 PMF runs 在 Q/Qexp 图中以蓝色突出显示。选框选择是累加的，可以在“**Select_PMFruns**”面板上以及弹出的图表中对任何图表进行操作。但只有在 Q/Qexp 图中，所选与未选的 runs 才能通过颜色区分。

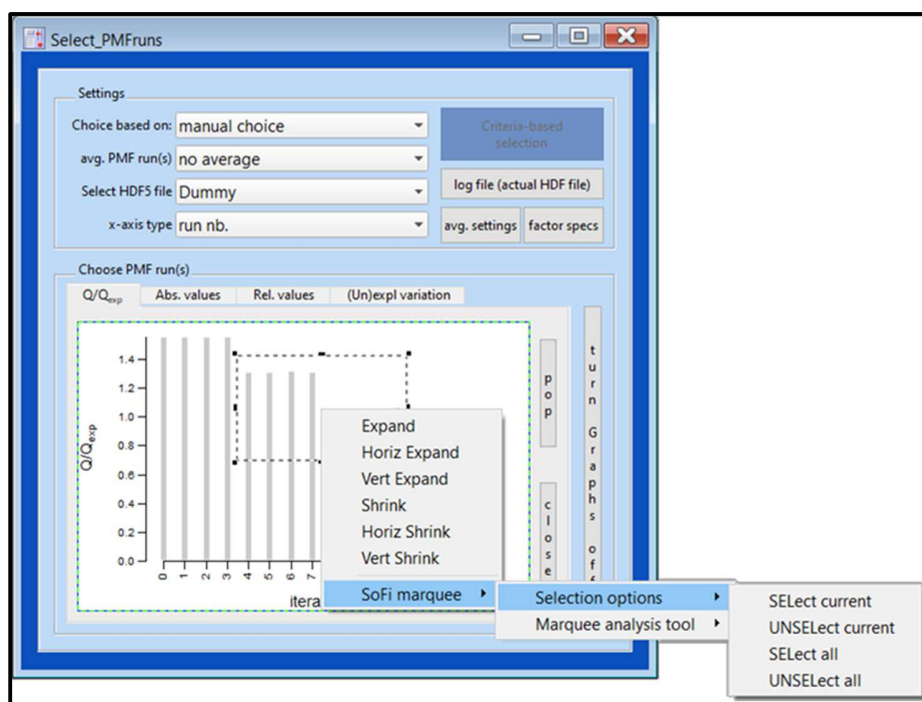


图 63: 从 HDF5 文件 “Dummy” 中使用选框选择一组解。

7.3. 基于特定条件的 PMF 筛选器

当 PMF 分析涉及数百甚至数千个单独的 run 时，例如在 bootstrap 法期间（见 6.46.4）或滚动窗口分析（见 6.5

用户可以定义这些特定条件，作为整体模型行为的代理，而不是检查每一个 PMF runs。这些条件可能包括，例如，相关性系数、所选变量之间的比率或某些参数的总和。同时还可以选择使用更客观选择适当 PMF runs 的高级统计方法。此外，SoFi Pro 还提供可以探索 1 维数组，例如特定因子源谱或日变化，以及以用户定义的 x、y 和颜色形式的 2 维数组，作为每个 PMF runs 的函数（详见下文）。

然后，每个 run 会被分配一组所选条件的分数，这些分数可以在 SoFi 中直接可视化和比较。要访问基于条件的选择，用户必须在下拉菜单“基于选择的选项 **Choice based on:**”中选择“自动选择 **automated selection**”选项（图 64

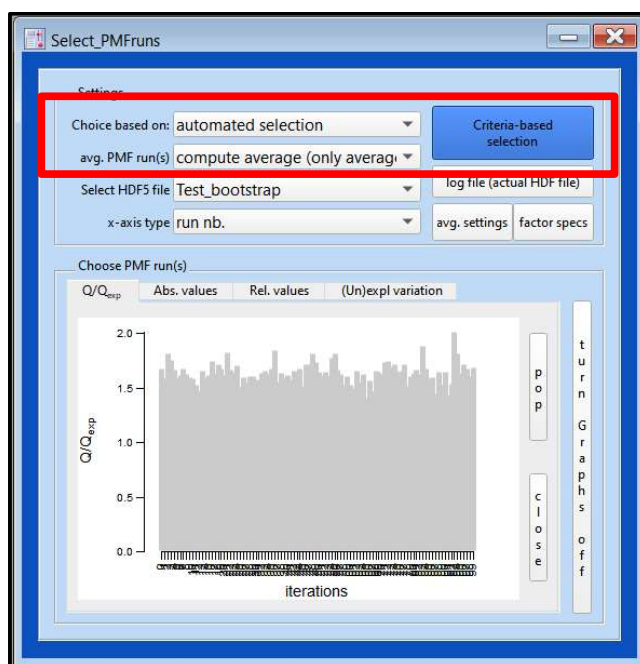


图 64

点击“基于条件的选择 **Criteria-based selection**”按钮会启动子面板“PMF_criteria”（见图 65

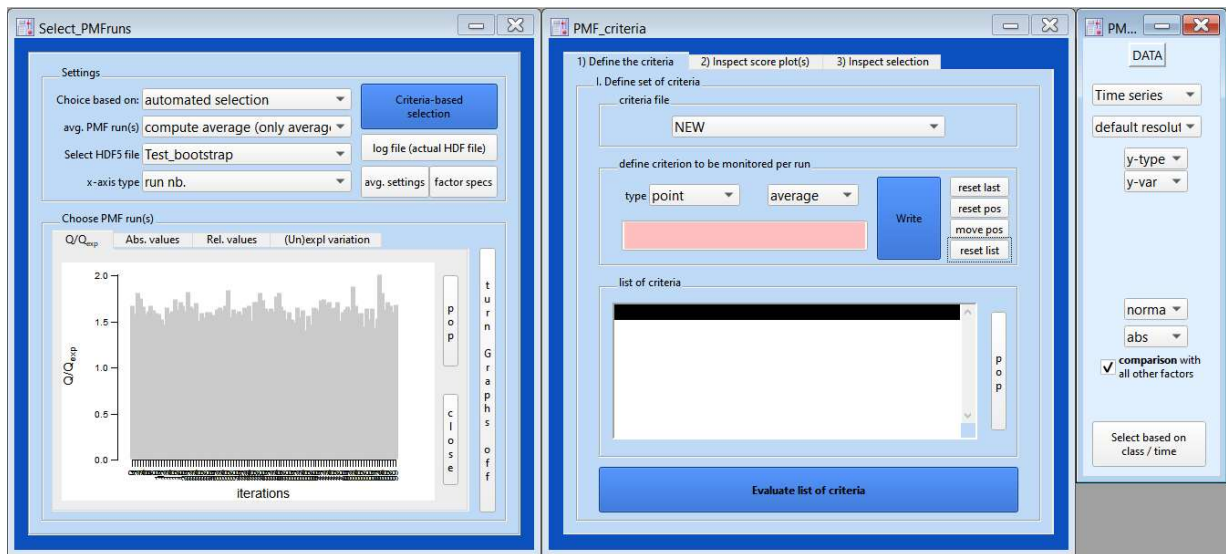


图 65

7.3.1. 条件类型

“PMF_criteria”中的第一个标签页管理用户定义条件的定义和概述（见图 66 点条件可用于因子排序。

“点 point”

每个 PMF runs 监测一个单点

-Average

-平均单一值（最终的 IGOR wave 可以是其他 wave 的和、积、比率等）被平均。

-Sum

-和单一值是最终 IGOR wave 的和。

-Multilinear regr.

对 PMF 结果进行多元线性回归。最多可以定义十个自变量

-R Pearson

R-Pearson 值对线性关系有用

-R Spearman

也称为秩相关，对任何类型的关系都有用。

-Kendall tau

分析一致对和不一致对的秩相关系数

--R uncentr

非中心相关系数，余弦相似度，用于比较因子源谱信息

“图像 image”

1 维数组在每次 PMF runs 中被监控。

“电影 movie”

2 维数组含有用户定义的颜色，即在每次 PMF runs 中监控的彩色散点图。

各种类型的条件被颜色编码，以便用户在概览列表“条件列表 **list of criteria**”中轻松区分它们（图 66

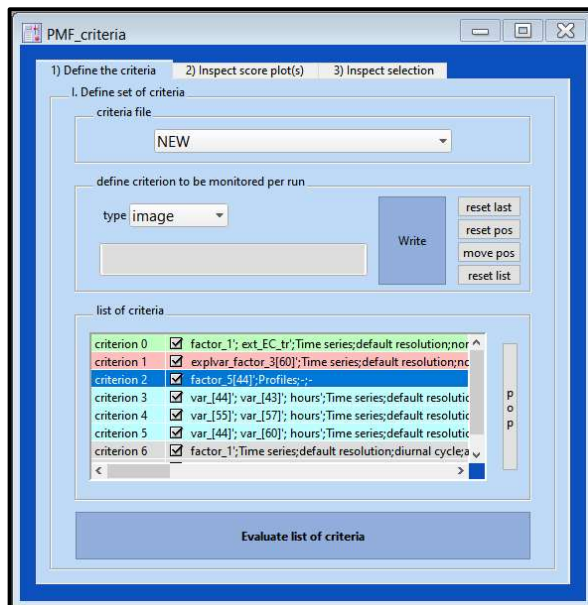


图 66: 不同类型条件的不同颜色有助于轻松区分它们。

7.3.2. 条件列表

在 SoFi 中可以轻松定义几种不同的条件，且容易忘记哪个条件是为哪个因子定义的。因此，一个小的概览表，包括条件类型的颜色编码，显示已定义的条件始终在第一个标签中可见。包含所有条件信息的更详细的批量表，如选择的确切时间点、数据类型（绝对值或占比）或时间分辨率，可以从第二个标签中查阅（见图 78

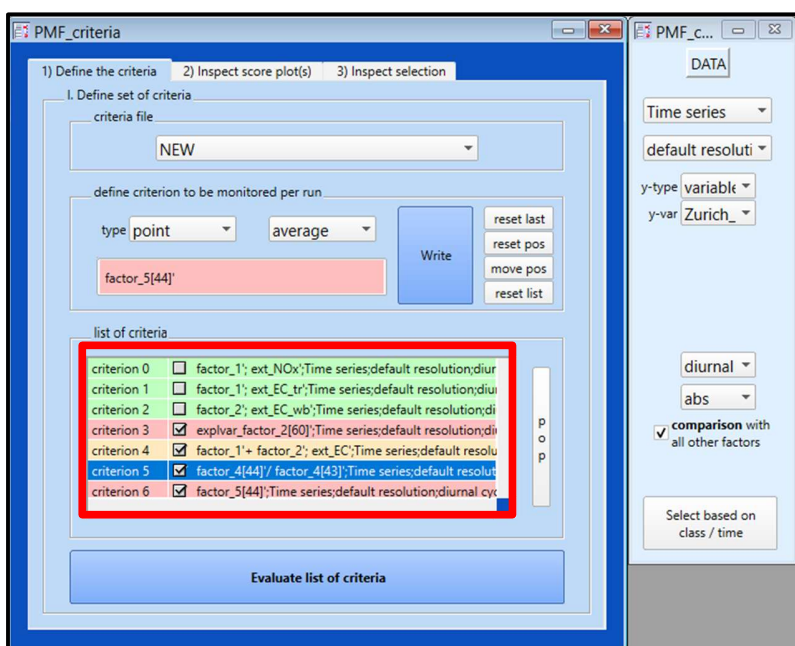


图 67

必须勾选评估的条件。如果某个条件在之前的轮次中已经被评估，则无需再次评估（除非它用于排序）。这节省了大量计算时间，使条件分析更快。必须选择用于重新定位（排序）不受限制因子的条件，并在条件列表中高亮（蓝色背景）显示（有关更多详细信息，请参见“关于不受限制因子的条件方法”和“主动与被动条件”下方（见图 67 此外，约束始终优先于排序条件，这意味着约束因子不能被排序，而是固定在它们被约束的位置。包含来自面板“基于类别/时间选择 **Select based on class / time**”的子选择的条件在条件列表中以加粗形式可视化。

7.3.3. 主动（排序）与被动（监控）条件

如果 runs 多个包含多个不受限制因子的 PMF，并且解没有手动检查，则会出现因子位置很可能会变化的问题。

表 6

PMFrun 1				
位置	1	2	3	4
因子	HOA	COA	BBOA	OOA
PMFrun 2				
位置	1	2	3	4
因子	HOA	BBOA	COA	OOA

在 PMF run 1 和 2 之间，位置 2 和 3 发生了交换。因子需要根据某些条件/条件重新定位，以确保位置 1 始终是 HOA，位置 2 是 COA，位置 3 是 BBOA，位置 4 是 OOA。

在可用位置上重新定位因子的问题是经典的“无替换抽样”问题，其规模与位置数量的阶乘成正比，例如，对于 6 个不受限制的位置，有 $6!$ 或 $6*5*4*3*2 = 720$ 种可能性。SoFi 计算所有组合并评估所有定义的条件。然后，分数在内部加权（z-score 后跟高斯排名转换 Gaussian rank transformation），以使它们具有相似的权重和范围。SoFi Pro 采用匈牙利算法 Hungarian algorithm 来优化排序因子。虽然此算法非常有效，但其计算成本随着不受限制因子数量的增加而迅速上升。例如，对于 10 个不受限制的因子，每次 runs 就有 3,628,800 种组合，因此约束已知且非常稳定的因子（例如固定的元素比例）会降低排序问题的维度，并允许 SoFi Pro 将计算工作集中在剩余的不受限制的因子上。SoFi 会返回得分总和最高的组合，用于最终排序因子。这确保了不受限制的因子以最佳方式进行得分排序。

用于重新定位无约束因子的条件称为主动条件 (**active criterion**)。对于每个因子，最多只能定义一个主动（排序）条件，除非它是一个受约束因子，此时约束将优先于排序。被动条件 (**passive criterion**)将仅为当前位置信息提供分数，而不会影响其排序。被动条件通常用于受约束因子，其中因子位置已经固定并提前已知，或作为正在排序的因子的进一步条件。用户可以通过从条件列表中直接选择主动条件来控制条件是主动/被动，方法是用鼠标点击“条件 x criterion x”。主动条件会被高亮显示（见图 67）

7.3.4. 条件语法

“数据 DATA”子面板（见）提供定义特定条件所需的所有选项。以下列表显示了一个时间条件（ts）。关于源谱的条件选项以（pr）为特征。每种数据类型在内部通过特征前缀或后缀进行区分（在下面的列表中以粗体括号显示）。这些标识符对于 SoFi 正确识别和评估相应条件至关重要。因此，用户务必不要修改这些前缀或后缀。当条件手动输入时，必须始终包含适当的前缀或后缀；否则，SoFi 将无法正确处理该条件。请注意，此子面板中的功能通常是可累加的。这意味着条件行中的现有信息不会被覆盖，而是将新信息附加到现有信息上。

“Time series, Profiles”	用于定义数据类型（贡献（ts）或源谱文件（pr））
“default, hourly, daily res.”	用于定义时间分辨率
“y (x)-type”	解 solution (ts, pr)，解的变量 (ts, ... [var.])，外部 external (ts, pr, ext_)，变量 variable (ts, var_)，残差 residual (ts, RES_)，绝对残差 absolute residual (ts, ABSRES_)，缩放（归一化）残差 scaled residual (ts, RESW_)，绝对缩放残差 absolute scaled residual (ts, ABSRESW_)，Q 残差 (ts, QRES_)，解释的变化程度 explained variation (ts, explvar_)，日变化 (ts, ... [hour])
“y (x)-var”	根据“y 类型”选择而变化
“normal, cycles”	用于定义是否应使用正常、每日、每周、每月或每年的周期
“abs, fraction”	用于定义是否应使用绝对数据或占比，即归一化到总和的数据
“comparison with all other”	可以计算来自显著性测试的额外 p 值和其他因子中最高判据得分

“select based on class/time” 根据类别或时间评估条件。注意，当选择了特定类别或时间时，条件的计算时间将增加。检查 5.3.15.3.2

条件输入行有以下几个语法规则：

- 1) 之前在“子面板数据 DATA”中描述的前缀和后缀必须严格遵循。
- 2) 每个包含波动规范的术语，例如因子 1 factor 1 或 RESW[total]，以及每个条件行的末尾必须以英文的单引号 ‘ 结束。
- 3) 每个术语必须以空格（第一个术语的情况）或某些数学符号开始，例如 “()”、“+”、“-”、“*”。

这是必需的，因为每个术语的评估具有以下结构：

“任何符号” - “术语” - “任何符号” - “引号”（终止）

如果不遵循这些规则，SoFi 在处理单个字符串时将失败，这会导致条件的评估不佳或不明确的 IGOR 错误消息。因此，在更改和定义新的手动条件之前，请确保您了解条件语法。

除了这个子面板，用户还可以直接更改、添加或删除条件行中的术语，并执行 IGOR 命令行通常支持的简单数学运算。

当前从条件行支持的数学运算

表 7: 7

数学运算	示例
加法、减法、乘法、除法	wave1' + wave2', wave1'-wave2', wave1'*wave2', wave1'/wave2'
指数运算	exp(wave1'), wave1'^2, 10^wave1'
平方根	sqrt(wave1)'
最小值，最大值	min(wave1', wave2')', max(wave1', wave2')'
直方图(x 最小值, x 最大值, 分箱数量, wave)	histogram(-5, 5, 1000, wave1)'
带括号的规则	(wave1'+wave2')*(wave3'+wave4')/wave5'
常量值	wave1'+wave2'+5', wave1'+wave2'+5'
一些组合示例	(wave1'+wave2')/2+sqrt(wave3'*wave4'^2)'

	<code>min(histogram(-5,5,100, wave1'), histogram(-5,-5,100, wave1'))'</code>
--	--

在这些示例中，wave1、wave2、wave3、wave4 和 wave5 可以是“子面板数据”下提供的任何选项。

对于表达式 `min(wave1, wav2)` 和 `max(wave1, wave2)`，比较是逐点进行的，针对这两个波。结果波的维度与 wave1 或 wave2 相同（它们必须具有相同的维度），并且对于每个点 *i*，执行 `min(wave1[i], wave2[i])`，以获得最小值。

`histogram(x min, x max, nb. of bins, wave1)` 为指定的 wave1 创建直方图，参数为 x min、x max 和 nb. of bins。选择“fraction”选项会将直方图的总和归一化为 1。请注意，直方图选项每个条件只能应用一次。此外，直方图特别适合与 `min()` 结合，对不同类别的残差逐个分箱进行比较。见几页后的示例。

组合条件

用户可以将两个或更多条件组合成一个新的条件。这在一个条件不足以捕捉所需表达时非常重要。例如，当需要组合各种时间或类别设置（子面板“基于类别/时间选择 **Select based on class / time**”）时，通常会出现这种情况。

组合条件的语法如下： `criterion X' + criterion Y'`

其中 X 和 Y 代表两个条件的位置。这个例子展示了求和，但任何数学组合以及普通条件和 GRAND 条件之间的组合在理论上都是可能的。

7.3.5. 如何设置条件列表的示例

以下是一些说明如何设置条件的示例。

7.3.5.1. 因子 1（假设为交通源贡献）与 NOx 之间的相关性（见图 68）

- 1) 在“类型 type”中选择“点 point”和“R-Pearson”或“R-Pearson^2”。请注意，框变为绿色，因为这是 Pearson 选项的颜色编码。
- 2) 在子面板“数据 DATA”中选择“时间序列 Time series”选项，“y 类型 y-type”的“结果 solution”，然后选择第一个条目，即 factor_1。此外，选择“外部 external”作为“x 类型 x-type”和 NOx 数据。
- 3) 在这个阶段，条件线显示 factor_1'; ext_NOx'。
- 4) 点击“写入 Write”将写入此条件，并传递所有数据（未选择时间子集）用于评估的信息。
- 5) 成功编写的条件在“条件列表 list of criteria”下报告。

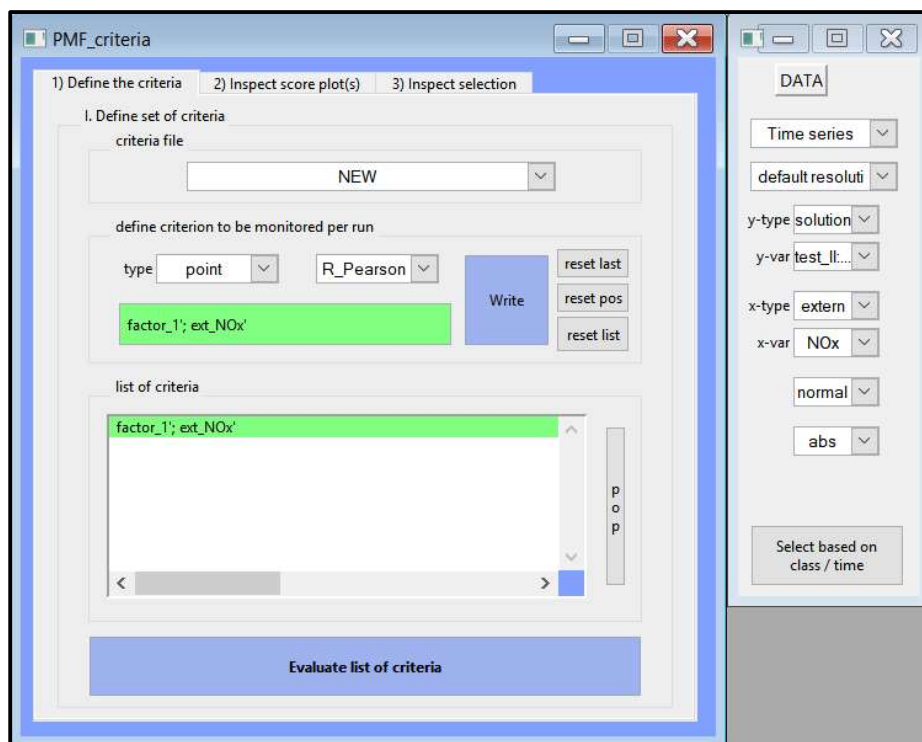


图 68

7.3.5.2. 因子 2 在某些时段的比例（见图 69）

- 1) 在“类型 type”下选择“点 point”和“平均 average”。请注意，框的颜色变为红色，这是平均选项的颜色编码。
- 2) 在“数据 DATA”子面板中选择“时间序列 Time series”选项，选择“小时 hours”作为“y 类型 y-type”，然后选择“factor_2[hr11]。”
- 3) 在此阶段，条件线显示为 factor_2[hr12]’。

用户可以直接添加其余信息，即 $((\text{factor_1}[\text{hr11}] + \text{factor_1}[\text{hr12}]) / 2) / ((\text{factor_1}[\text{hr6}] + \text{factor_1}[\text{hr7}] + \text{factor_1}[\text{hr8}]) / 3)$ ’

该表达式将评估中午（上午 11 点和 12 点）峰值与上午 6 点、7 点和 8 点的平均值之间的比例。

- 4) 点击“写入 write”将写入此条件，并传递所有数据（未选择时间子集）用于评估的信息。
- 5) 成功编写的条件在“条件列表 list of criteria”下报告。

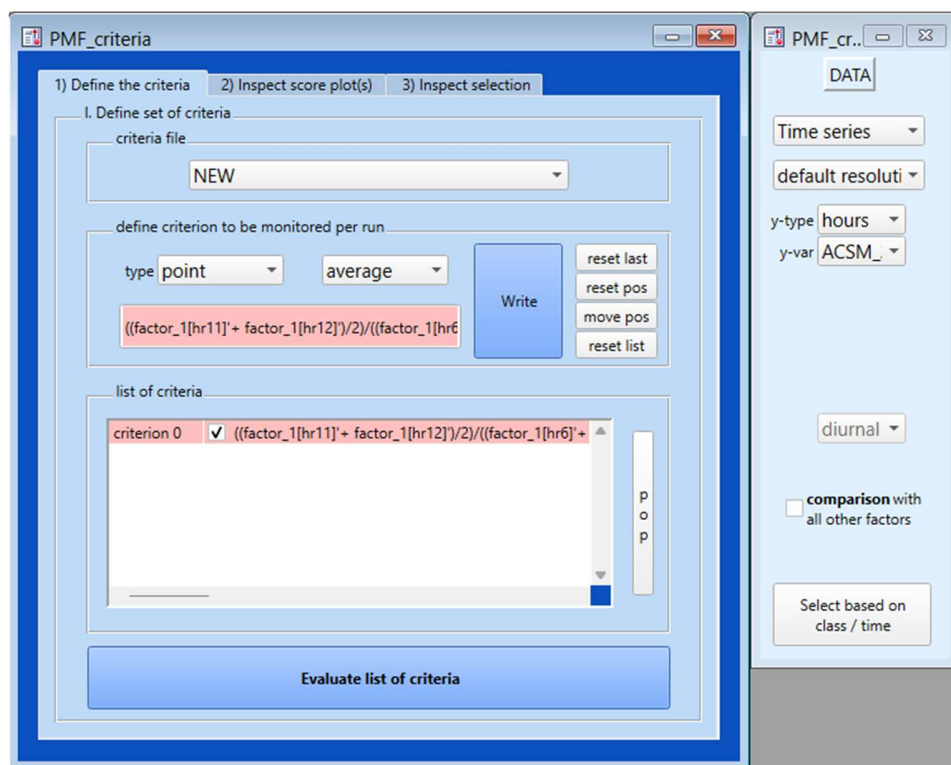


图 69

7.3.5.3. 监测第三因子的 m/z 60 的质量分数 (f60) (见图 70)

- 1) 在“类型 type”下选择“点 point”和“平均 average”。请注意，框的颜色变为红色，这是平均选项的颜色编码。
- 2) 在“数据 DATA”子面板中选择“时间序列 Time series”选项，“y 型 y-type”的“变量解 Solution of var.”，然后选择“factor_3 60:60”。此外，还要选择“分数 fraction”。
- 3) 在此阶段，条件线显示为 factor_3[60]’。
该表达式将评估第三因子的 m/z 60 的质量分数。
- 4) 点击“写入 write”将写入此条件，并传递所有数据（未选择时间子集）用于评估的信息。
- 5) 成功编写的条件在“条件列表 list of criteria”下报告。

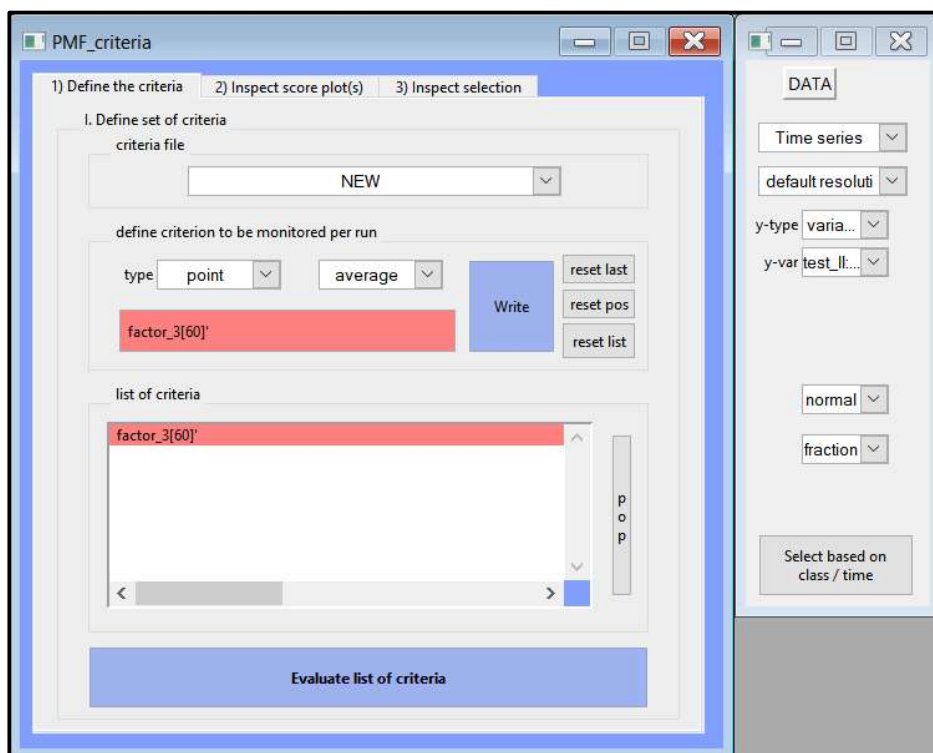


图 70

7.3.5.4. 监测多元线性回归中的相关值 R（见图 71）

线性回归是在因子 1（假设交通）和因子 3（假设生物质燃烧）与黑碳 EC 的时间序列之间进行的，假设只有交通和生物质燃烧对黑碳有贡献。

- 1) 在“类型 type”中选择“点 point”和“多元线性回归 Multilinear regression”。请注意，框的颜色变为橙色，这是多元线性回归选项的颜色编码。
- 2) 在“数据 DATA”子面板中选择“时间序列”选项，选择“y 类型 y-type”的“结果 solution”，然后选择“factor_1”。此外，选择“外部 external”作为“x 类型 x-type”，然后选择“EC”。
- 3) 在此阶段，条件线显示为 `factor_1' ; ext_EC'`。用户可以直接将剩余信息，即 `+ factor_3'` 添加到字符串的第一部分。最终表达式为：`factor_1' + factor_3' ; ext_EC'`。该表达式将执行多元线性回归，并找到斜率 a 和 b 的最佳值： $EC = a \cdot factor_1 + b \cdot factor_3$ 。EC 是因变量，`factor_1` 和 `factor_3` 是两个自变量。当前选项允许用户添加最多 5 个自变量，并始终只有一个因变量。对于排序而言，该条件的实用性有限。对于排序，所有因子组合都应进行测试。然而，SoFi 只能迭代组合中最后提到的因子，这意味着在当前条件下，`factor_1` 保持固定，而评估其他因子的组合。因此，我们建议使用其他条件进行排序。
- 4) 点击“写入 write”将写入此条件，并传递所有数据（未选择时间子集）用于评估的信息。
- 5) 成功编写的条件在“条件列表 list of criteria”下报告。

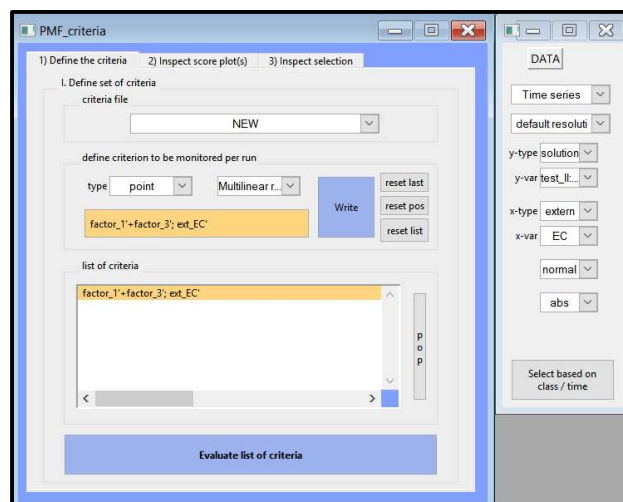


图 71

7.3.5.5. 监测第二个因子的 m/z 57 与 m/z 55 的比率（见图 72）

- 1) 在“类型 type”下选择“点 point”和“平均 average”。请注意，框的颜色变为红色，这是平均选项的颜色编码。
- 2) 在“DATA”子面板中选择“Profiles”选项，“solution”作为“y-type”，然后选择“factor_2 57:57”
- 3) 在此阶段，条件行显示为 **factor_2[57]’**。

用户可以直接添加其余信息，即 **/factor_2[55]’**。

该表达式将评估第二个因子的 m/z 57 与 m/z 55 之间的比率。

- 4) 点击“写入 write”将写入此条件，并传递所有数据（未选择时间子集）用于评估的信息。
- 5) 成功编写的条件在“条件列表 list of criteria”下报告。

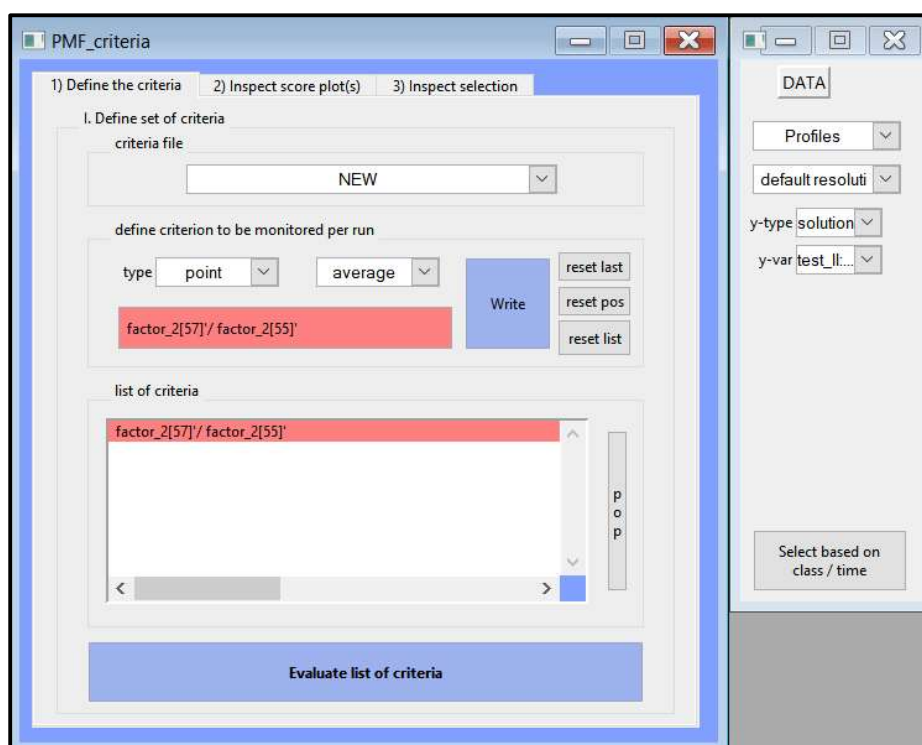


图 72 m/z 57 与 m/z 55 的比率。

7.3.5.6. 仅针对工作日的因子 1（假设交通）与 NOx 之间的日变化相关性（见图 73

- 1) 在“类型 type”中选择“点 point”和“R-Pearson”或“R-Pearson²”。请注意，框变为绿色，因为这是 Pearson 选项的颜色编码。
- 2) 在子面板“数据 DATA”中选择“时间序列 Time series”选项，“y 类型 y-type”的“结果 solution”，然后选择第一个条目，即 factor_1。此外，选择“外部 external”作为“x 类型 x-type”和 NOx 数据。
- 3) 在这个阶段，条件线显示 factor_1'; ext_NOx'。
- 4) 从“基于类别/时间选择 Select based on class / time”面板中选择日变化 diurnal cycle 和工作日 weekdays
- 5) 点击“写入 write”将写入此条件并传递用于评估的选定信息。
- 6) 成功编写的条件在“条件列表 list of criteria”中报告。请注意，该条件是加粗的，因为应用了来自“基于类别/时间选择 Select based on class / time”面板的时间子选择。

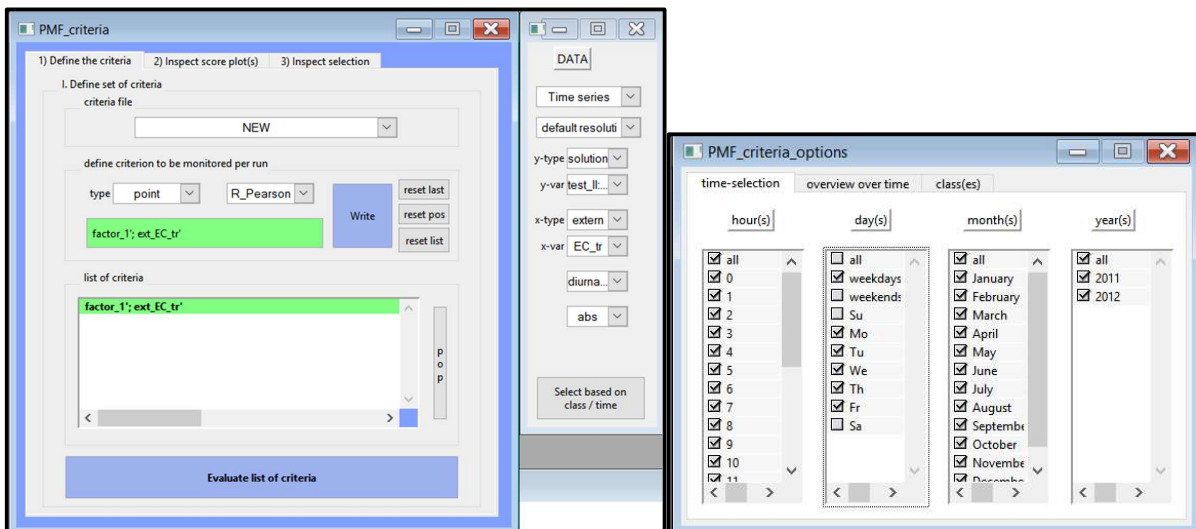


图 73

7.3.5.7. 监测第四因子的 m/z 44 的质量分数与 m/z 44 的缩放残差之间的比率（见图 74

- 1) 在“类型 type”下选择“点 point”和“平均 average”。请注意，框的颜色变为红色，这是平均选项的颜色编码。
- 2) 在“数据 DATA”子面板中选择“时间序列 Time series”选项，“y 类型 y-type”的“结果 solution of var.”，然后选择“factor_4 44:44”。
- 3) 在此阶段，条件行显示为‘factor_4[44]’。用户可以直接添加其余信息，即/ RESW_[44]’。

该表达式将评估第四因子的 m/z 44 的质量分数与 m/z 44 的缩放残差之间的比率，该因子预计对 m/z 44 有很高的贡献。

- 4) 点击“写入 write”将写入此条件，并传递所有数据（未选择时间子集）用于评估的信息。
- 5) 成功编写的条件在“条件列表 list of criteria”下报告。

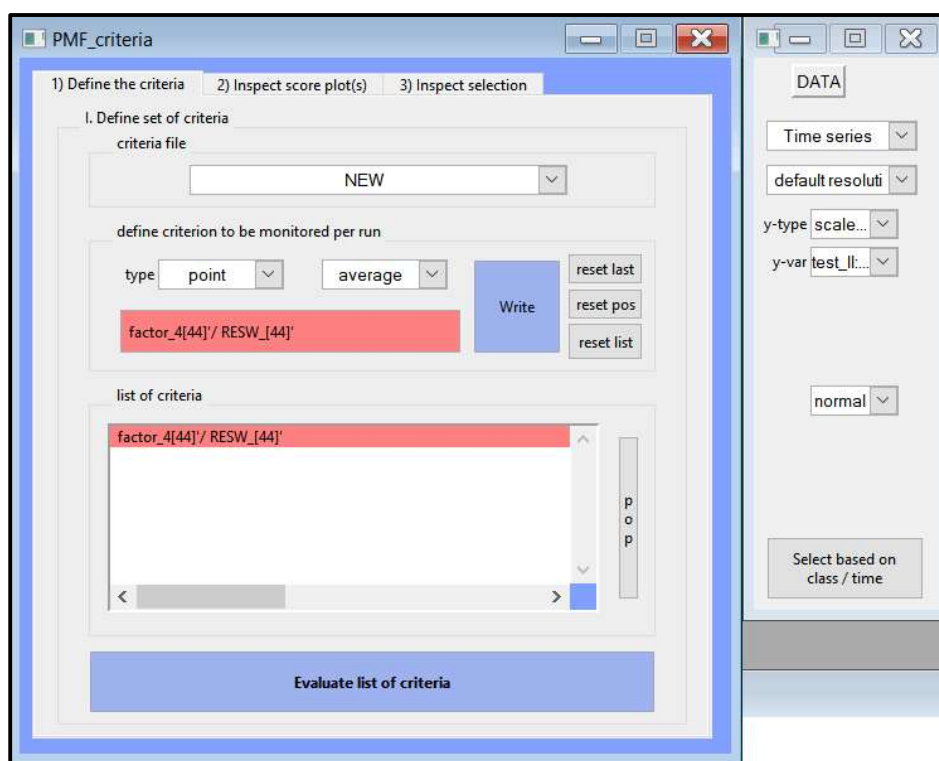


图 74

7.3.5.8. 使用组合条件监测 C 值 runs

当来自两个不同仪器的数据（例如 AMS 和 PTR-MS）被合并时，使用 C 值方法进行相对重标定（见 6.2.4.4

- 1) 在“类型 type”中选择“点 point”和“平均 average”。在子面板“数据 DATA”中选择“时间序列 Time series”选项，“缩放残差 scaled res”作为“y 类型 y-type”，然后选择“total_matrix”。此时，条件为 RESW[total_matrix]’。
- 2) 将直方图选项（最小值、最大值和分箱数）添加到条件行中，使其读取为 e.g. **histogram(-5,5,1000,RESW[total_matrix])**’选择“分数 fraction”，并从“基于类别/时间选择 select based on class/time”选项中选择类别 1（例如 AMS）。然后将条件写入列表中。
- 3) 重复，但这次从“基于类别/时间选择 select based on class/time”选项中选择类别 2（例如 PTR-MS）。
- 4) 现在创建条件，使 SoFi 实际比较直方图。在“类型”中选择“点 point”和“总和 sum”。确保两个类别都被选中。
- 5) 手动写入 min(criterion 0’,criterion 1’)’ 并将条件添加到列表中。

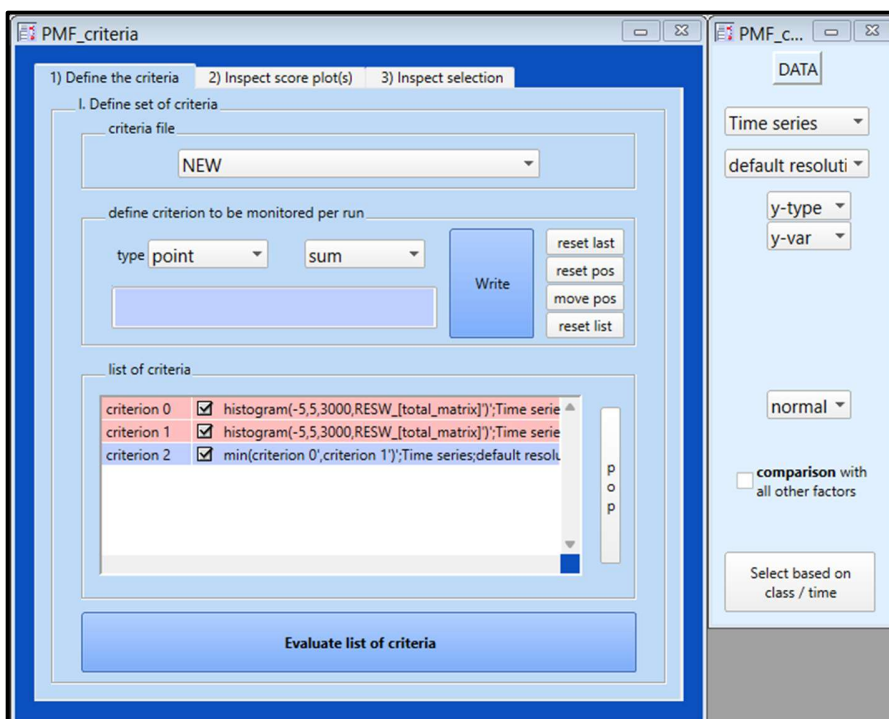


图 75

7.3.6. 评估条件

按下“评估条件列表 **Evaluate list of criteria**”按钮可以让 SoFi 遍历当前 HDF 文件所属的所有 PMF runs，并评估用户定义的条件列表（见图 76）

此外，“条件文件 **criteria file**”下的下拉菜单控制条件列表的保存/加载选项。

“NEW”

创建一个新的条件列表。如果当前列表不为空，它将被删除。

“LOAD list of criteria”

可以导入存储为 IGOR txt 文件的条件列表

“LOAD list of criteria and evaluation”

导入同时包含判据列表和判据得分的 IGOR txt 文件。只有当待加载文件属于同一个 HDF 文件时，该选项才有意义。

“SAVE list of criteria”

条件列表将保存在一个 IGOR txt 文件中

“SAVE list of criteria and evaluation”

包含其分数的条件列表将作为 IGOR txt 文件保存

如果加载了条件列表，它应该来自同一个实验，因为索引也会被加载。如果这些不匹配，评估的分数将是 NaN。

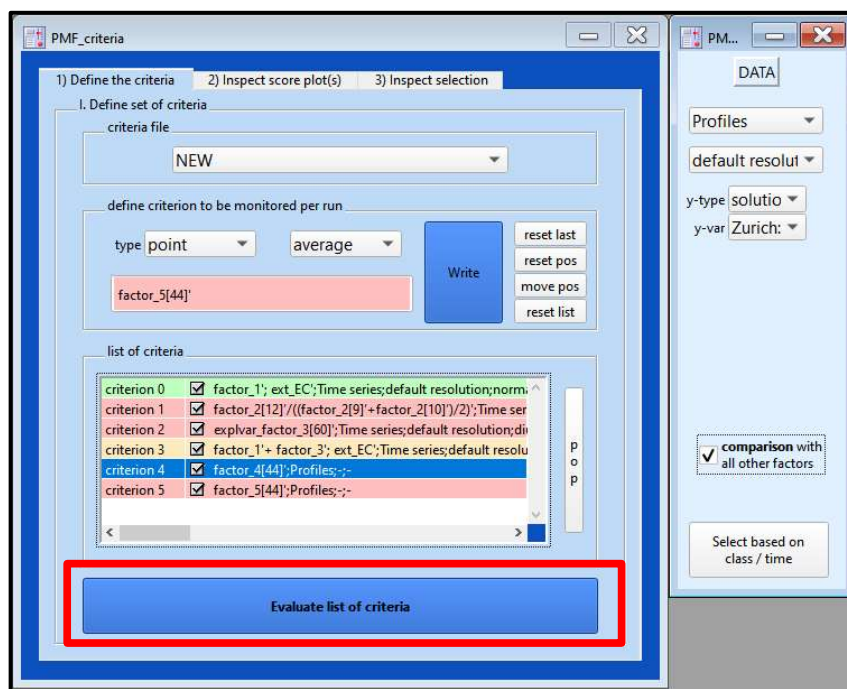


图 76

7.3.7. 检查条件分数

SoFi 在 “PMF_criteria” 面板的第二个标签 “2) 选择 PMF runs 2) Select PMF run(s)” 中报告每个条件和 PMF runs 的分数（见图 78

按钮 “弹出批处理表 **pop batch table**” 将打开一个表格，其中包含在上一个标签中为所选条件选择的所有信息（见图 77

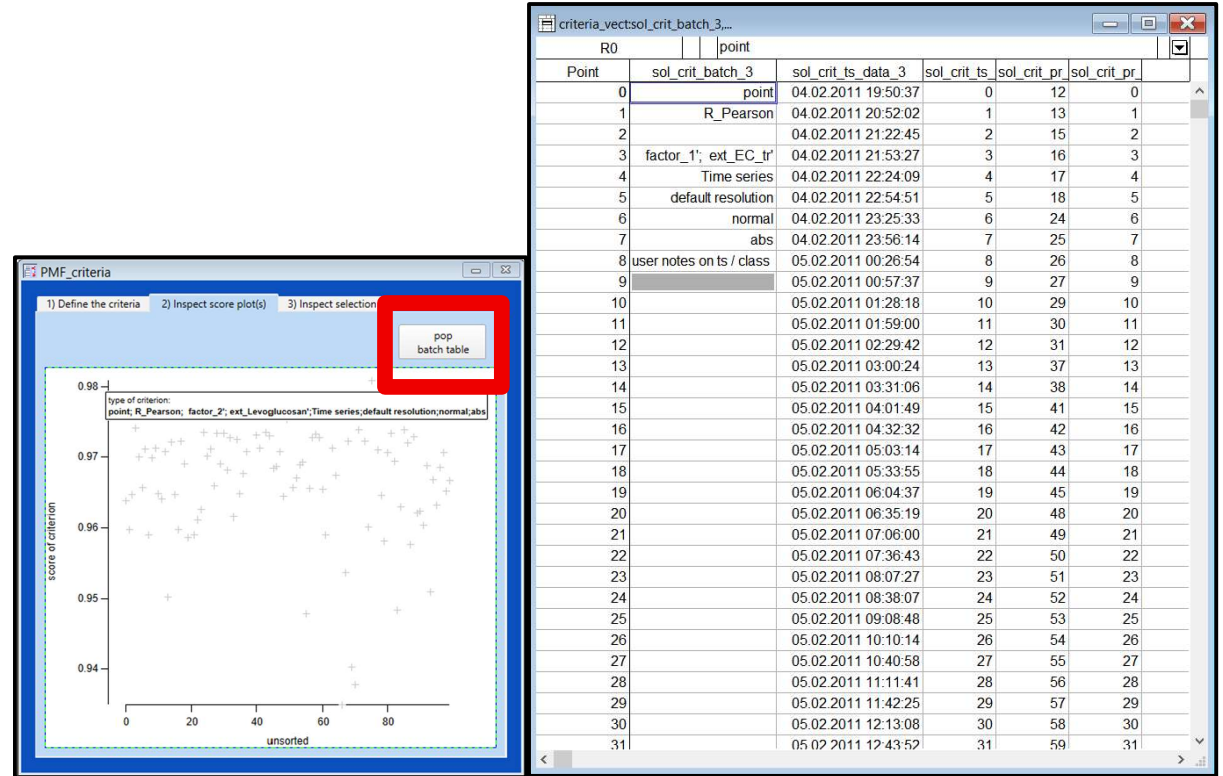


图 77

子面板 “数据 DATA”

- “条件 **criteria**”

“因子 **factors**”

“排序 **sorting**”

“其他最高 **other highest**”
- 选择应调查的条件

如果 PMF 调用包含不同因子数量的 runs，用于选择因子数条件的分数可以是 “未排序的 **unsorted**”，按 “升序 **ascending order**” 或 “降序 **descending order**” 排序（详见下文）

如果当前因子的得分已经是最高值，则显示第二高得分；如果当前因子的得分不是最高值，则显示全部其他因子中的最高得分。（详见下文）

“p 值 p-values”

基于 T 检验 t-test 绘制 p 值，并且如果可用，还绘制相关系数的 p 值。此信息可以直接用于排除不具有统计显著性的相关系数。

“... MARQUEE 选择...”

在图中使用精准的限制，而不是框选

“Find PMF runs...”

SoFi 查找满足用户在所有条件上选择的 PMF runs。此选项在选择排序分数时至关重要。

“nb. of common runs:”

这里显示共同满足筛选条件的 PMF runs 数量

“Apply selection”

共同的 runs 直接传递给框选（见图 63）

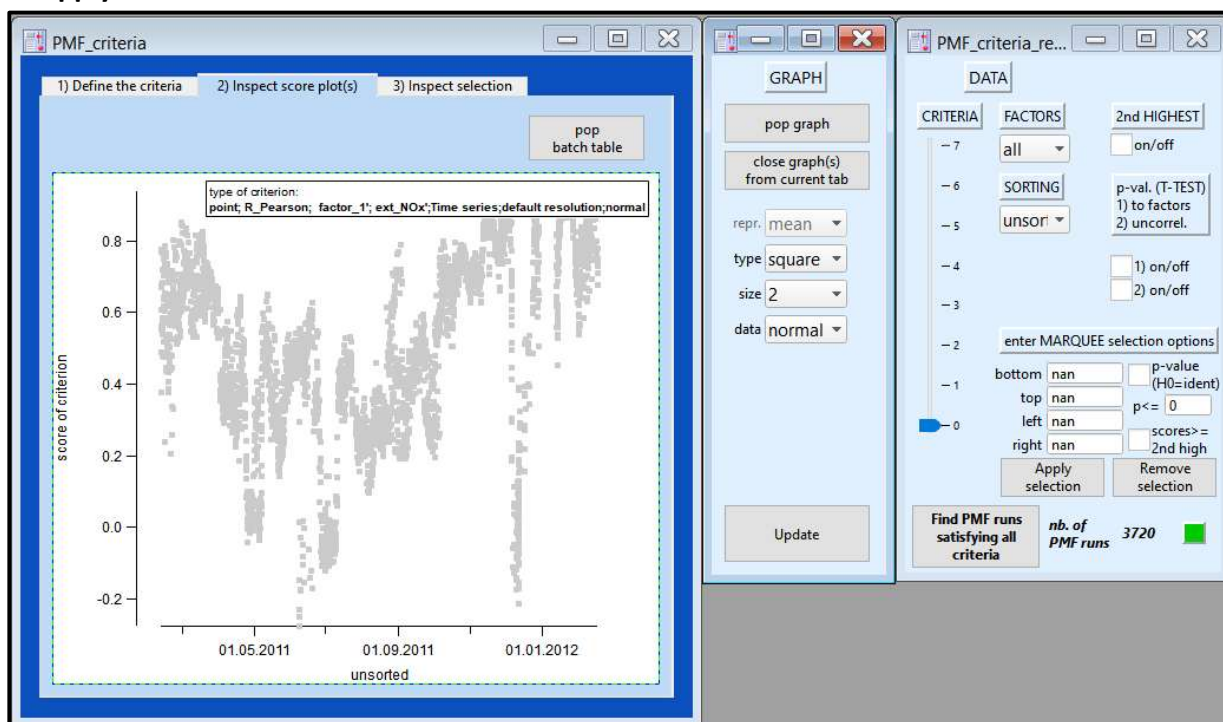


图 78 因子 1 与外部数据 NO_x 时间序列的 Pearson 相关性随时间的变化（灰线）。相邻的子面板控制与 PMF run 选择相关的所有选项。

用户的任务是使用框选工具选择应作为完整 PMF 结果进行调查的得分（PMF runs）。为了更好地可视化得分的范围和密度，分数图可以按升序或降序排序。例如，当预期高分时，用户可以更容易地在排序图中选择适当的分数。

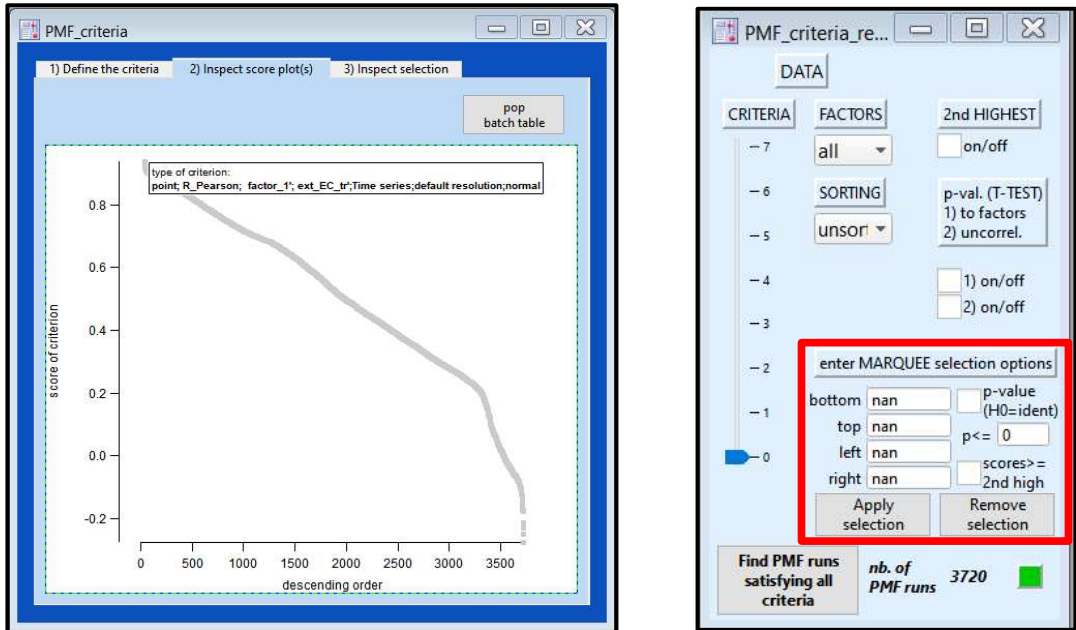


图 79: 条件 f 43 用于因子 5 的源谱（灰线），包括得分图（红曲线）的斜率。绘制选框并选择感兴趣区域周围的点会将它们高亮显示为蓝色。

筛选过程中最关键的问题是：应在何处停止选择。这可以基于已知的绝对值或从其他统计分析中得出的阈值（见条件的稳定性）7.3.99.2

另一种可能性是仅选择与所有其他因子相比可以实现最高分的解。为此，可以使用“其他最高 **other highest**”选项。如果选择了此选项，SoFi 将绘制下一个较低的（如果此 run 的分数已经是最高）或最高的（如果此 run 的分数不是最高）。这也可以用于自动选择（分数 $\geq$ 其他高）**scores $\geq$ other high**。

此外，排序后的得分图形状也可以进行研究。作为第一个指标，得分图的强烈下降表明因子之间的混合程度很高（见图 79 图 792019 年 4 月 21 日和 2019 年 6 月 21 日分别对应左侧和右侧。此外，用户还可以根据显著性测试的 p 值以及第二分类得分值排除不满足条件的数据点（见 7.3.9

关于各种条件的得分图的一些印象在接下来的几个图中展示。

a) 点图 Point representation	b) 点图 Point representation
----------------------------	----------------------------

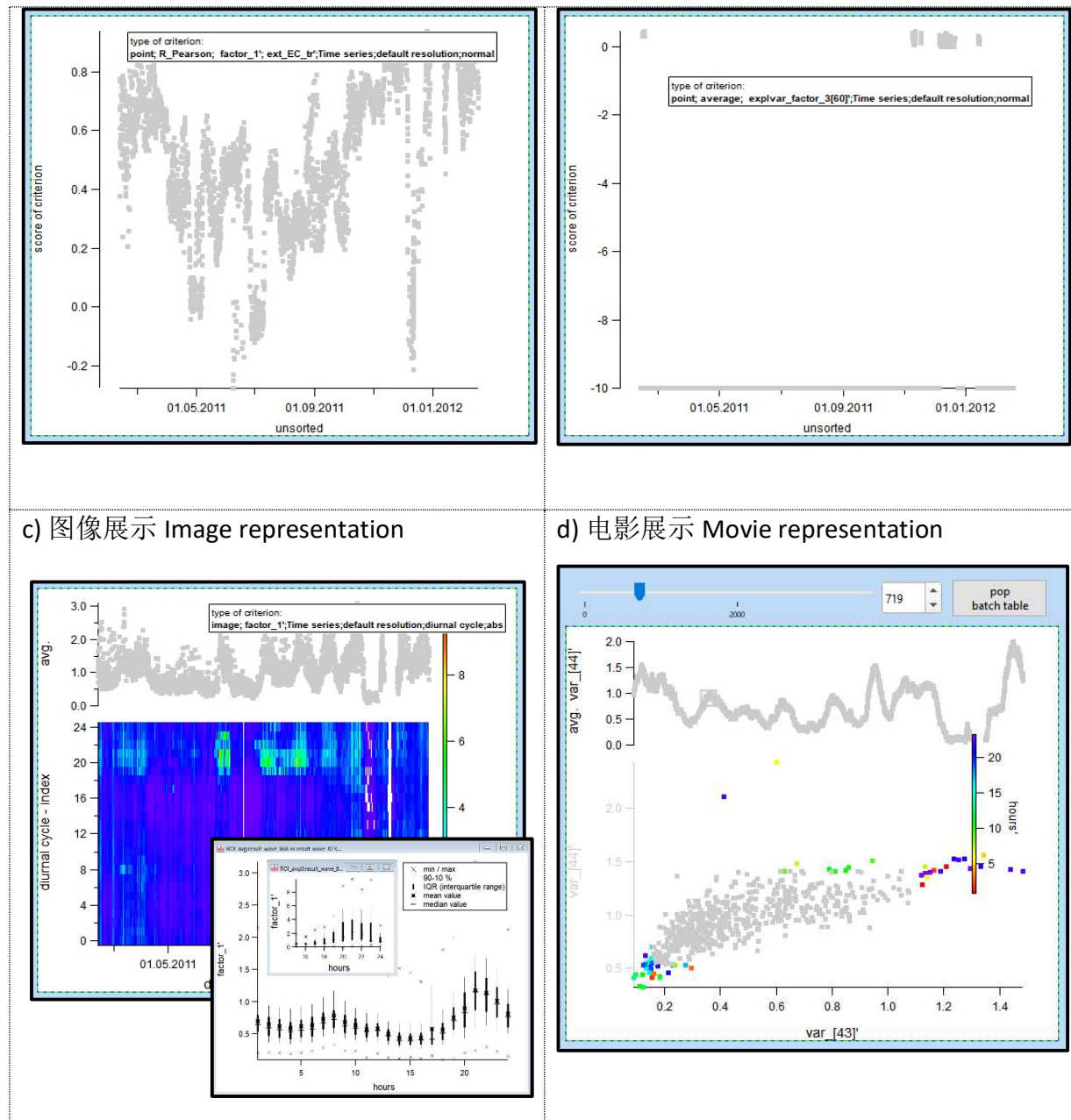


图 80

一旦为所有条件选择了得分图中的区域，满足所有选择的 PMF runs 将被识别并突出显示，以便进一步分析“查找满足所有条件选择的 PMF runs **Find PMF runs fulfilling selection for all criteria**”和“应用选择 **Apply selection**”（见图 816.5 图 82

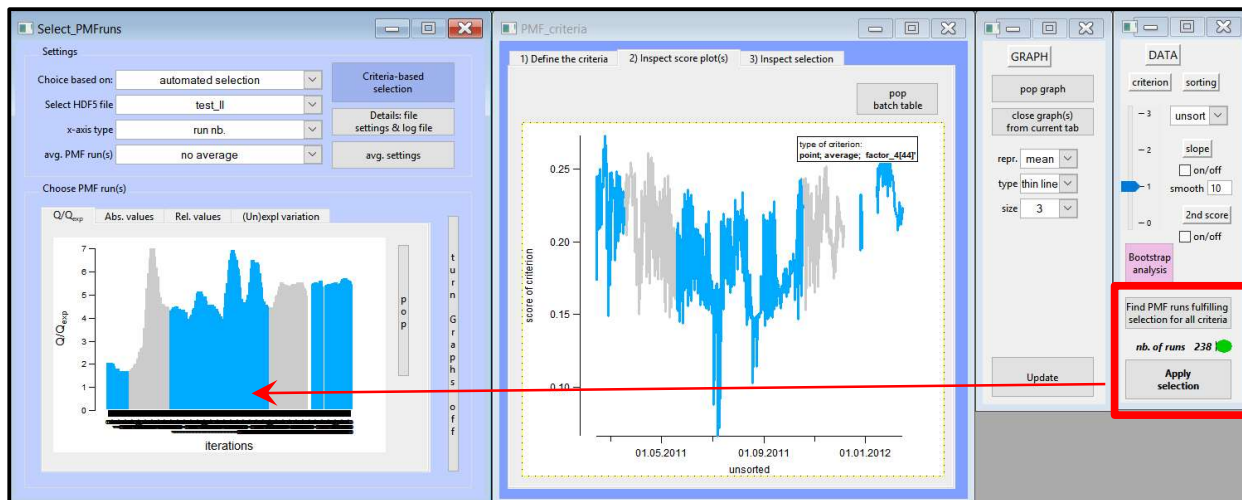


图 81

7.3.8. 检查选定的 PMF runs

选择（见图 81 图 82 图 79

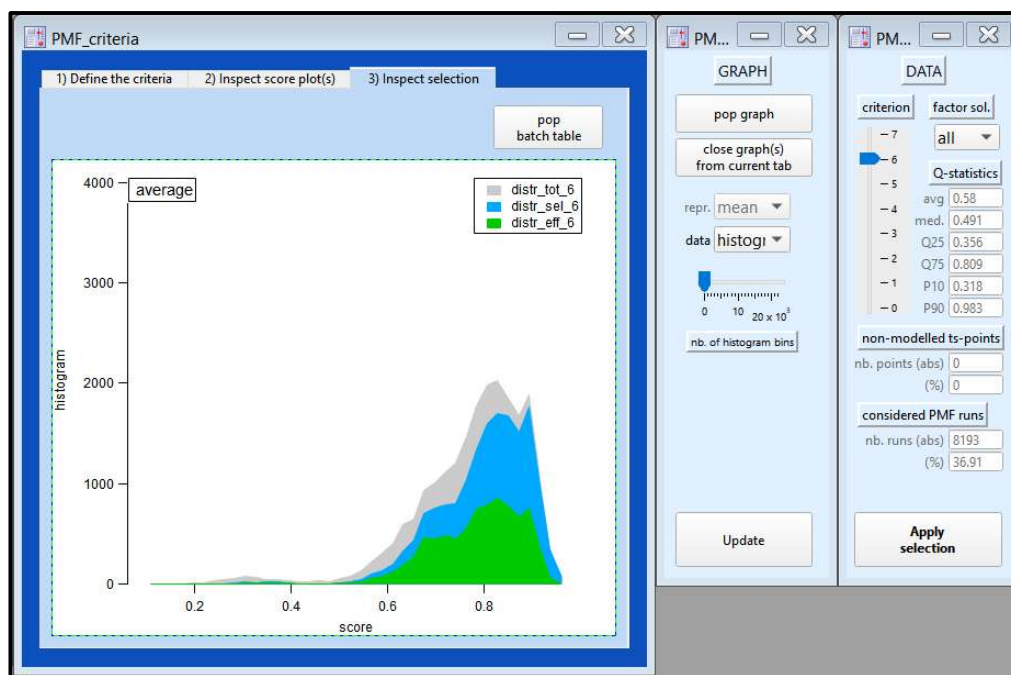


图 82

相邻子面板上的可用选项

“data”

选择正常 normal（得分图）、直方图（正常） histogram (normal)和直方图（pdf）之间的选项

- “**sorting**” 在正常 normal 情况下：条件的得分可以是“未排序 unsorted”、“升序 ascending”或“降序 descending”
- “**nb. of histogram bins**” 在直方图情况下：滑块用于设置直方图的箱数
- “**criterion**” 定义应调查的条件
- “**Q-statistics**” 在快速扫描 PMF runs 后对有效 PMF runs 的 Q 统计
- “**non-modelled ts-points**” 缺失时间点的统计数据
- “**considered PMF runs**” 对最终选择的 PMF runs 的数量和占总 PMF runs 数量的统计

7.3.9. 检查条件的稳定性

可以通过比较另一个因子的最高得分来判断 SoFi 中得分选择的稳定性，无论该因子是否受到约束。如果另一个因子的得分高于当前因子的得分（该得分本应是最高的），则该解可以被视为混合并被丢弃。此外，还可以检查得分值及其显著性检验的 p 值（仅当选“与所有其他因子比较 **comparison with all other factors**”复选框时，见图 66 图 83

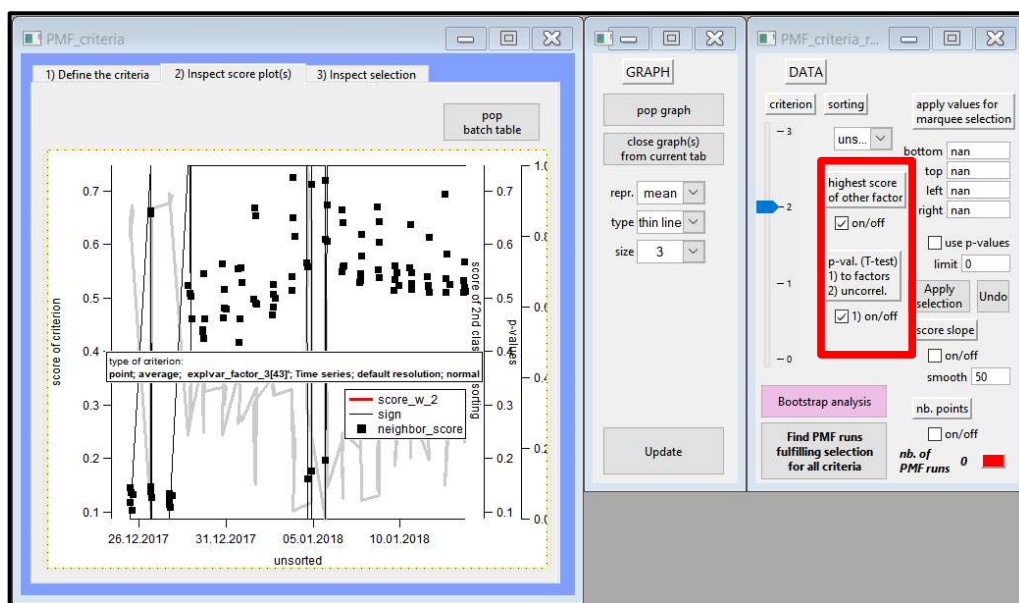


图 83

7.3.10. 一个 HDF 文件中的不同因子数量

当使用判据条件面板检查包含不同因子数的模型解时，该功能尤为重要。例如在一个 HDF 文件中进行的 PMF runs，例如五个和六个因子。当为第六个因子定义条件时，该条件的分数图在一半的时间内将是空的，即对于所有五因子的 runs。这将导致所有五因子的 PMF runs 在最终解中缺失，因为它们具有未定义的值（NaN），并且在寻找满足所有条件选择

的 PMF runs 时，它们将被全部丢弃。因此，SoFi 将 0 传递给这些未定义的值，允许用户选择这些 runs（零值在图 84:

84

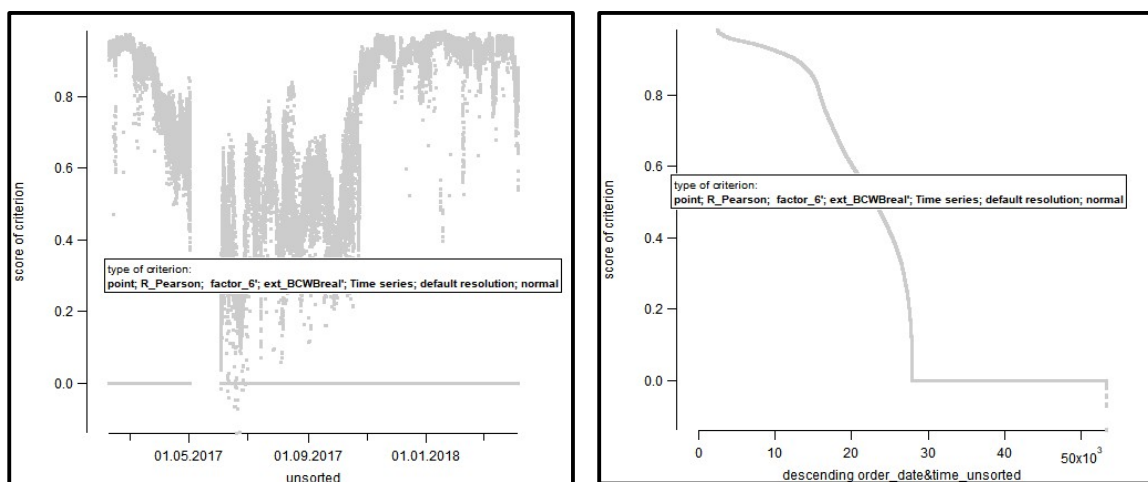


图 84: 84

此外，从多因子数的解中检查特定因子 runs 时，例如仅在所有 PMF runs 中提取 5 因子解的因子并用其在整个得分图上做特定因子与其 tracer 之间的相关系数是困难的。用户可以指定应提取哪个因子解（图 85 图 85

SoFi 防止对具有不同因子数量的 PMF runs 进行平均。当用户通过选框 marquee 或点击“应用选择 **Apply selection**”按钮时（图 79）一些 PMF runs。SoFi Pro 也会自动关闭不同因子数量的选项当在同一个 rolling 的时间窗口运行 rolling PMF 时，（6.5）会自动禁用。（如果在具有多因子解的 HDF 文件中选择的 PMF runs 导致重叠期，SoFi 将仅考虑出现频率最高的因子解。

当对来自不同 HDF 文件的 PMF runs 进行平均时，具有相同自定义名称的因子（见 6.3.1

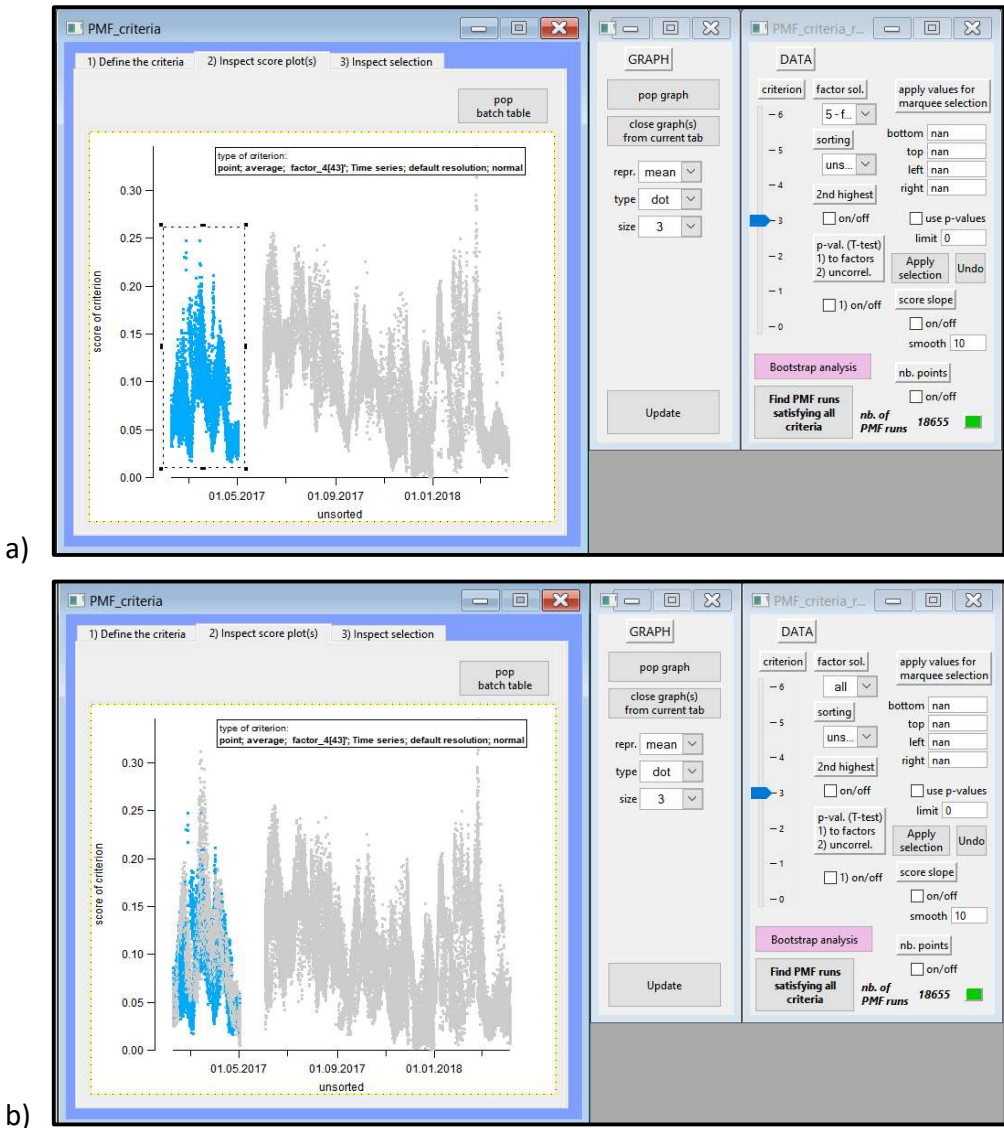


图 85: a) 仅针对 5 因子 PMF runs 的得分图。用户已选择了所有在选框内的 PMF runs (蓝点)。b) 所有因子解的得分图。之前选择的 5 因子 PMF runs 为蓝点，而在 a) 中标记的相同时间段内，灰点为未选择的 6 因子 PMF runs。所有其他在 a) 中选框外的灰点分别为 5 因子和 6 因子 PMF runs。

7.4. 文件设置和日志文件

用户可以通过“日志文件（当前 HDF 文件）log file (actual HDF file)”按钮弹出 PMF 调用的主要设置（图 86 图 87

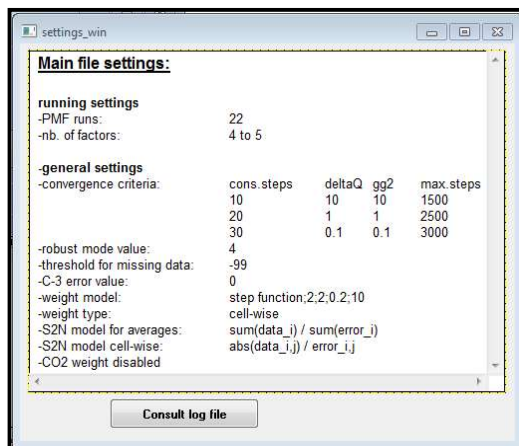


图 86

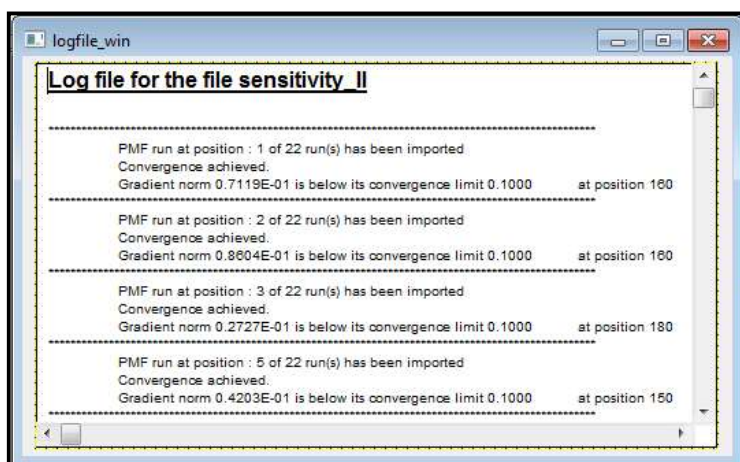


图 87

7.5. 关于如何导入所选 PMF runs 的选项

通过“II.b 在 IGOR 中导入所选 HDF5 文件 II.b Import chosen HDF5 file(s) in IGOR”，仅导入了 PMF 结果的子集，足以创建概览图。为了进一步检查，从所选 PMF runs 中导入所有信息 7.2.7.3 图 61

7.5.1. 平均

SoFi Pro 允许用户对多个 PMF runs 进行平均。这是通过下拉菜单“avg. PMF run(s)”控制的。

可用选项:

“no average”	仅导入和评估单个 PMF runs
“compute avg. (save single runs)”	计算这些 runs 的平均值并保存单个 run。此选项仅适用于少量 PMF runs
“compute avg. (only average)”	计算并存储 PMF runs 的平均值，在 SoFi 中。这个“平均值”选项在对许多 runs 进行平均时相关，这会导致在 IGOR 中存储为单个 PMF runs 时出现“内存不足 out of memory”问题。

请注意，在对 **bootstrap runs** 进行平均时（见 6.4.1

7.5.2. 平均（分辨率）和数据选项

用户可以控制结果导入的分辨率类型。默认情况下包含“原始 raw”时间分辨率的平均值。

“每小时 hourly”、“每日 daily”、“每周 weekly”、“每月 monthly”和“每年 yearly”的分辨率可以添加，并将针对每个单独的 PMF runs 以及单个 runs 的平均值进行处理。此外，用户可以控制在 SoFi 中应评估哪些数据（见图 88

默认情况下，所有数据都被导入，包括所有统计数据。

请注意，对大量 PMF runs 进行平均可能会导致平均过程非常缓慢。当从平均选项“存储的数据 data stored of”列表中禁用“评估分位数 evaluate quantiles”复选框时，不会评估分位数统计。结果将仅包含平均值和标准偏差，可将计算时间降低 2-3 倍。

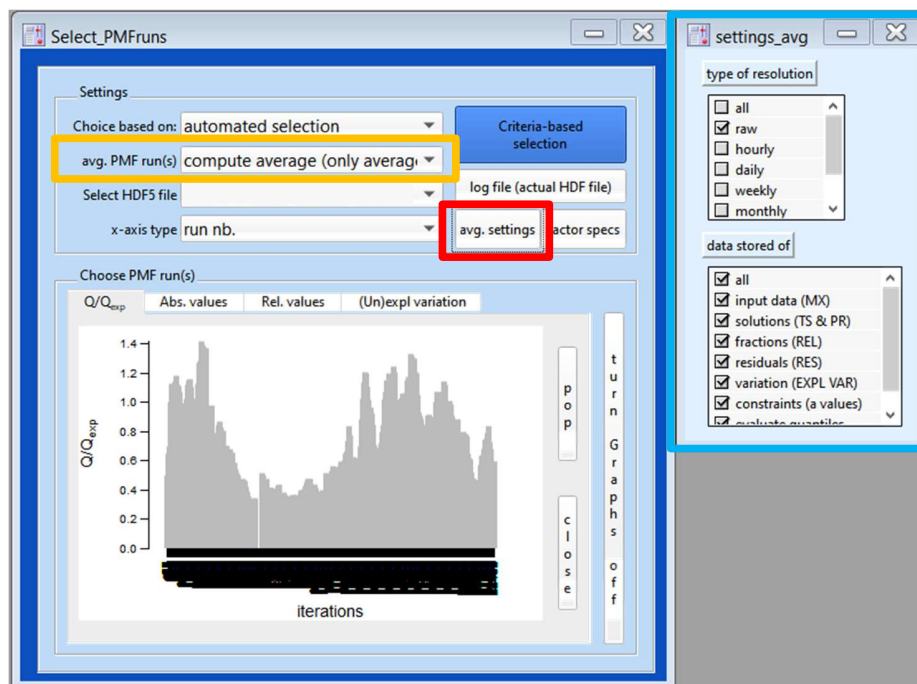


图 88

7.5.3. 自定义因子名称

自定义因子名称（有关更多详细信息，请参阅 6.3.1 图 89

请注意，组合因子必须遵循真实因子的列表，即它们必须位于列表的末尾。

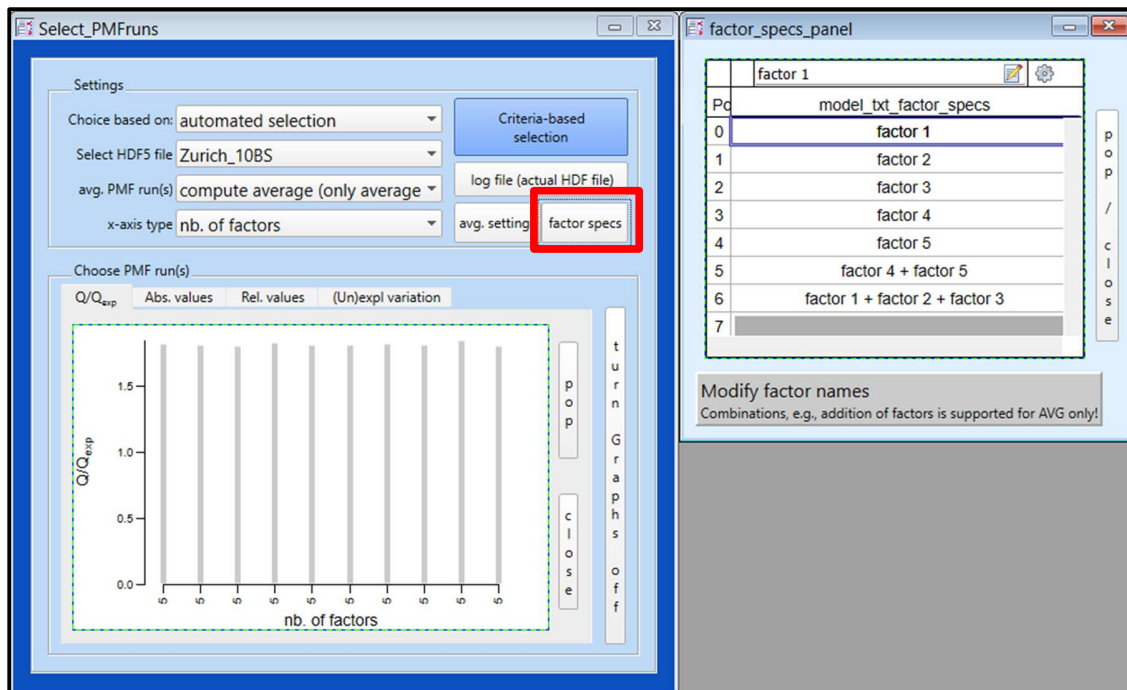


图 89: 因子规格选项用于修改因子名称。两个及以上因子的组合也是可能的，例如在此示例中，SOA（因子 4 + 因子 5）和 POA（因子 1 + 因子 2 + 因子 3）。

7.6. 结果文件夹的技术信息

下面介绍 SoFi 结果文件夹的结构，以及保存当前 PMF call 主要设置的两个参数文件，并说明单个 PMF runs 结果文件夹中保存的信息。这只是为了更好地理解，这里不应更改任何波形，因为这可能导致致命错误。

7.6.1. 结果文件夹的结构

根目录:SoFi:结果 root:SoFi:Results

- “:Variables” 与主结果面板相关的所有变量。用户不得在此处更改、删除或添加任何内容。
- “:Overview” 与平均结果相关的 IGOR wave（见 7.5）
- “:Final” 此文件夹包含最终的时间序列和因子源谱矩阵，包括均值和标准差，以及中位数和分位数。这些数据对于后续的 SoFi 分析非常有用，例如使用 ZEFIR 进行风和轨迹分析（Petit et al., 2017）
- “:Plot” 绘图用的 IGOR wave
- “:Graph_macro” 用于从结果面板标签页重建图形的宏

- “[HDF filename]” 结果文件夹在此列出，例如，灵敏度_I (sensitivity_I)，以下简称 sens_I（见图 90）
- “:[HDF filename]:Parameter” 参数文件（一般数据和误差矩阵、时间序列、变量信息、参数文件、ME-2 的收敛标准、名称字符串、单位字符串、日志文件、组信息）
- “:[HDF filename]:Overview” 与 PMF 调用相关的所有表示（见 7.2）
- “:[HDF filename]:Solutions” 包含 PMF 调用的各个 PMF runs，如果这些 runs 已被选中进行分析（见 7.2）

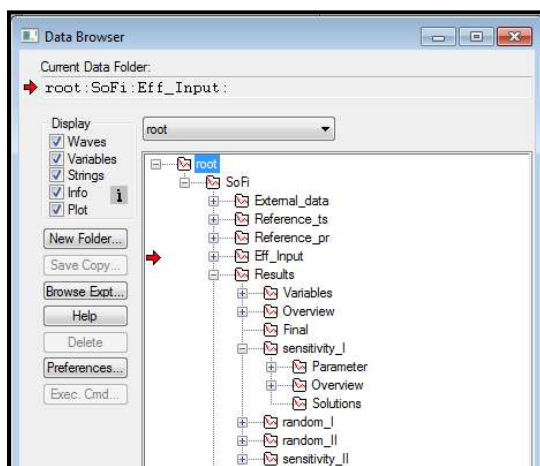


图 90

7.6.2. 参数文件

有两个主要的参数文件声明了 PMF 调用的所有相关信息，solution_para 和 batch_file。它们都存储在 root: SoFi:Results:[RUNNAME]:Parameter:。

7.6.2.1. 解参数 Solution para

Solution_para 是一个文本 IGOR wave，保留了 17 个位置用于各种一般信息：

- ” solution_para[0]” 求解器 solver 信息
- ” solution_para[1]” PMF 迭代过程中的归一化
- ” solution_para[2]” Robust 模式及其阈值
- ” solution_para[3]” 缺失数据的阈值
- ” solution_para[4]” 矩阵缩放因子
- ” solution_para[5]” C3 值
- ” solution_para[6]” 权重模型 weight model 类型，包括参数
- ” solution_para[7]” 权重应用 weight application 类型

" solution_para[8]"	与 CO2 相关的权重
" solution_para[9]"	质量闭合
" solution_para[10]"	相对权重 (C 值)
" solution_para[11]"	fpeak 值
" solution_para[12]"	a 值类型和约束信息的 a 值
" solution_para[13]"	保留用于拉动 pulling 类型
" solution_para[14]"	平均 averaged 信噪比公式
" solution_para[15]"	逐单元 cellwise 信噪比公式
" solution_para[16]"	保留用于统计误差传播
" solution_para[17]"	滚动机制
" solution_para[18]"	多种时间分辨率 multi-time
" solution_para[19]"	仪器类型

7.6.2.2. 批处理文件 Batch_file

批处理文件是一个数值矩阵，包含每次 PMF runs 的信息。

其结构为：batch_file[PMF run 次数][根据约束数量而变化]

batch_file [0][0]	因子数量
batch_file [0][1]	固定因子数量的迭代索引
batch_file [0][2]	滚动索引
batch_file [0][3]	总索引，即批处理文件的行维度
batch_file [0][>4]	约束信息（随时间和/或因子源谱文件变化）。每个约束（时间序列、因子源谱文件）都有一个单独的列。

7.6.3. 主要信息

用户定义哪些 PMF runs 应被考虑用于分析，并在按下“III.b 分析选择的 PMF runs”“III.b Analyze chosen PMF run(s)”后，单个 PMF runs 被导入（见 8.1

“:[HDF 文件名]:Solutions:run_0”

其中 0 代表 PMF run 的编号，即此案例的第一次 run。结果文件夹通常包含以下信息：

“curr log”	当前 PMF runs 的日志文件（ME-2 的日志文件 ME-2 log file）
“Input data”	显示当前 PMF runs 的 PMF input 文件。这与例如 bootstrap 策略（6.46.5
“ini_file”	包含传递给 ME-2 的指令文件的文本 IGOR wave。
“fpeak_matrix”	如果执行过 fpeak runs，则保存相关 fpeak 信息。

下一个结果波存储在相应的分辨率文件夹中，即“raw”、“hourly”、“daily”、“weekly”、“monthly”和/或“yearly”。

“sol_ts/pr”	是 PMF 结果
“pr/ts_constr”	包含 a 值约束的信息。它是一个 3D 矩阵，y 轴上有因子的数量，x 轴上是前两层的 a 值下限和上限。第三层包含每个受约束模型条目的 a 值。
“REL_pr/ts”	时间序列和因子源谱矩阵的相对矩阵。相对因子源谱矩阵是按贡献加权的。
“RES_mx”	残差矩阵
“RESW_mx”	缩放残差矩阵
“RESABS_mx”	绝对残差矩阵
“RESWABS_mx”	缩放绝对残差矩阵
“Q_mx”	Q 矩阵
“EXPL_mx”	随时间变化的解释和未解释变异矩阵因子，x 代表时间，y 代表变量，z 代表因子数量。
“RES_ts/pr”	所有不同类型的残差随时间/变量变化的平均残差矩阵
“EXPL_tseries”	随时间变化的解释和未解释的变化程度
“EXPL_profile”	随因子源谱变化的解释和未解释的变化程度

8. SoFi – PMF 分析

8.1. 选择一个 PMF run

PMF_analyze 面板 (图 91 图 61)

- “General information” 包含四个部分：“选择用于探索的 PMF runs **Select PMF run for exploration**”、“相邻列表中所有 PMF runs 的信息 **Info for all PMF run(s) in adjacent list**”、“所选 PMF runs 的设置 **settings for chosen PMF run...**”和“导出 PMF 解 **Export PMF solution**”
- “Overview plots” 因子源贡献时间序列、因子源谱以及概述残差图（绝对残差、缩放残差、绝对缩放残差、Q）
- “Detail plots” 占比图、因子图、散点图、残差分析、残差直方图、相关矩阵、HR 数据图
- “Averaged solution” 用于评估 PMF 误差的图，时间依赖的因子源谱，随时间变化的 α 值。仅适用于平均解。

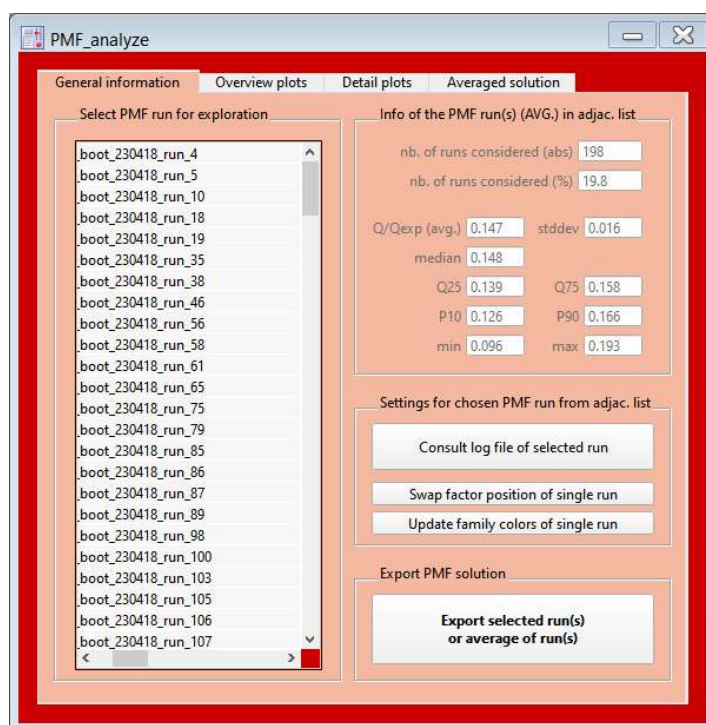


图 91

8.2. 一般信息

8.2.1. 选择 PMF runs 以进行探索

用户选择应该检查的 PMF runs（一次一个 run）。在“数据 DATA”子面板中也提供了相同列表的下拉菜单。这允许快速浏览，以便在需要调查多个 PMF runs 的情况下使用。

当没有选择任何内容时（按下 SHIFT 键会取消当前选择），如果选择了“计算平均值 compute average”（见 **Fehler! Verweisquelle konnte nicht gefunden werden.**），则平均解将作为结果呈现。

8.2.2. 相邻列表中所有 PMF runs 的信息

如果启用了“compute average”，该部分会显示参与平均的 PMF runs 的一些概览统计信息。（见 **Fehler! Verweisquelle konnte nicht gefunden werden.**）

8.2.3. 从相邻列表中选择的 PMF runs 的设置

“consult log ...” 会弹出 ME-2 为当前 runs 写入的日志文件。

“Swap factor position...” 如果对不受约束或松散约束的因子进行了位置交换，用户可以重新定义因子的顺序。 交换仅影响 IGOR 中当前/选择的 PMF runs（图 92）

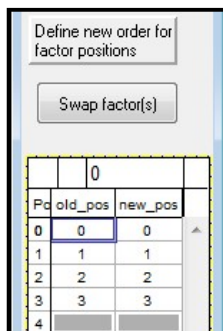


图 92

“Update family colors...” 此按钮仅在 SoFi 检测到用户定义的 family 时出现（见 3.4.9 这仅影响 IGOR 中当前/选择的 PMF runs。）

8.2.4. 导出 PMF 的最终解

一组具有合理环境意义的 PMF runs 结果，定义为 PMF 的最终解，可以导出为 .itx、.txt 和 .dat 文件。根据用户在概览中选择的是单个解还是平均值（如果未选择单独解），相应

的解将被导出。如果用户选择将平均解导出为 .itx 文件，可以选择仅导出平均时间序列/因子源谱或包含所有统计数据的解。按下“导出选定或平均 PMF runs **Export selected or average PMF run(s)**”按钮会打开一个对话框，用户可以在其中定义文件名和文件格式。因子源谱和时间序列将分别导出。时间序列将以在分析 **analyze**（红色）面板的数据子窗口中选择的时间分辨率导出。按下继续 **continue** 后，将打开一个文件路径对话框。将很快添加以 Zefir 友好格式（包括平均功能）导出数据的选项。

8.3. 子面板图形 Graph 和数据 Data

一旦 PMF 结果 Result 面板（红色面板）弹出，两个子面板“图形”和“数据”也会弹出。

“**Graph panel**” 包含与图形相关的选项，例如均值 **mean** 或中位数 **median** 表示、表示模式（线 **line**、点 **point**、交叉 **cross** 等）、符号大小等。

“**Data panel**” 包含与数据类型相关的选项，例如时间序列或日变化曲线。

这些子面板（图 935.3.15.3.25.3.3

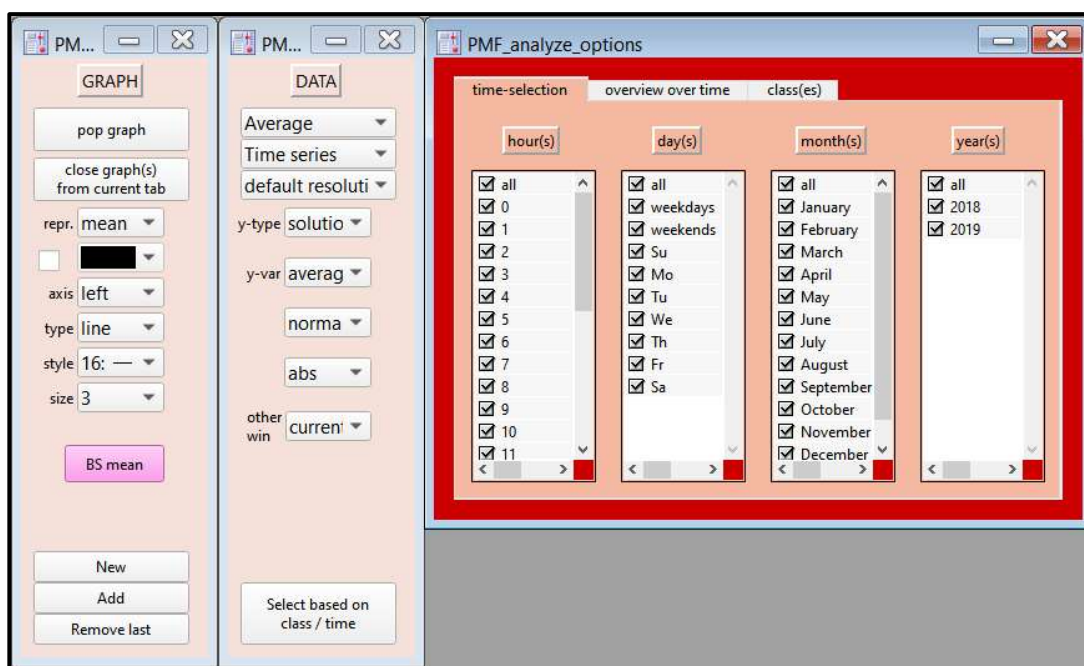


图 93

数据面板的顶部下拉菜单允许选择 PMF runs，或者在计算了平均值的情况下，选择应绘制的平均值，而无需来回切换到“常规设置 **General settings**”标签页。

注意：在检查完特定标签页上的结果后，不要忘记点击“关闭图表 **close the graphs**”。否则，SoFi 会积累过多 IGOR waves 从而消耗内存，变得缓慢。

8.4. 概述图

因子时间序列、因子源谱和概述残差可以被可视化。如果没有其他定义，SoFi 将使用默认的 Igor 颜色（见 3.4.8.2 图 94 图 94

对于受限信息，可以通过检查“a 值”进行可视化。例如，因子源谱在所示的解中受到约束。图 94

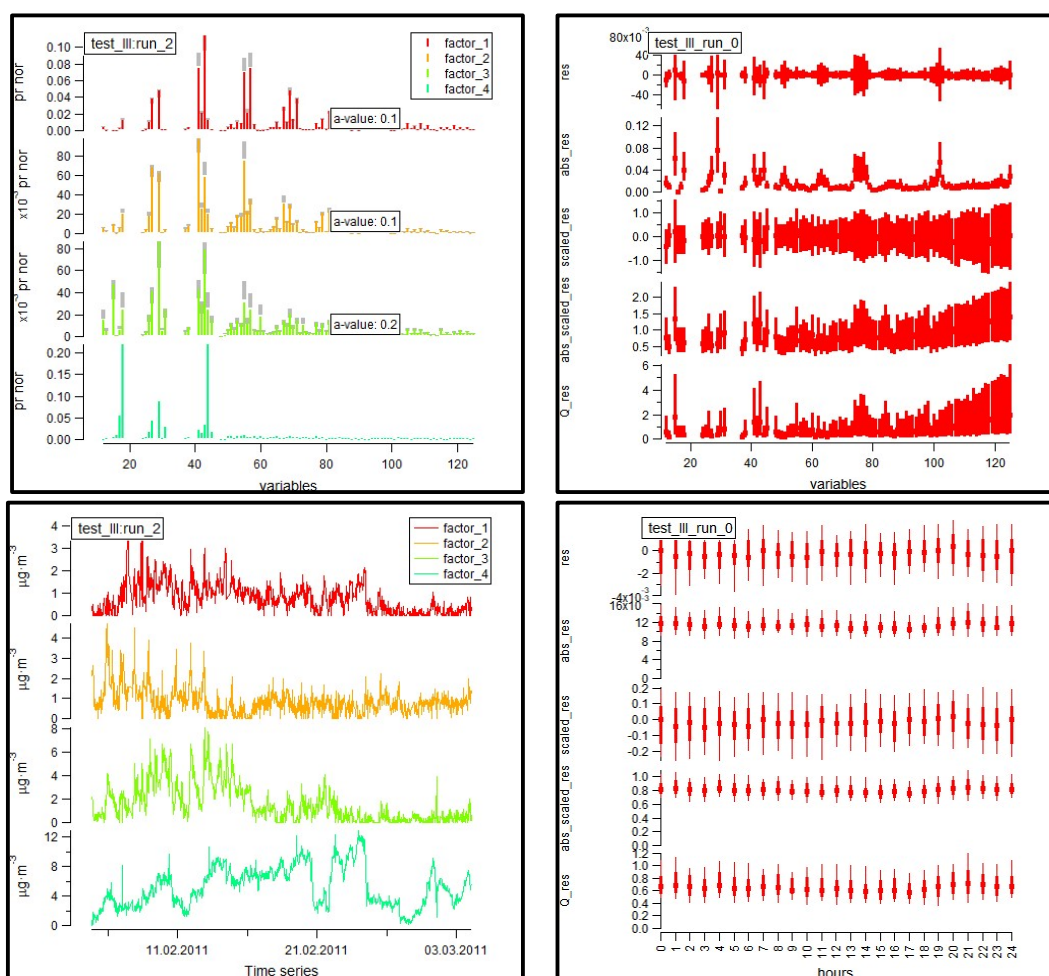


图 94: 因子源谱（左上）和变量的残差（右上）。变量的残差由均值和四分位间距（IQR）表示。因子贡献（左下）和日变化残差（右下）用中位数表示，IQR 的条形图和 10%到 90%的须状图。

8.4.1. 时间序列特定选项

如果导入到 IGOR 中（见 7.5.2

如果数据分辨率允许，所有时间序列图可以绘制为“正常 normal”或周期 cycle 图。可用的周期图包括日变化、每周的日变化 diurnal cycle over a week（即日变化被分为工作日，

以更好地理解不同工作日的平均日间行为)、每周周期、每月周期和每年周期。对于所有周期图,第一个值在末尾重复。与工作日相关的图从星期天开始。

对于默认分辨率下的 **Figaero TD PMF** (即时间序列是温度维度),这些周期图显然不可用。相反,可以将热图平均为一个单一的热图,表示一个因子的平均波动行为,使用“解吸温度 **desorption temp.**”选项。

对于 **Figaero TD PMF**,选项“基于滤膜的分辨率 **filter-based resolution**”也可用。使用此选项,单个热图(即每个测量的滤膜)被平均为一个点。此时,时间序列现在表示输入中定义的给定时间序列,而不再是温度。在这里,所有来自“正常”**PMF**的周期图再次可用。

用户可以在时间序列的绝对值和占比两种显示方式之间切换。

8.4.2. 因子源谱特有选项

在“**GRAPH**”面板中,前面三个可用于因子源谱文件的选项与图表中的 **x** 轴表示相关。选择“**x dim continuous**”时,将使用连续的 **x** 轴,这对于 **x** 轴上没有质量的数据显示非常有用,例如金属或传统的离线 **PMF**,以及某些元素/变量被列入黑名单的情况。当勾选“**x-label**”时,SoFi将使用定义的“**pr_txt**”中的文本信息,而不是 **x** 轴上的数字表示。

- | | |
|------------------------------|--|
| “Factor specs” | 将把定义的因子名称添加到相应的因子中。如果没有定义因子名称,则因子将被称为“ factor [nb] ”。 |
| “contr. weighted var” | 选项将在右侧轴上添加每个变量的加权贡献,使用黑色标记 (图 95) |

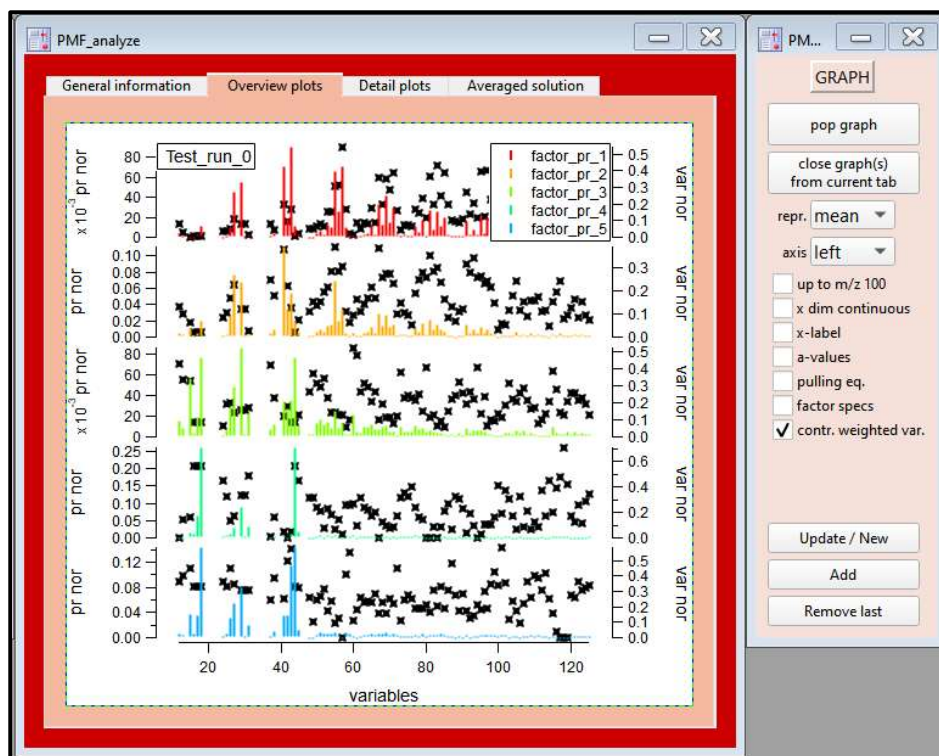


图 95

如果为所选解指定了用户定义的 family，则可以使用“按 family 着色 color with families”选项。与 HRfamily 图（8.5.6.1

8.5. 详细图

8.5.1. 占比 fraction

SoFi 可以通过柱状图 bar 和饼 pie 图，以占比形式显示：源贡献，已解释的变化程度，未解释的变化程度。这些结果既可以沿时间维度，也可以沿因子源谱维度展示。对于与质量相关的柱状图，可以按照以下方式进行分箱：按分位数分箱，按质量分箱。按质量分箱时，还可以显示某个质量区间被记录的频率。对于时间序列饼图，还可以加入外部数据。例如，可以加入无机组分，使饼图表示总 PM 质量，而不仅仅是 PMF 得到的有机组分。此外，用户还可以针对每个因子，以饼图方式显示 HR family 信息。用户可以自由选择：bar; pie

chart

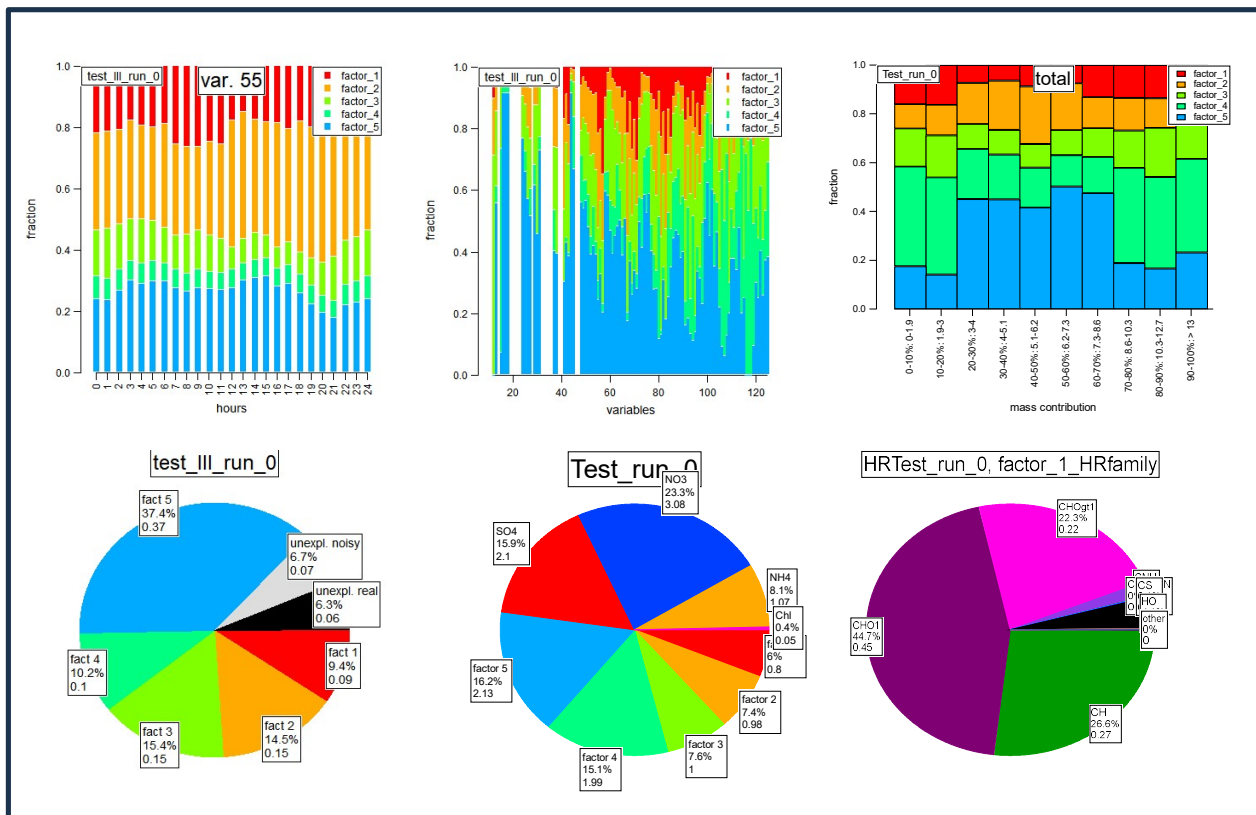


图 96: 变量 m/z 55 的日变化条形图（左上），源贡献加权条形图（中上），基于质量的贡献条形图，按分位数分箱（右上），在用户定义的时段上以饼图表示的（未）解释的变化程度（左下），总 NR-PM 的饼图，包括有机 PMF 结果和无机物质，以及在 HR-AMS 分析中对因子 1 的 family 贡献的饼图。

8.5.2. 标准 Standard

标准标签页允许绘制用户选择的多种示踪物 tracer 的信息随时间变化（时间序列图 time series plot）或随剖面变化（变量图 variable plot）。图 97

用户可以向同一张图中加入多个 trace，这些 trace 可以来自：当前 PMF runs，其他任意 PMF runs。只需返回选择标签页，选择另一个 PMF runs，然后将对应 trace 添加到已有图形即可。

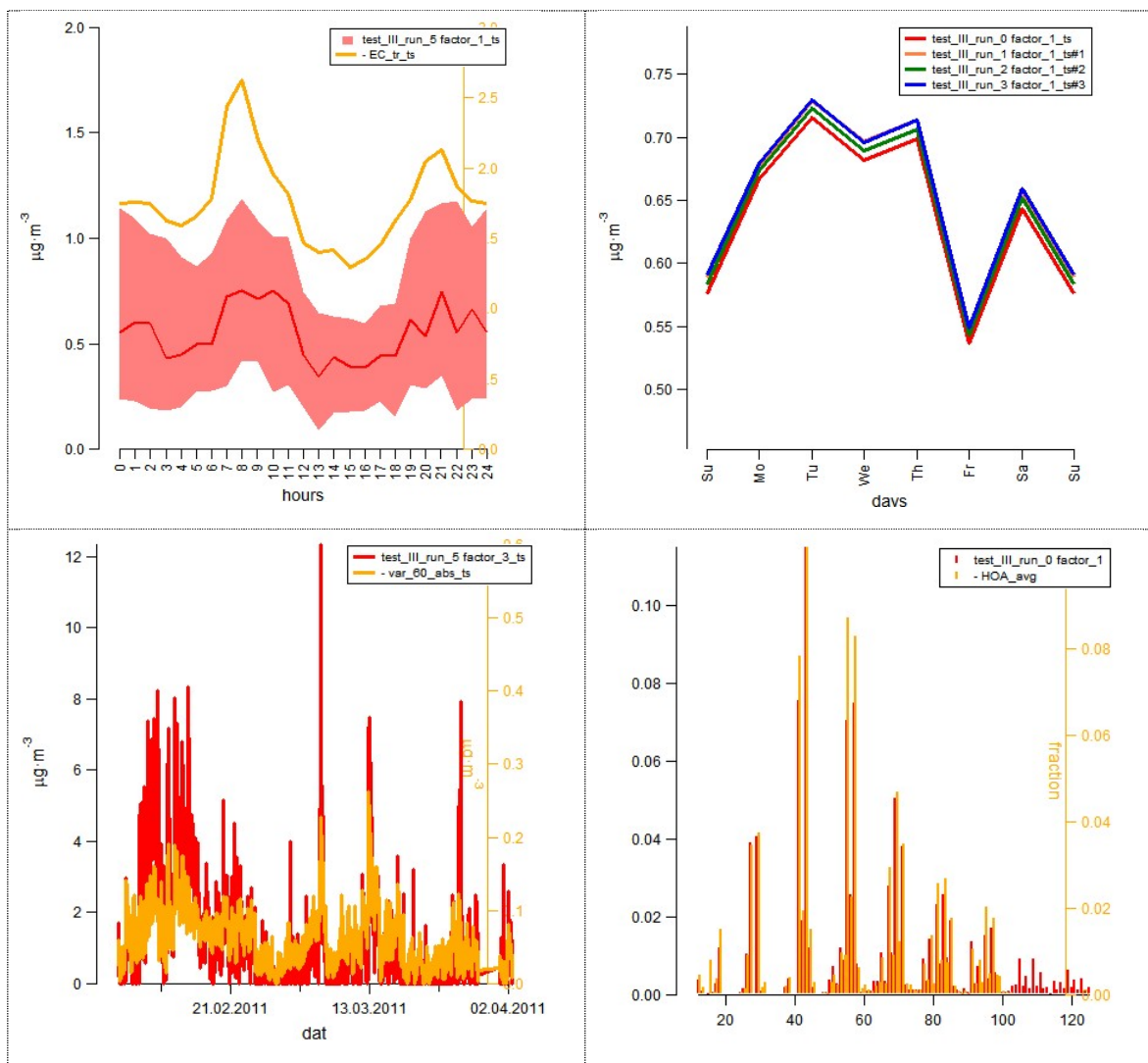


图 97: 因子 1 (假设交通源) 的日变化及其示踪物 EC (左上), 来自多个 PMF runs 的因子 1 的周变化 (右上), 因子 3 (假设 BBOA) 的时间序列及其代理 m/z 60 (左下), 以及因子 1 的因子源谱及相邻的外部交通源参考谱 (右下)。

此外, 从“概览图 **Overview plots**”或“占比 **fraction**”标签页生成的图可以加载到“标准”标签页中。这允许用户为现有图形添加其他 **trace**, 以便更好地可视化和比较。在选项窗口温度下选择“概览图”或“占比图”以添加来自概览或分数图的图。按下“标准 **standard**”标签页或“新建 **new**”按钮将最终添加预先存在的图 (图 98 图 99

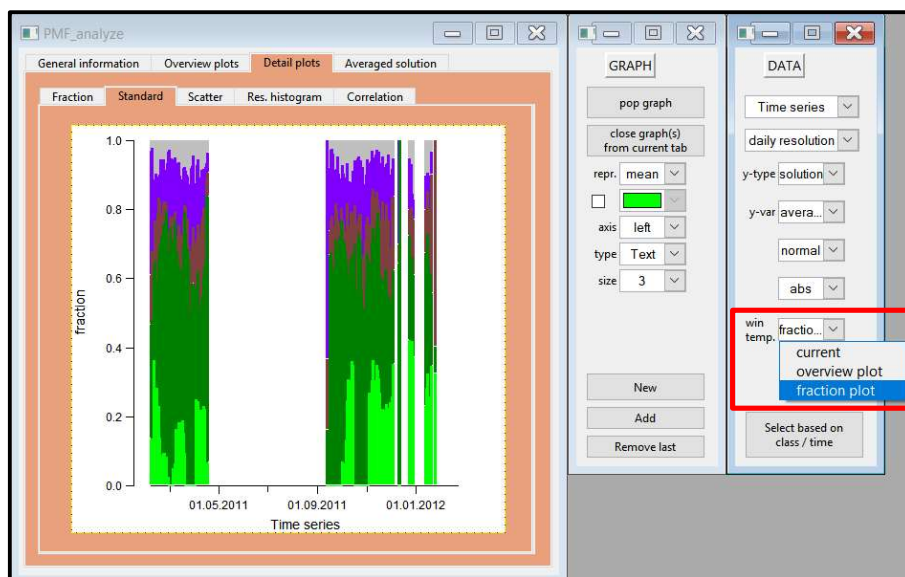


图 98 从“概览图”或“占比”标签页中绘制现有图形。

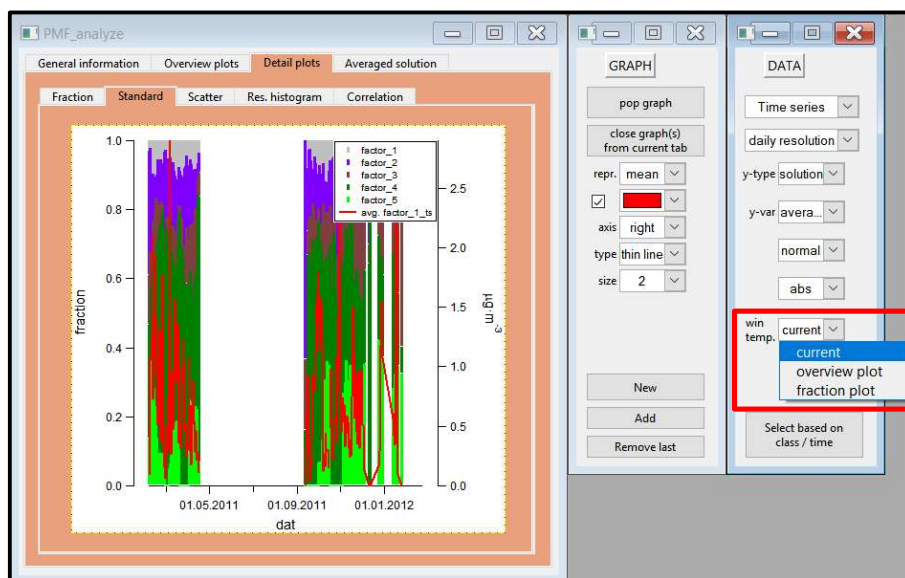


图 99 从“标准”标签页向现有图形添加新轨迹。

8.5.3. 散点图 Scatter

“Scatter” 标签页允许将数据绘制为散点图，并根据第三个变量进行颜色编码。散点图既可以沿：时间维度，变量维度绘制，并支持多种用户自定义 tracer。图 100

用户既可以比较当前 PMF runs 中的不同 trace，也可以比较不同 PMF runs 之间的 trace。通过在以下轴上勾选：“hold”。可以将对应 trace 固定在此前选择的 PMF runs：y-axis；x-axis；color-axis。这样，在切换至另一个 PMF runs 时，被“hold”的 trace 会保持不变，从而方便不同 PMF runs 之间的直接比较。

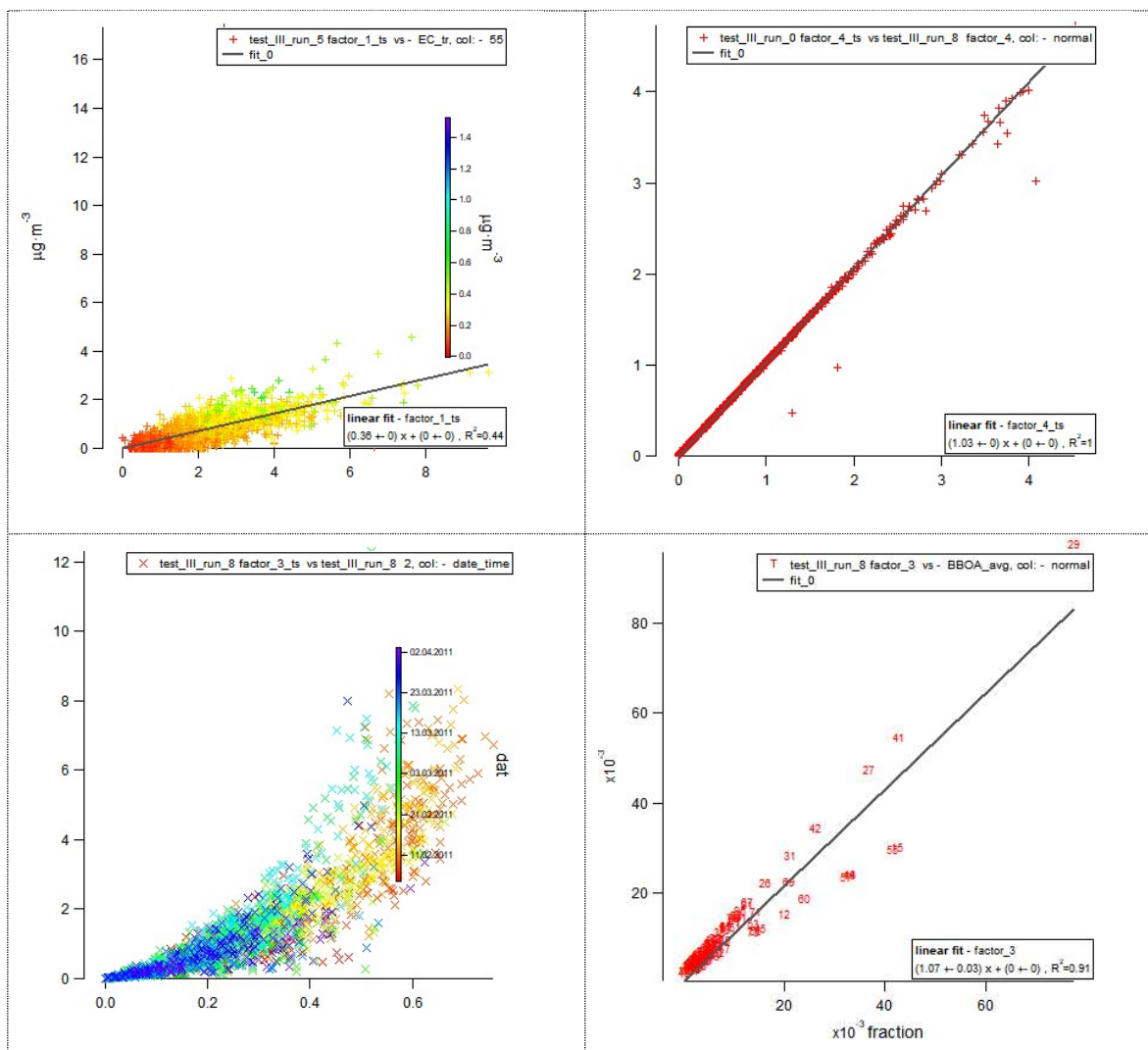


图 100: 因子 1（假设交通）与其 tracer EC 按其代理 m/z 55 进行颜色编码（左上），因子 4 与另一个 PMF 的因子 4（右上），因子 3（假设 BBOA）与其代理 m/z 60 的解释的变化程度按时间进行颜色编码（左下），以及因子 3（假设 BBOA）与外部 BBOA 源谱（右下）。

8.5.3.1. 模型减法 Model subtraction 选项

SoFi Pro 提供了将特定模型结果与模型理想上应重现的测量量进行比较的可能性。在 ACSM 数据中，针对一次和二次有机气溶胶源，用户可以例如去除特定变量的一次源贡献，并将二次源的模型结果的占比（通常是因子源谱的条目）与二次测量贡献进行比较。有关其应用的更多细节，请参阅 Canonaco et al., (2015) 了解有关其应用的更多详细信息。

示例

模型中二次因子的 m/z 44 和 43 的占比，即 f_{44} 和 f_{43} 应与仅来自二次源的测量 f_{44} 和 f_{43} 的占比进行比较。

y-type 和 x-type 必须显示为“模型减法 **model subtraction**”。y-var 和 x-var 分别为 m/z 44 和 m/z 43，并且应勾选“占比 **fraction**”选项。右侧的表格在图 101

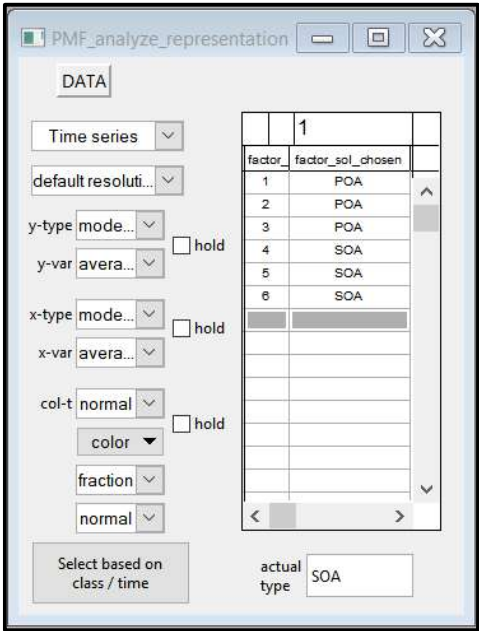


图 101: 模型减法面板，前面三个因子属于“POA”组，最后三个因子属于“SOA”组。正在进一步分析的因子是定义在“实际类型”下的，即 SOA。

鉴于 PMF 中的线性关系，测量点的变化程度（红点在图 102 图 1026.5

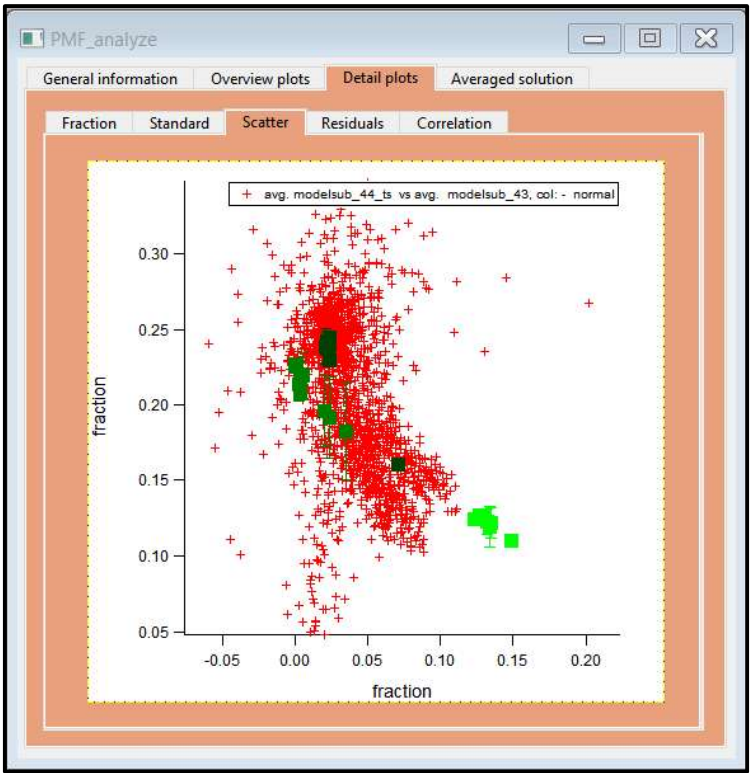


图 102: 模型减法选项的结果, f_{44} 对比 f_{43} 。红色点为扣除 POA 对 f_{44} 和 f_{43} 贡献后的实测值, 浅绿色和深绿色点为 SOA 因子源谱中的相应质量分数。

请注意, 在对绝对值进行“模型减法”绘图时, 模型因子源谱中的“占比”不能直接与实测贡献进行比较。

因此, 在这种情况下, 只绘制从实测变量中扣除了不需要进一步分析的因子贡献后的结果。此外, 检查减去测量点的贡献形状 (红点在图 102)

8.5.3.2. 选项: 多元线性回归 Multilinear regression

SoFi Pro 提供在外部因子与 PMF 因子之间或外部因子与 PMF 因子加外部因子之间进行多元线性回归的选项。这可以用于例如对 eBC 与 PMF 燃烧因子 (如交通和生物质燃烧) 进行多元线性回归。 y 类型和 x 类型必须读取为“多元线性回归”。对于 y 类型, SoFi 列出了所有外部时间序列数据。对于 x 类型, 用户必须选择 (勾选) 应用于多元线性回归的因子 (和外部数据)。为了方便导航, 用户可以在 x -var 的下拉菜单中选择仅 **solution** 或 **solution & external**。SoFi 支持最多 10 个变量的多元线性回归。

SoFi 将结果呈现为散点图, 其中外部数据与拟合参数乘以因子的总和进行绘制。拟合参数显示在右下角。多元线性拟合可以强制通过零 (原点), 或者可以允许有截距 (复选框“MLR 通过零 MLR through zero”), 在这种情况下, 这个截距也会与拟合参数一起列出。

8.5.4. 残差 Residual

此标签页包含与残差矩阵相关的功能。一方面, 可以访问整个 (缩放的) 残差矩阵。另一方面, 可以为全部数据以及单个变量或类别 (如果之前定义过) 绘制残差直方图。在 C 值方法的情况下, 可以在这里检查和比较不同仪器类别的直方图。

8.5.4.1. 整个残差矩阵 Entire residual matrix

可以直观地检查整个 (缩放的) 残差矩阵。控制矩阵颜色编码的相邻滑块允许动态检查 (缩放的) 残差矩阵的各个区域, 并验证矩阵的条目是随机的, 即在时间和变量的所有范围内都是无结构的 (见图 103)

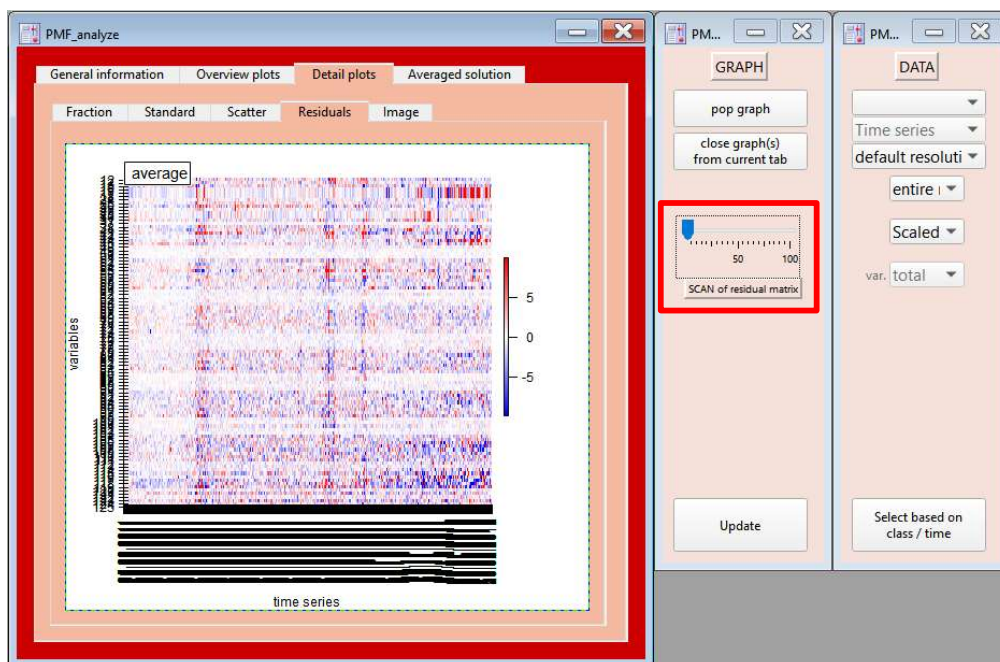


图 103: 基于用户定义的滑块“残差矩阵的扫描 SCAN of residual matrix”值，在子面板“图形 graph”中，缩放的残差矩阵的子部分在-0.4 到+0.4 之间进行颜色编码。

8.5.4.2. 残差直方图 Residual histogram

残差直方图是一种简单快速的方法，用于调查 PMF 结果是否包含某些系统性的低估或高估（见图 104： 104

可用的残差直方图包括：

- “residual” 仅考虑残差。这揭示了数据建模的好坏。形状应该是单峰的。预计不会呈现高斯形状，因为 PMF 并不直接拟合残差，而是缩放残差（见下文）
- “abs. residual” 绝对残差直方图。直方图将仅位于正象限。
- “scaled residual” 缩放残差直方图。该量揭示了 PMF 模型对数据的拟合程度。理想情况下，概率密度函数是单峰且类似高斯。此外，根据 Paatero 和 Hopke (2003)，缩放残差直方图主要应在 ± 3 之间散布。
- “abs. scaled residual” 绝对缩放残差直方图。直方图将仅位于正象限
- “Q residual” Q 直方图。直方图将仅位于正象限。

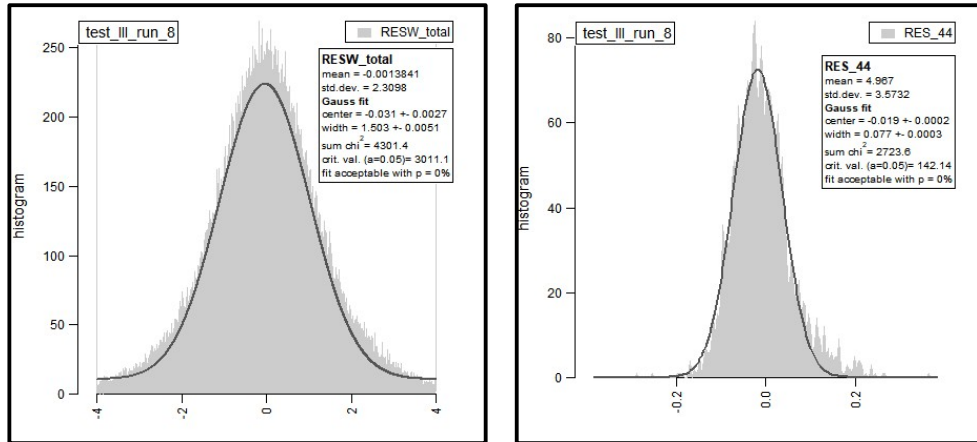


图 104: 104

如果实际残差的形状与此不同，表现为双峰或多峰，或在负侧或正侧出现一些高峰，则需要进一步调查 PMF 模型。如果施加的约束与数据集的自然变化程度不匹配，可能是因子的数量过高/过低等。

类别的相对权重（C 值方法）

如果研究者对某些类别的不确定度进行了相对重新加权，则可以分别绘制这些类别的残差或缩放残差直方图，并通过统计检验判断这些分布是否可能来自相同的总体分布。

以下统计分析可以直接从 SoFi 访问：

“ANOVA (equal variance)” 一种单因子 ANOVA 检验，假设各类别缩放残差的抽样分布均服从正态分布；各类别具有相同方差。（经典 ANOVA 检验使用 F 分布或 T 分布用于两个类别）。此外，还计算了 Bartlett’s test 的等方差检验，结果在相邻文本框中表示。

“ANOVA (unequal variance)” 假设所有（缩放的）残差采样分布服从基础正态分布一元方差分析测试（Welch 测试）

“ANOVA (non-parametric)” 进行非参数或秩的一元方差分析测试。（进行考虑不同类别成员数量的 Dunn-Holland-Wolfe 测试）

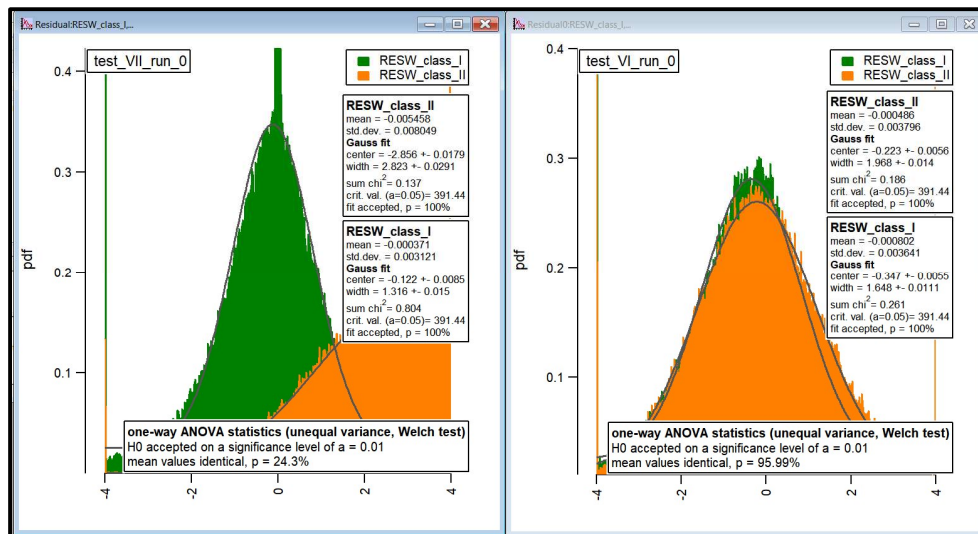


图 105 左图为误差未重新加权的情况；右图使用 C-value 方法对两个类别的缩放残差进行优化，即“equilibrate group scaled RES”。

根据每个统计测试报告的 p 值，分析师可以判断测试假设 H0 真实的可能性，即每个类别的（缩放的）残差分布的均值来自相同的基础分布。

给出了一个例子图 105

8.5.5. 图像 Image

在第一个窗口的下拉菜单中，用户可以选择图像绘图 image plot 的类型（位于分析（红色）面板旁边）。目前，SoFi 提供相关性 correlation、风分析 wind analysis 和层次聚类图像绘图 hierarchical clustering。

8.5.5.1. 相关性 Correlations

因子时间序列以及因子源谱可以与外部数据和/或变量进行相关，并且可以相互之间进行相关。以下选项可用：

- “R_Pearson (squared)” R-Pearson 值适用于线性关系
- “R_Spearman (squared)” 也称为秩相关，适用于任何类型的关系
- “Kendall tau (squared)” Kendall 秩相关系数，分析一致对和不一致对
- “R_uncentr (squared)” 余弦相似度，非中心相关系数
- “Fisher transformation” Pearson 相关系数的 Fisher z 转换

对于所有这些选项，表示形式可以是图像、表格或组合格式（图 106）

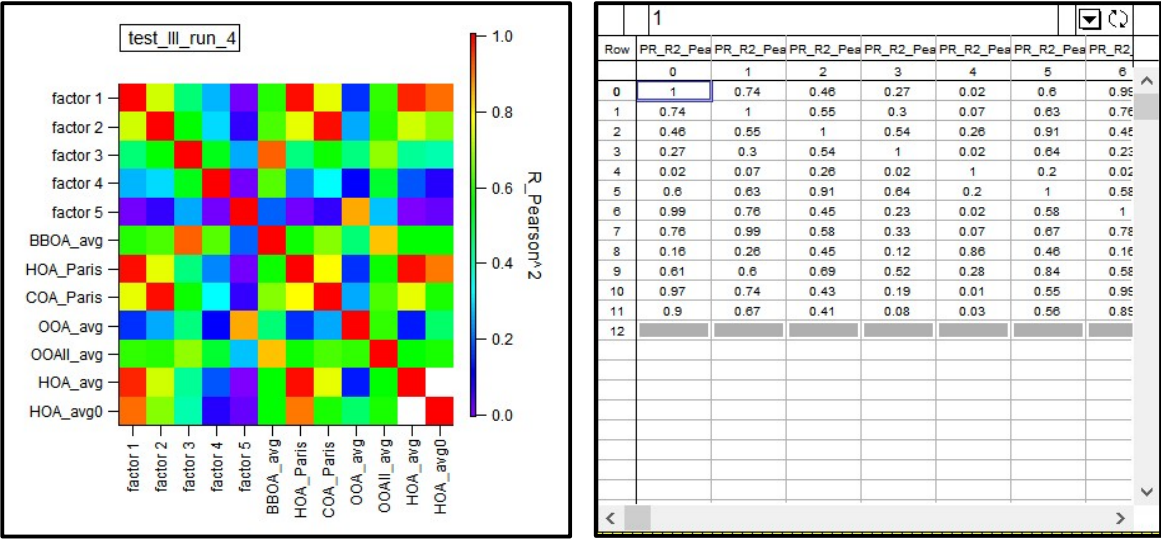


图 106external tracers 之间的相关矩阵图和相关系数表。

在“DATA”子面板上的表格允许用户选择适当的因子和外部或变量（图 107

子面板“GRAPH”还包含表示类型（表格 R 值、表格 p 值、图像、图像+R 值、图像+p 值）和基于阈值的显著性检验的临界值。用户可以在同一面板上调整阈值，并立即检查可接受的最小 R 值。此检验仅在选项 R-Pearson 和 Kendall tau 中可用。

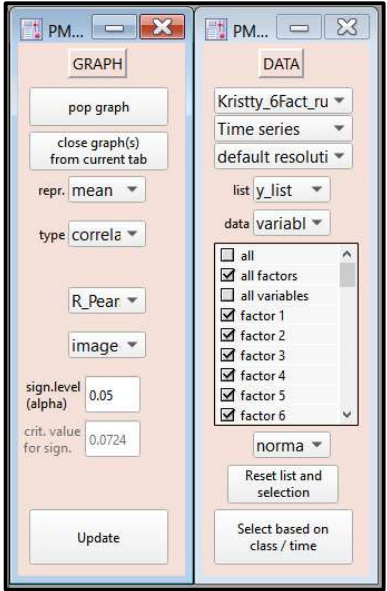


图 107：“相关性 correlation”部分的图形（左）和数据（右）子面板。

8.5.5.2. 风分析 Wind analysis

SoFi 提供风玫瑰图 wind rose、污染玫瑰图 pollution rose、联合概率 joint probability 和非参数风回归 non-parametric wind regression（NWR）图的选项。联合概率和 NWR 的图形相似，

可以在 Zefir 中找到。(Petit et al., 2017)。所有这些图表都需要风速和风向。SoFi 将在处理外部 external（黄色）面板中查找生成的 IGOR wave 名称。因此，默认情况下，使用风速（wind_vel）和基于风向平均（wind_dir_degdep，优选）或风向（wind_dir，普通平均，不基于方向）。如果这些 IGOR wave 不存在，将出现一个弹出窗口，用户可以选择风速和风向 IGOR wave。风速可以是 m/s 或 km/h。结果图将以相同的单位绘制。风向数据需要以度为单位，0° 代表北方方位。用户可以选择五种不同的颜色渐变。

一般来说，非参数回归是一种估计给定一个或多个预测变量值时因变量均值的方法。在 NWR 的情况下，因变量是污染物浓度 C ，预测变量是平均风向 θ 和风速 u 。因此，平均污染物浓度 C 是作为风速 u 和风向 θ 的函数计算的，使用 Kkernel 函数。（亨利等人，2009）

$$E(C|\theta, u) = \frac{\sum_{i=1}^N K_1\left(\frac{\theta - W_i}{\sigma}\right) K_2\left(\frac{u - U_i}{h}\right) C_i}{\sum_{i=1}^N K_1\left(\frac{\theta - W_i}{\sigma}\right) K_2\left(\frac{u - U_i}{h}\right)} \quad (27)$$

这里 C_i , U_i , 以及 W_i 是给定时间点的污染物观测浓度和相应的风速和风向。平滑参数 σ 和 h ，根据经验法则估计(Silverman, 1986)。

高斯 Kkernel 用于风向

$$K_1(x) = (2\pi)^{-\left(\frac{1}{2}\right)} \exp(-0.5x^2), -\infty < x < \infty \quad (28)$$

，而风速则使用 Epanechnikov kernel

$$K_2(x) = 0.75(1 - x^2), -1 < x < 1 = 0 \quad (29)$$

然而，没有证据表明结果对 Kernel 函数的选择特别敏感。

风玫瑰 Wind rose

风玫瑰显示了风的吹向，圆周上每个“辐条”的长度表示风从该方向吹的频率。用户可以选择速度分箱 bin 宽以及最大速度。此外，平静速度 clam speed，即被视为平静/无风的风速上限，可以进行调整。

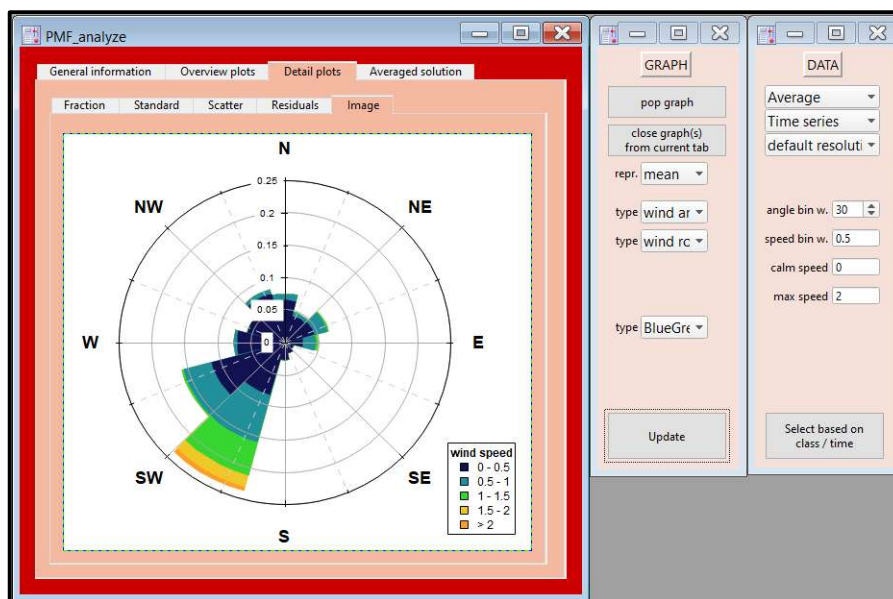


图 108

污染玫瑰 Pollution rose

污染玫瑰采用与风玫瑰相同的绘图结构，但用浓度时间序列替代风速。与风玫瑰类似，角度的分箱大小必须定义。此外，浓度的箱宽和最大浓度 **max conc** 也必须设置。对于下拉菜单，用户可以在结果、外部数据和变量的浓度之间进行选择。

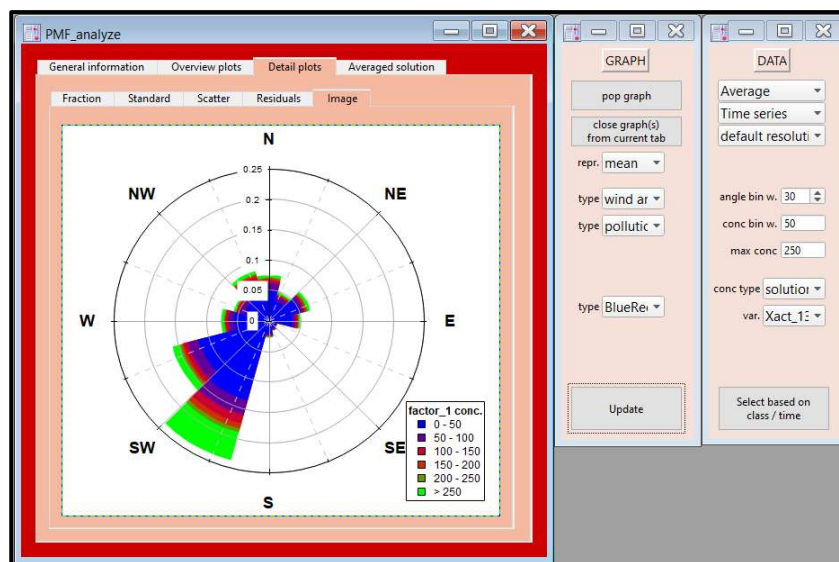


图 109

联合概率 Joint Probability

顾名思义，联合概率显示了风速和风向的分布。它类似于风玫瑰图，后者是可视化风向和风速联合分布的传统方式，但联合概率图更加详细。用户必须给出绘图的最大速度（上限

max speed)。由于这是一个极坐标图，用户必须定义角度 **angle** 和径向 **radial** 分辨率 **res**，即绘制多少个点以使其看起来平滑。

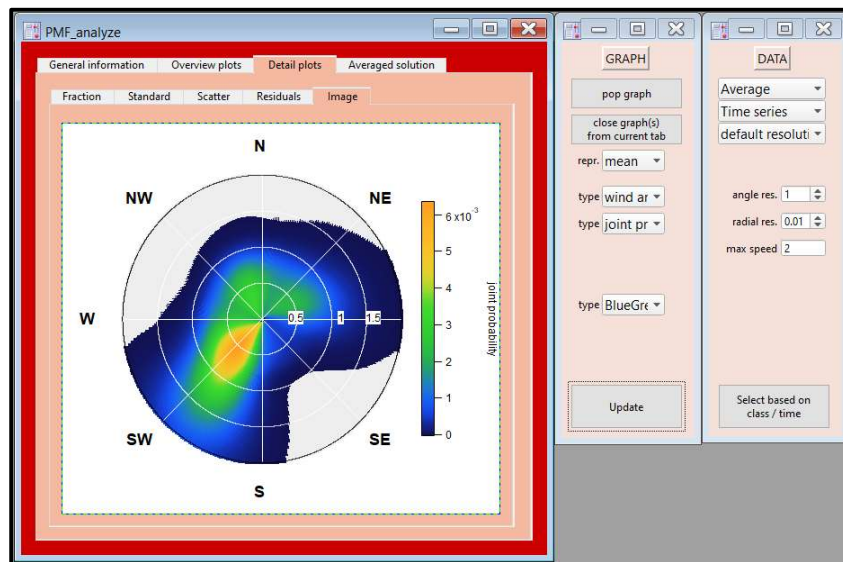


图 110

NWR（非参数风回归 nonparametric wind regression）

NWR 显示所选污染物与风向和风速之间的详细关系。除了在分离同一大致方向的源方面表现更好外，NWR 方法比污染玫瑰图更高效，因为它不需要选择风向分箱数。对于 NWR，用户只需定义最大风速。从下拉菜单中，可以选择所需的污染物浓度（结果、外部数据、变量）。与联合概率类似，必须定义角度和径向分辨率。值越小，分辨率越高，计算时间越长。灰色区域是联合概率过小的区域，即点数太少，无法进行可靠估计。一些用户可能对这个图形比较熟悉，因为它与 ZeFir 中提供的图形相似。（Petit et al., 2017）。

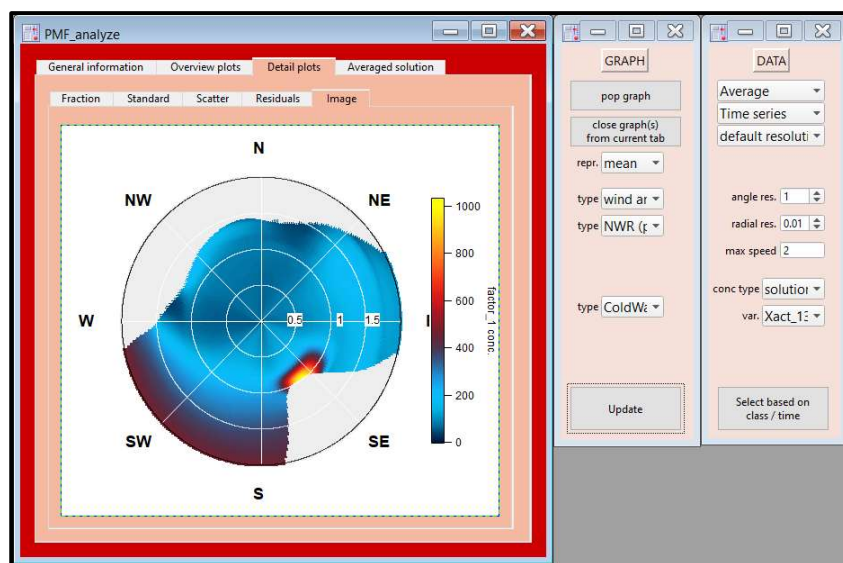


图 111

8.5.5.3. 层次聚类 Hierarchical clustering

此选项仅在 Igor 9 或更新版本中可用。

此选项对所选因子和外部或变量时间序列进行层次聚类，所选自列表中。默认情况下，聚类使用欧几里得距离 **Euclidean distance**。因子和外部的时间序列可以在内部预处理（推荐）：均值中心化 **center**、归一化 **normalize [0,1]**、中心化和归一化 **[-1, 1]**、 $X_i/SD(X)$ 、 $(X_i - \text{mean})/SD(X)$ 。在 SoFi 中有 3 种方法可用于确定在表示多个数据向量的树状图 **dendrogram** 中节点之间的差异性（“连接 **linkage**”方法）。默认情况下，选择单一方法 **single linkage**，其中差异性是每个聚类 **cluster** 中两个最近元素之间的差异性（最近点算法 **Nearest point algorithm**）。完全连接方法 **complete linkage** 也称为最远点算法 **farthest point algorithm**。平均方法 **average linkage** 是每个聚类中点之间所有成对差异性的平均值。

用户可以选择不同的差异性 **dissimilarity** 进行聚类：欧几里得 **Euclidean** 距离或曼哈顿 **Manhattan** 距离，或基于相关性（**R Pearson**）的差异性矩阵。层次聚类在处理 **NaN** 时是不稳定的。因此，对于前两者，SoFi 需要忽略所有至少有一个外部/变量为 **NaN** 的数据点，对于后者，SoFi 可以选择遵循此方法（更稳定）或基于可用数据计算所有对的相关性（可能导致无意义的相关性，因为并非所有相关性都是基于相同数量的点）。

层次聚类图（图 112

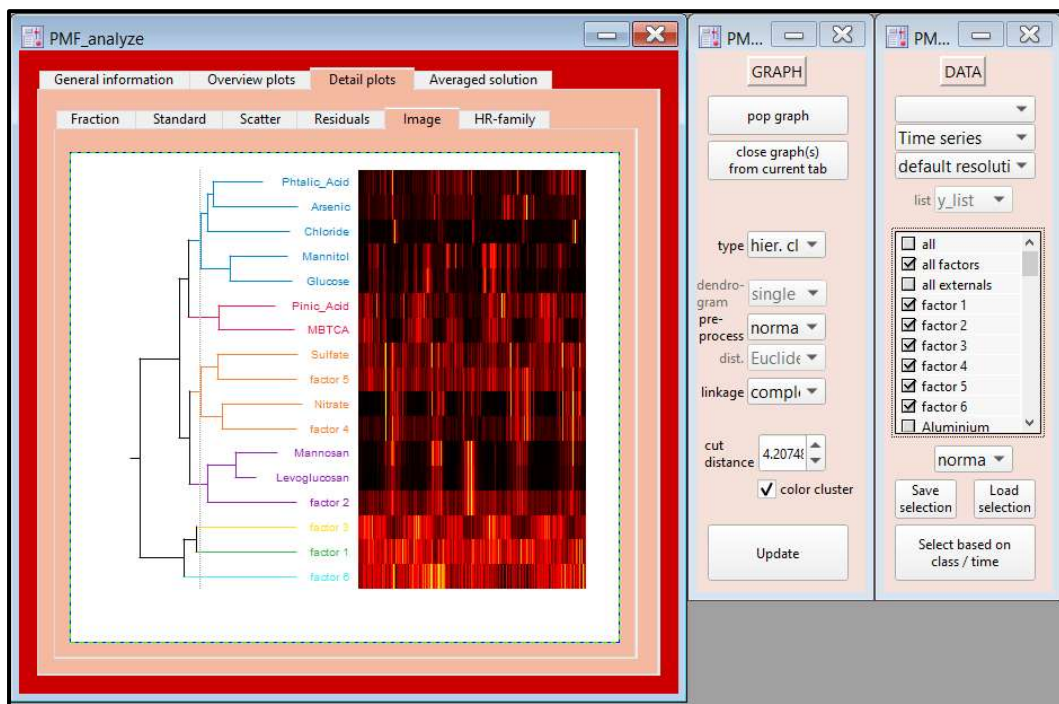


图 112

8.5.6. HR 图

如果存储了包含变量 **family** 信息的文本向量，名称为 “**speciesfamilytext**”，则可以访问 “**HR-family**” 标签页，此外还需存储 **PMF** 输入数据（导入数据，见 4.18.2.3）

8.5.6.1. 堆叠源谱 Stacked profiles

对于 **HR-AMS** 数据分析，**SoFi** 支持因子源谱图的 **family** 堆叠版本。这意味着每个 **UMR** 质量上，所有 **family** 都堆叠在一起，这使得比概览图中的选项更易于阅读，后者中的离子也可以根据 **family** 进行着色。此外，所有因子都配备了 **OM:OC**、**O:C**、**H:C**、**N:C** 和 **S:C** 计算（图 113(Aiken et al., 2008) 和 “**Canagaratna ambient**” (Canagaratna et al., 2015) 方法用于可视化这些比率。

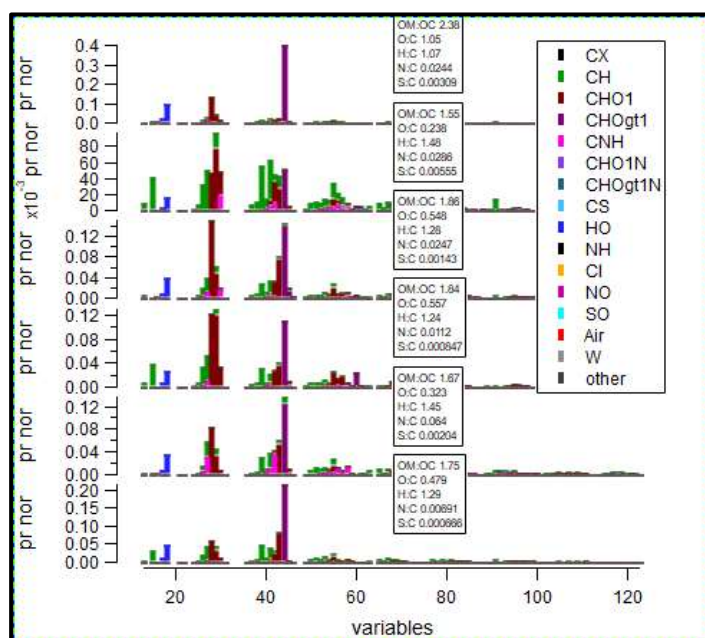


图 113

8.5.6.2. 质量缺陷 Mass defect

在高分辨率质谱中，**Kendrick** 质量(Kendrick, 1963) 通常用于识别仅相差几个基本单位的同系化合物，而 **Kendrick** 质量缺陷 (**KMD**) 分析通常用于可视化化学组成特征。因此 **SoFi** 可以根据 **Kendrick** 原始的质量缺陷定义绘制质量缺陷或更广泛使用定义在大气化学中由 **Aerodyne** 和 **Alton** (Alton 等, 2023):

$$KMD_{\frac{m}{z}} = \frac{m}{z} * \frac{Kendrick_{base}}{IUPAC_{base}} - \text{round}\left(\frac{m}{z} * \frac{Kendrick_{base}}{IUPAC_{base}}\right) \quad (30)$$

KMD 的计算方法是测量的精确质量与 **Kendrick** 基本单位与 **IUPAC** 基本单位的比值相乘的差值，再减去四舍五入的测量精确质量与 **Kendrick** 基本单位与 **IUPAC** 基本单位的比值相乘的

结果。Kendrick（1963）提出 CH₂ 作为基本单位，从那时起也使用了其他单位。SoFi 目前提供 4 个基本单位供选择（表 8

表 8

基本单位 Base unit	Kendrick 基本单位 Kendrick base unit	IUPAC 基本单位 IUPAC base unit
¹² C (= 仅四舍五入)	12	12
¹² C ¹ H ₂	14	14.01565
¹⁶ O	16	15.9949
¹² C ¹ H ₂ ¹⁶ O	30	30.01055

质量缺陷图（图 114

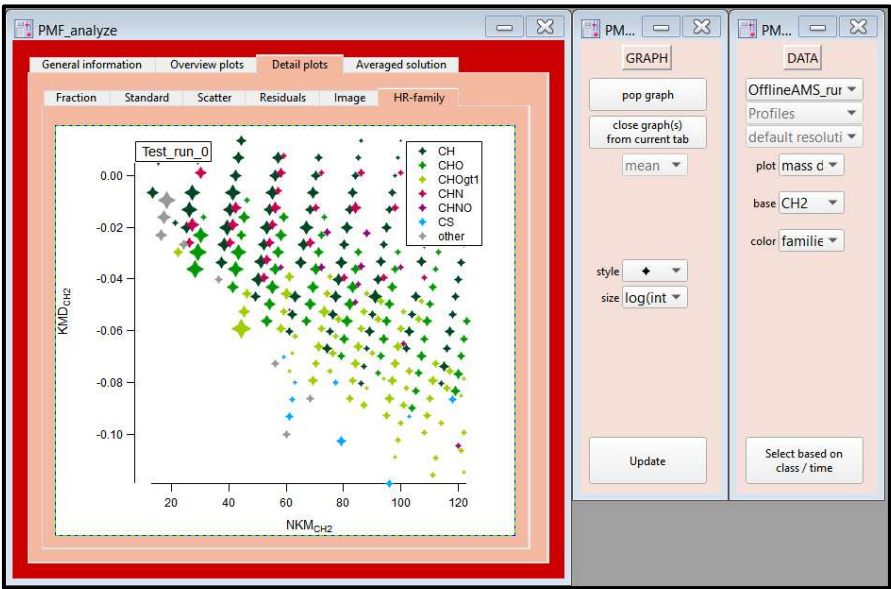


图 114

8.5.6.3. 范·克雷维伦图 Van Krevelen plot

范克雷维伦图(Van Krevelen, 1950) 常用于高分辨率质谱数据的图形解释，其中单个化合物的元素比率被绘制，而不是整体样品的比率。在图中 (图 115

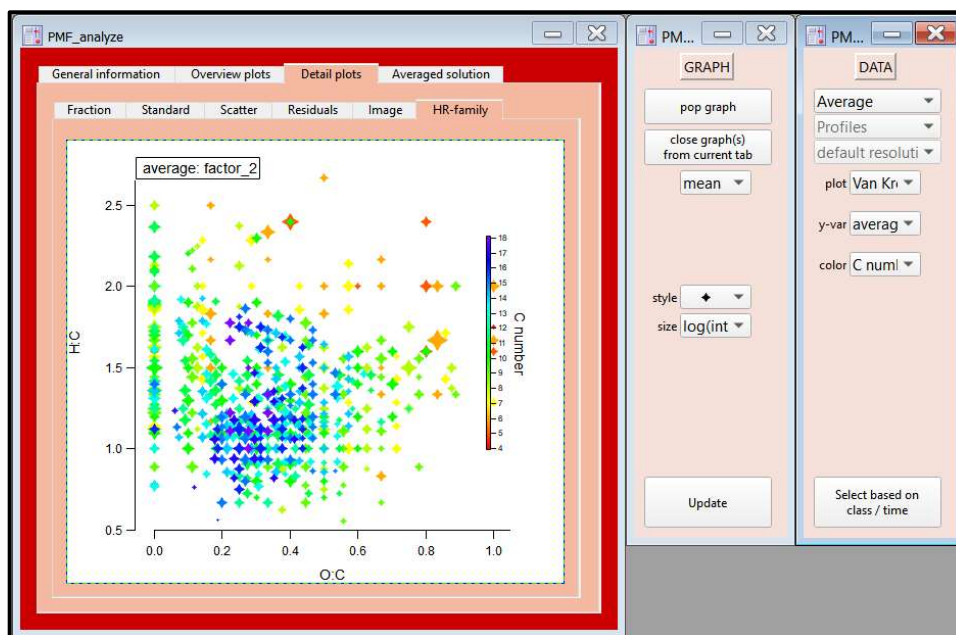


图 115

离子的强度也可以通过对其数进行缩放，并且可以选择使用最小强度（下限）minimum intensity (floor)。这是为了在数据跨越几个数量级时稳定对数缩放。

8.5.6.4. 碳原子数分布 Carbon number distribution

碳原子数分布图根据碳原子数对离子进行分箱。具有相同碳数的离子进一步分为不同的子类别，通过颜色编码显示。因此，SoFi 支持基于用户定义的 family、氧的数量、H:C 和 O:C 的分类。H:C 和 O:C 的范围可以在颜色选择下方的表格中定义（见样本在表 9）。SoFi 可以读取最多带有一个括号的化学式（例如， $(H_2O)_3H_3O$ ）。

表 9

SoFi 中的表格	SoFi 读取
< 1.1	< 1.1
1.1 - 1.3	>= 1.1 && < 1.3
1.3 - 1.5	>= 1.3 && < 1.5
1.5 - 1.7	>= 1.5 && < 1.7
> 1.7	>= 1.7

由于可能很难掌握每个子类别对总因子组成的贡献，用户可以添加摘要饼图 summary pie chart（图 116 图 116，右）。饼图仅在图形被弹出时才能正确显示。

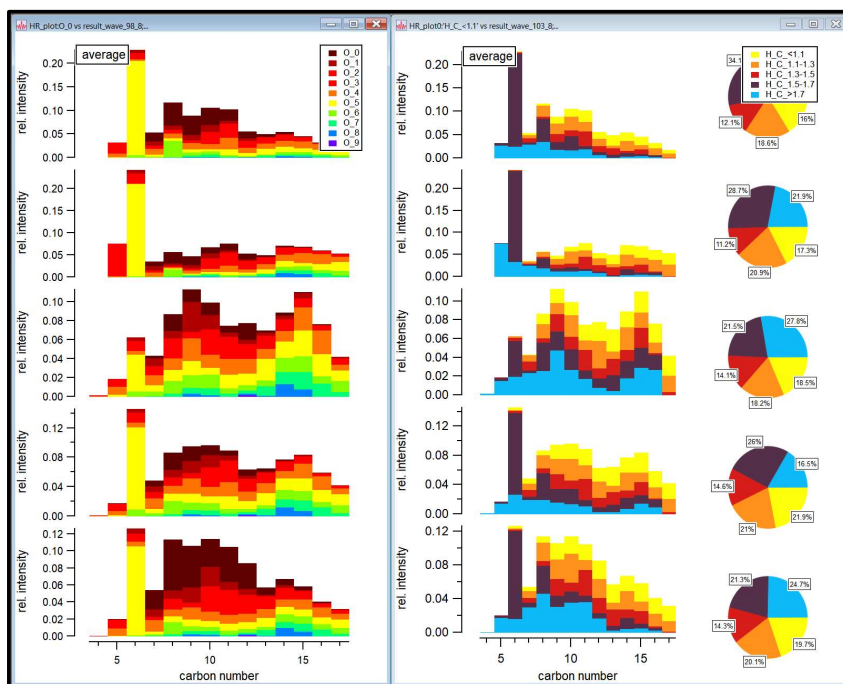


图 116

8.5.7. 类别图 Category plot

这些图提供了对解的详细检查，按预定义类别 classes（分类）组织，以便于更好的分析和比较。浓度、贡献等按类别显示。类别图以柱状图形式展示，允许对不同类别的贡献进行清晰的视觉比较。这些图对于例如来自不同站点或不同粒径的数据显示最为有用。

标准图中可用的所有时间序列可视化选项在这里也可以访问。此外，本节还包括一个额外功能，允许在单一视图中绘制整个模型解（图 117

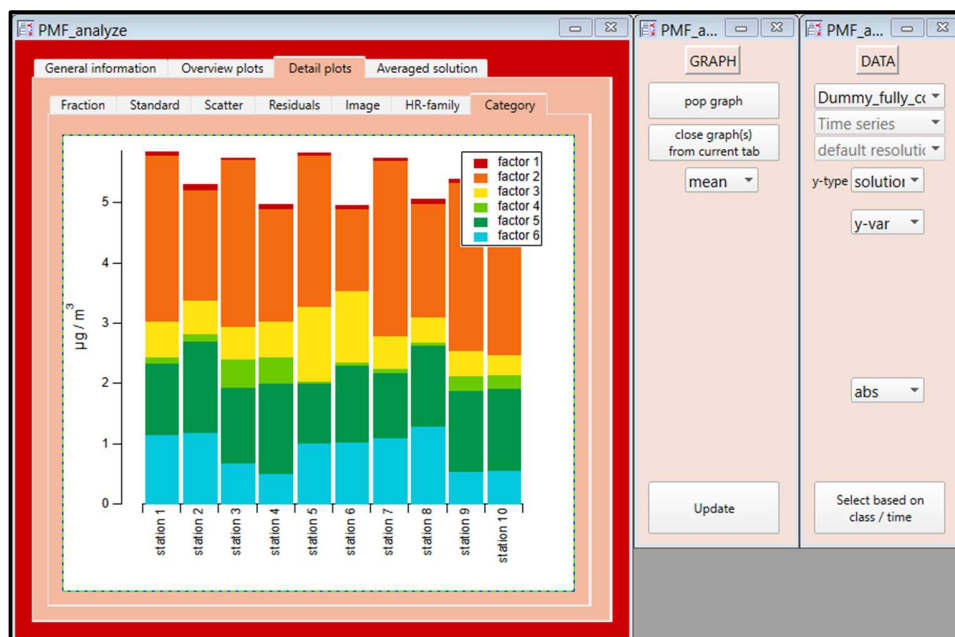


图 117

8.6. 平均解 Averaged solution

该标签页对于分析平均解特别有用，尤其是对于滚动 PMF runs。它包含关于 PMF input 数据覆盖程度的统计信息、量化 PMF 误差的方法、随时间变化的 a 值统计和随时间变化的因子源谱统计。然而，“误差估计 error estimation” 标签页在量化常规（非滚动）PMF runs 的 PMF 误差时也会很好地发挥作用。

8.6.1. 重复 Repeats

进行 bootstrap 或 rolling PMF 时，部分被建模的数据点可能没有出现在最终结果中。这种情况可能发生在：bootstrap 重复次数较少；criteria 设置了非常严格的筛选阈值。“repeats” 标签页可以显示：PMF input 数据的覆盖程度；平均 PMF 结果对原始 PMF input 数据的代表程度；不同时间点的 repeats 数量。（见图 118

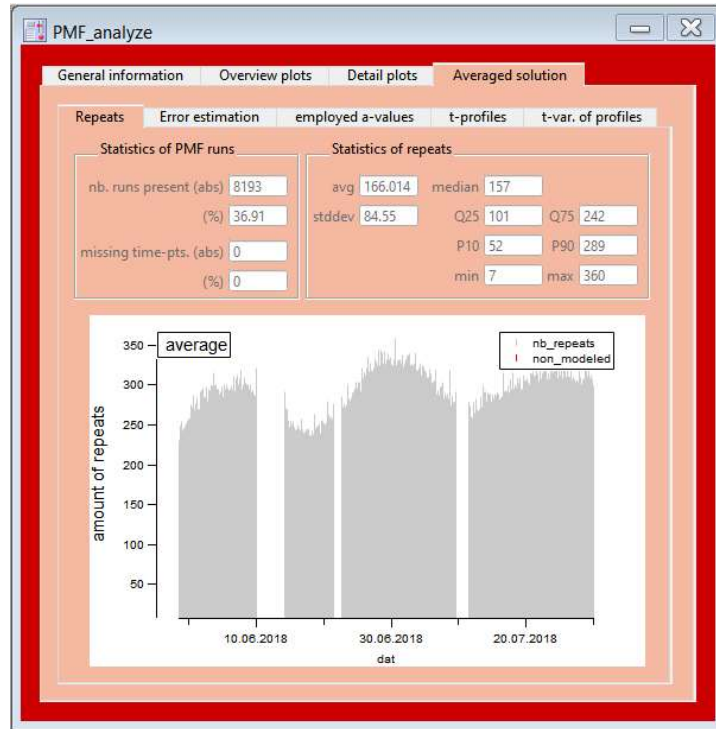


图 118

8.6.1. 质量误差估计 Mass error estimation

应用重采样策略，如 bootstrap（见 6.4.12.4

$$PMF_{error,i} = 100 \cdot \left(\frac{spread}{mean/median} \right)_i \quad (31)$$

其中的差异可以是上四分位数和下四分位数之间的差异，或者是某个时间点 i 的标准差。当将差异 $spread$ 与均值或中位数值绘制时，线性拟合的斜率代表相对误差。图

$$119 \left(\frac{q3-q1}{mean} \right)_i$$

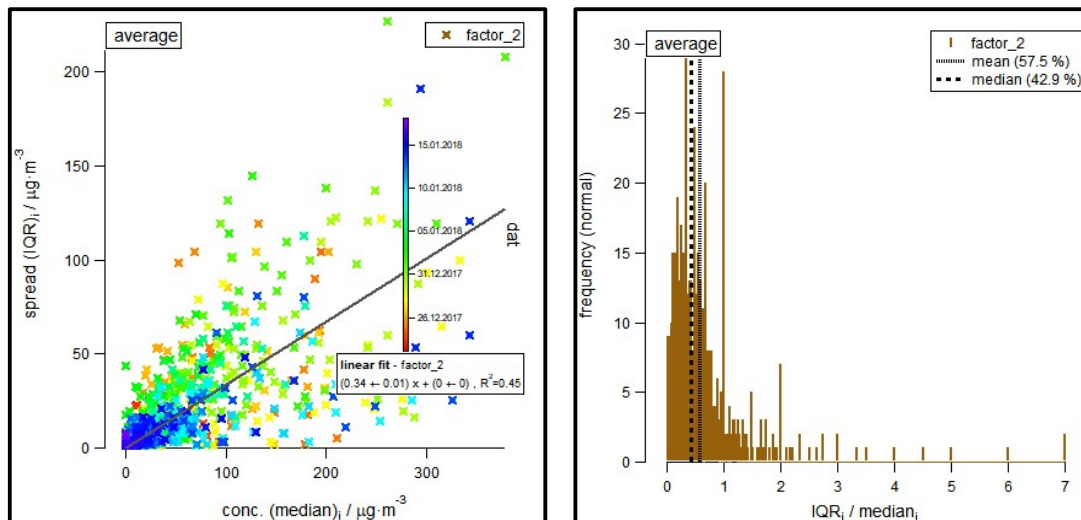


图 119: 以 IQR 表示的差异与因子 2 的中位数值 (左) 以及其对应的概率密度函数图 (右) 绘制。该因子的质量误差是斜率的一半, 以百分比表示, 即散点图上的 $\pm 17\%$ 。在频率图上, 误差更能通过整体中位数值 42.9% 来表示。因此, 误差为 $\pm 21\%$, 因为某些时间点 i 的小均值可能导致极大的值影响整体均值 57.5%。

8.6.2. 使用的 a- 值

在执行随机 a- 值 random a-value 时, 例如在滚动 PMF runs 中, 此标签页可视化了所使用的 a- 值随时间变化的统计数据, 并以时间序列和直方图形式呈现 (见图 120 图 121)

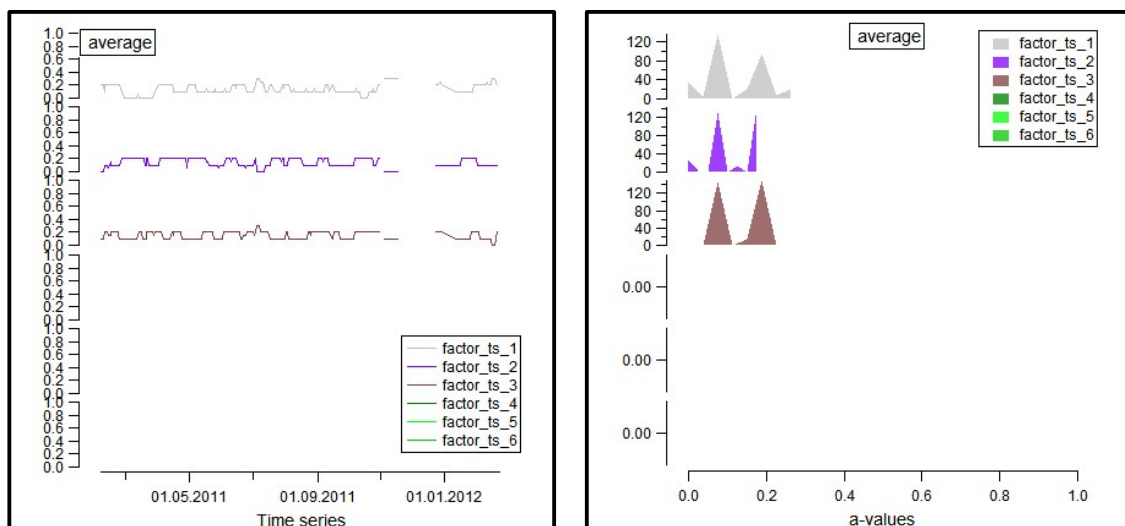


图 120: (左) 有效的 a 值在时间上的约束因子源谱和 (右) 这些 a 值的直方图表示。

除了均值、标准差、中位数和分位数统计外, 概率密度函数 (pdf) 也可以进行可视化。

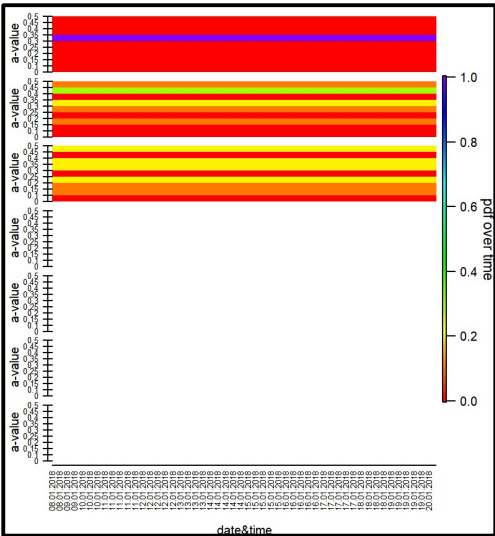


图 121: a 值随时间变化的概率密度函数。

8.6.1. t-谱 t-profiles

rolling PMF 旨在评估随时间变化的因子源谱。此信息通过将因子源谱随时间绘制为图像图来显示在此标签页中。因子源谱可以采用两种归一化方式：沿因子源谱方向归一化，使所有变量之和等于 1；沿时间方向归一化，即每个变量除以自身的平均值。此外，SoFi 还会执行双向 two-way ANOVA，用于检验所有因子源谱是否相互独立。零假设 (H0) 描述了 t-profile 矩阵中的所有因子源谱均来自同一个分布。检测结果在子面板“图形 GRAPH”中可可视化。（见图 122

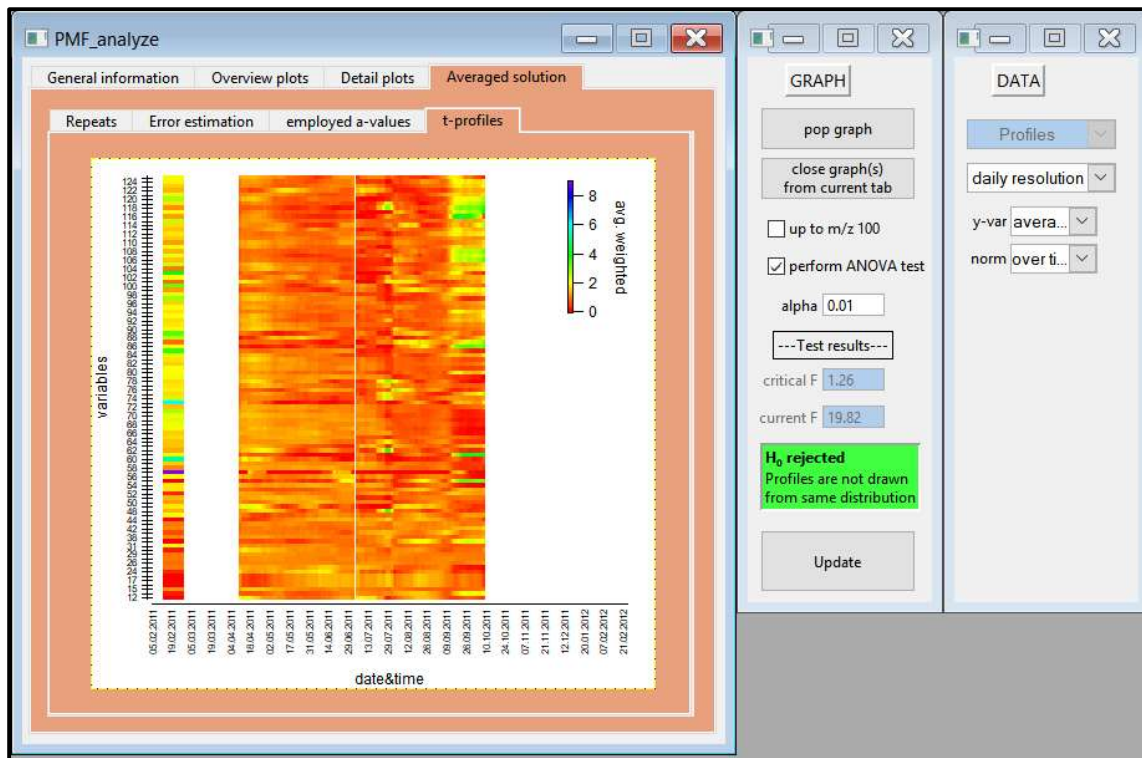


图 122

用户还可以通过绘制不同时段平均因子源谱，进一步检查因子源谱随时间的变化。在感兴趣的时间范围内拖动 **marquee** 进行框选。需要注意的是，此处只有 **x** 轴范围会影响选择，也就是说，即使 **marquee** 没有覆盖 **y** 轴上的全部变量，**SoFi** 仍会使用完整因子源谱进行平均。框选后点击鼠标右键，即可调用 **marquee** 相关选项并绘制所选时段的平均因子源谱（见图 123：图 124

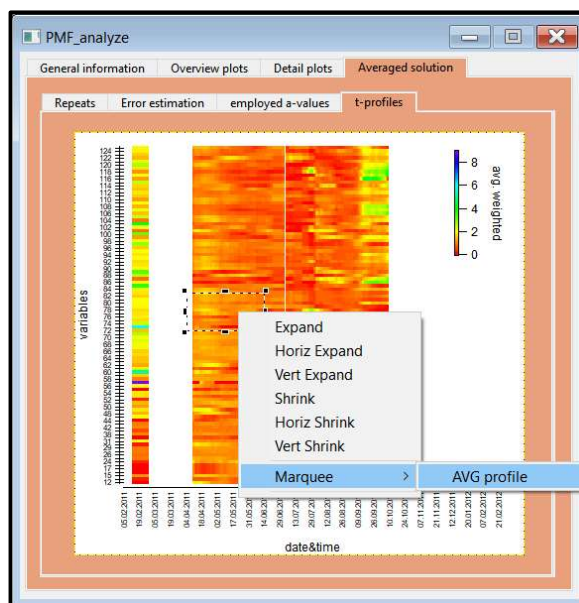


图 123: 123

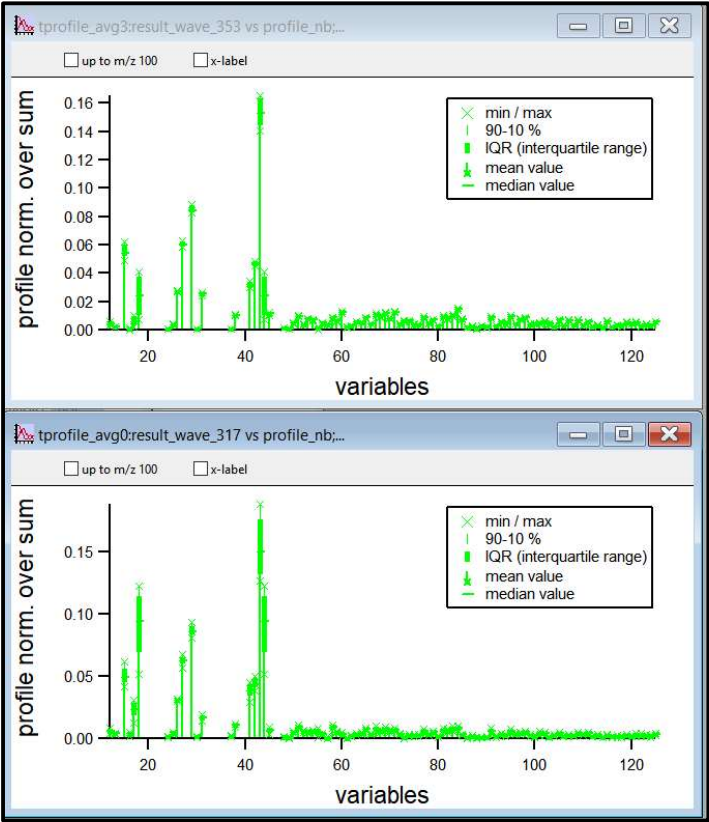


图 124

9. 独立分析

9.1. 高亮有趣的时间段

高亮 **Highlight** 面板提供了一种清晰、灵活且有效的方法，用于突出显示重要的数据时段。例如，可以用于标记：烟花燃放时段；采暖期与非采暖期；严重雾霾时段。用户通过指定一个连续的时间范围及其名称来定义需要高亮的时间段。此外，还可以调整用于高亮显示的颜色；透明度；线条大小。这样可以确保不同高亮区域在图中清晰可辨。同一个 **period** 可以由多个子时段组成。

根据需要高亮的图形类型，**Highlight** 面板可以接受不同形式的输入：普通日期，采用欧洲日期格式，即日/月/年；数值，例如日变化图或年变化图中的数值；文本，例如周变化图或月变化图中使用的相同缩写。**Highlight** 面板可以通过：“突出时期（带色背景）**highlight period(s) (tinted background)**”按钮打开。该按钮可在 **overview plot** 和 **standard plot** 中使用。

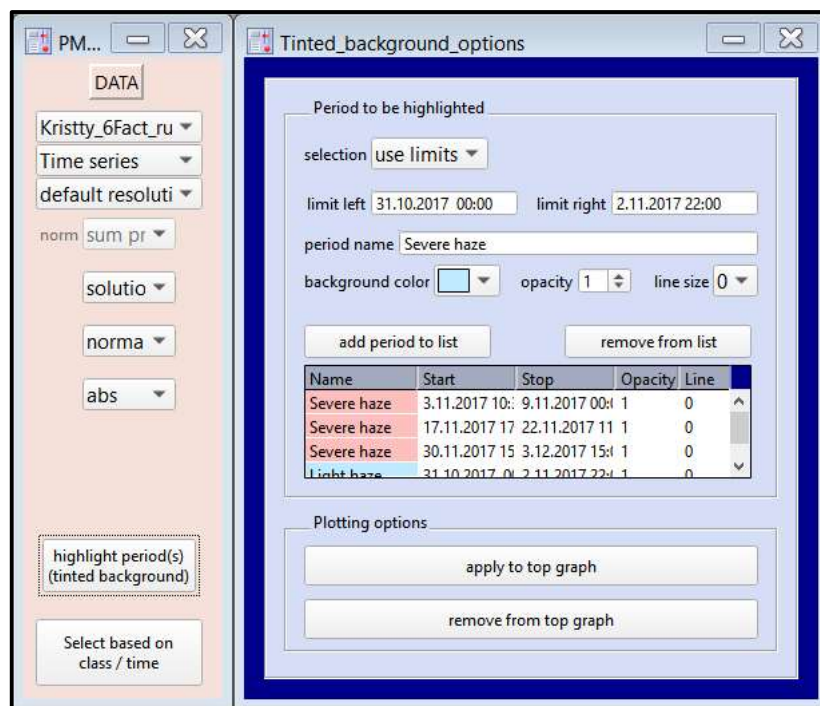


图 125

9.2. 独立 bootstrap 分析

对于探索 **explore** 和分析 **analysis** 面板，**SoFi** 提供了对（多）线性拟合参数的独立 **bootstrap** 分析。该方法可以在不对数据总体分布作特定假设的情况下，评估某些统计量的离散程度

或方差。例如，不需要假设数据服从正态分布。当 scatter plot 中启用 linear fit 后，会出现一个紫色的小按钮：“BS” 点击该按钮即可打开 bootstrap 面板。在第一个标签页中定义 bootstrap 设置。用户可以采用两种方式指定输入数据。自动设置：SoFi 自动使用 explore 或 analyze 面板中当前已经选择的：variables；externals；其他相关数据。当 explore 或 analyze 面板中的变量发生变化时，bootstrap 面板也会自动更新。手动设置：用户可以手动指定：x wave；y wave；time wave。这三个 wave 必须具有相同长度。

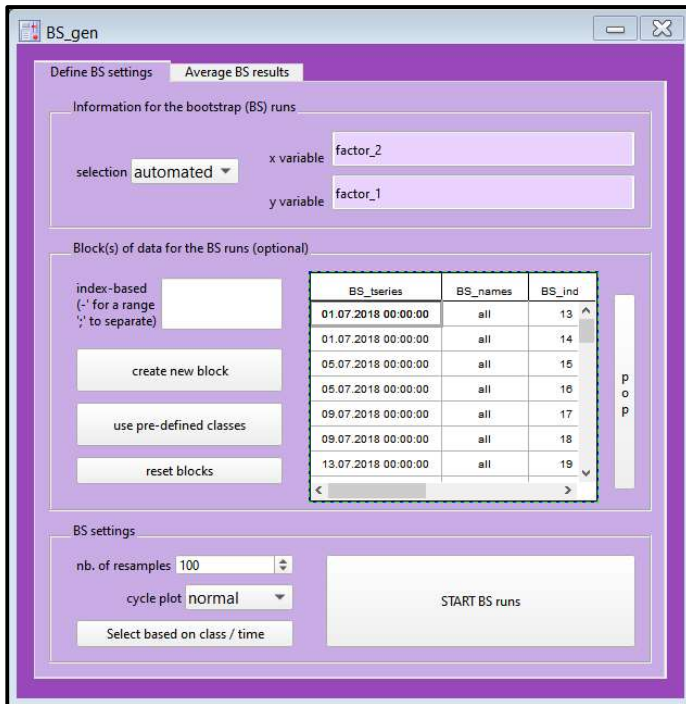


图 126: bootstrap 面板可以自动（如这里所示）或手动选择要进行 bootstrap 分析的波。此外，可以定义类别 class（可选）。

用户可以选择性地定义块 block。对于每个块，这些点会单独进行 bootstrap 分析，然后将所有块重新组合以重建输入。这样可以确保重要事件、季节性等仍然保留在数据中。SoFi 支持以下语法来创建块：

“Index based blocks” 使用 “-” 表示范围，使用 “;” 分隔单个索引或范围，然后按 “创建新块 create new block” 来创建块并命名它们

“Class supported block” 如果用户之前定义了类，SoFi 可以使用这些类来创建块。

“Monthly blocks” 使用关键字 “monthly” 按月创建块。块将自动以相应时间点的月份和年份命名

“Hours blocks” 使用关键字 “hours”，后面跟着应用于块的小时。例如，“hours17-20” 将创建一个包含下午 5 点到 8 点之间所有点的块。

“Minutes blocks”

使用关键字 “minutes”，后面跟着应用于块的分钟。例如，“minutes10-20” 将创建一个包含第 10 分钟到第 20 分钟之间所有点的块。

“Fixed length”

使用关键字 “block”，后面跟着您希望分组在一起的点数。例如，“block10” 将每 10 个点创建一个新块。请注意，此格式仅支持低时间分辨率的数据。

通过弹出 **BS_names**，块列表也可以手动调整。

在开始 **bootstrap** 分析之前，用户可以定义重抽样 **resampling** 的次数，即分析应重复多少次。目标是评估重抽样的次数，以便统计量的分布不再显示显著差异。此外，用户可以选择在哪个时间点（基于类/时间选择）“**select based on class/time**” 进行分析。此外，**bootstrap** 分析还可以在周期 **cycle** 上进行，以估计周期的不确定性（除了可以直接从红色分析面板访问的平均统计量）。

结果可以在第二个标签中查看（图 127

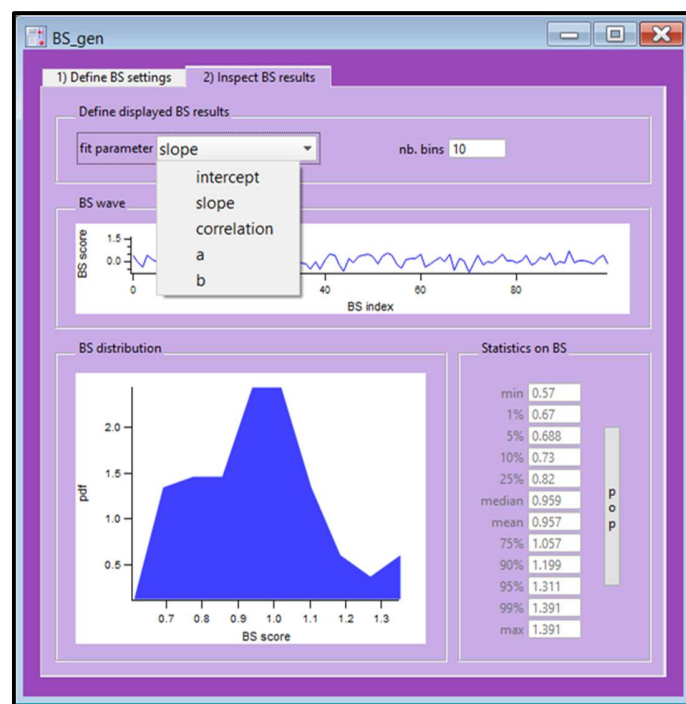


图 127: 在 **bootstrap** 法面板中检查结果。用户可以检查斜率、截距和相关性的 **bootstrap** 法结果。如果是多线性回归（如这里所示），其拟合参数也可以访问。

如果执行 **bootstrap** 重采样以估计循环不确定性，则会自动提供第三个标签。在此标签中，循环 **cycle** 图与 **bootstrap** 法数据可视化。用户可以选择标准表示 **standard representation**（类似于标准图中可用的内容）或箱形图 **box plot** 和小提琴图 **violin plot**。用户有广泛的选

项来最佳展示箱形图，包括但不限于应绘制的内容（例如，单个数据点、均值、whisker 表示的内容）以及使用的颜色（图 128

要以非图形方式访问统计数据，可以弹出一个统计表。

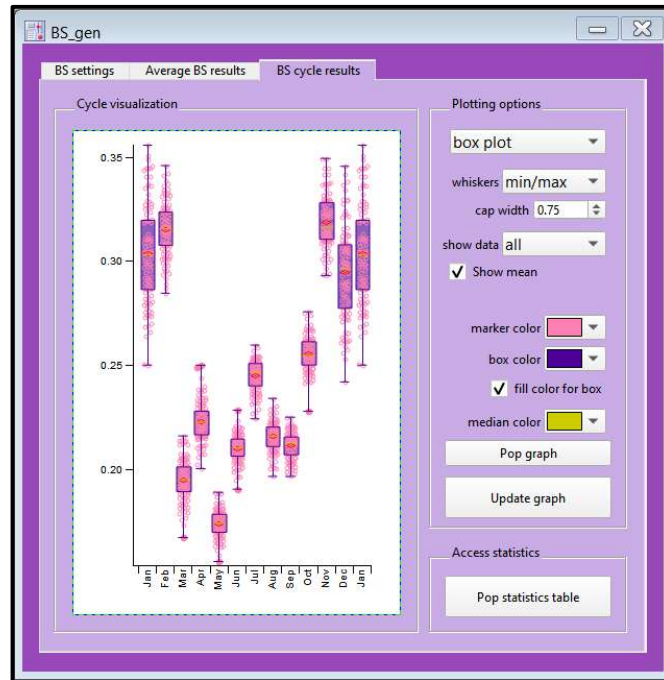


图 128: 对于周期的 bootstrap 分析，用户还有一个标签页可以检查周期图。

9.3. 选框 Marquee 工具

标记面板允许用户在图表中选择一组点，并在另一个图表中突出显示相同的点。例如，用户在因子时间序列中发现一个有趣的事件，并可以在温度、相对湿度等图表中突出显示相同的时间点，即使这些图表中并未绘制时间序列。

主要选项

选框工具可以从所有以 “var_expl”、“sol_nor” 或 “sol_scatter” 开头的图形窗口访问，即在 SoFi 探索 explore（绿色）面板中创建的所有图表（弹出或在 SoFi 面板中）或来自 SoFi 分析 analysis（红色）面板的标准和散点图。注意：当关闭红色面板并保持图表弹出时，这些图表会被重命名为 “userwin”，并且无法通过选框工具访问。可以通过拖动选框并右键在 “SoFi 选框 SoFi marquee” > “选框分析工具 Marquee analysis tool” 下选择 “选择时间段 select period” 或 “设为活动窗口 set as active window” 来访问选框面板。这将打开如图所示的选框面板。图 129

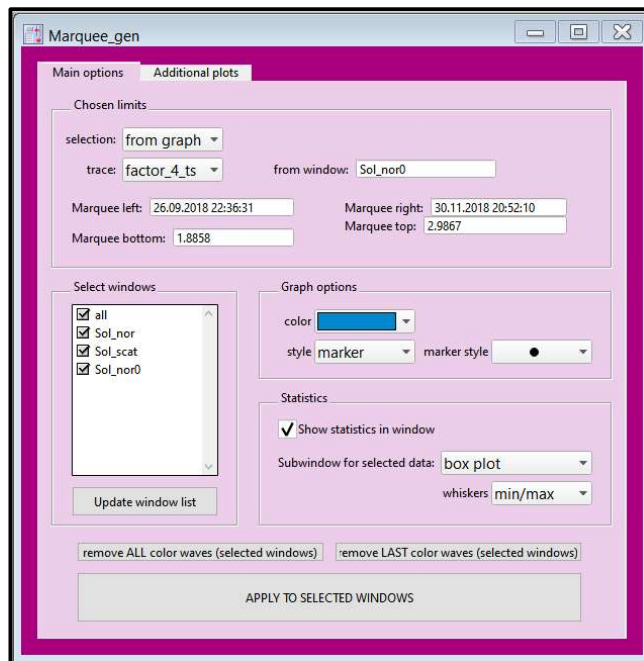


图 129

用户需要定义哪个是应作为“基准 **base**”来为其他窗口着色的感兴趣时间段。基准波的窗口显示在“从窗口 **from window**”下，可以通过点击任何“**sol_nor**”或“**sol_scet**”图表、拖动选框并选择“设为活动窗口 **set as active window**”来更改。trace 会自动更新为所选窗口中的 trace，并可以通过下拉菜单进行更改。

用户有三个选项来选择应使用的选定窗口和 trace 中的数据。

- “从图表 **From graph**”：使用执行“**set period**”时 marquee 所确定的边界。这些边界会显示在面板中，但只能通过重新拖动 marquee 修改。
- “使用类别 **use classes**”：用户可以选择在 PMF 之前创建的类别或添加自己的类别。PMF 之前定义的类别以及可供其他分析使用的类别会以列表形式显示。一次可以选择一个类别。用户可以选择添加类别。对于这些，用户必须提供一个带有索引的波形（“从波形新建类别 **new class from wave**”）或将索引写入新创建的波形（“手动添加新类别 **add new class manually**”）。请注意，这些类别仅适用于包含用于 PMF 计算的时间序列/因子源谱的图表，当使用“基于类别/时间选择 **select based on class/time**”选项进行绘图时将无法使用，因为索引将不再匹配。
- “使用限制 **uses limits**”：用户可以自行输入限制。接受日期和数字限制。

在“选择窗口 **select window**”下，可以选择数据应着色的窗口。列表显示所有包含与选定基线轨迹 **base trace** 相同长度数据的窗口。如果选择了不同的基线窗口，列表会更新。如

果在此期间生成了新窗口，列表不会自动更新，用户需要点击“更新窗口列表 **update window list**”。

用户可以通过“图表选项 **graph options**”下的下拉菜单为高亮数据选择不同的颜色和标记/线条。

可以在每个窗口中显示所选数据的一些统计信息。如果选中“在窗口中显示统计信息 **show statistics in window**”框，则在每个窗口中显示点数、均值、标准差、中位数和四分位间距 **interquartile range (IQR)**。显示的是最新添加点的统计信息。如果需要显示其他统计信息，用户需要在其中一个图表上拖动一个选框，并在选框分析菜单中选择“更新显示的统计信息 **update shown statistics**”。可以显示不同层 **layer**（顶层为层 0）的统计信息。用户可以仅更新当前窗口，也可以同时更新所有已经显示统计信息的窗口。此外，可以将包含所选数据的子窗口 **subwindow** 添加到图表中。子窗口的选项包括线性回归、箱线图和小提琴图。

当点击“应用于选定窗口 **apply to selected windows**”后，SoFi 会在每一个被选中的窗口中提取符合条件的数据点，并将它们作为新的 **wave** 添加至对应窗口。如果需要移除高亮数据：“**remove ALL color waves (selected windows)**”删除所选窗口中的全部高亮 **wave**。“**remove LAST color waves (selected windows)**”仅删除所选窗口中最后一次添加的高亮 **wave**。

附加图表

用户可以选择附加图表。在主菜单中的图表中，子窗口中显示的数据始终是所选的数据。在这里，用户还可以选择绘制高亮数据的周期、散点图和饼图。为了使这些图表可用，必须至少应用 **marquee** 选择一个 **IGOR wave**。

对于所有选项，可以更新子窗口或在第一个子窗口的位置旁边绘制一个新的子窗口。子窗口可以在工具模式下移动和调整大小 (**ctrl+t**)。然而，图表只能在工具模式 **tool mode** 被停用时代添加（再次按 **ctrl+t**）。对于周期图和散点图，还有将图表弹出到单独窗口的选项。此外，图表顶部的控制栏 **control bar** 始终会更新为最新添加的子窗口数据。

- **周期图 Cycle plots**。如果主图包含时间轴（因此不适用于散点图），则可用。从下拉菜单中，用户可以选择感兴趣的图表。下一个下拉菜单显示所有可能的标记波（用户高亮显示的不同周期）。用户可以选择从每日到每年的周期，以及周期的显示方式（均值、标准差、中位数、四分位距 **IQR**、10 百分位 **percentile/90** 百分位和最小/最大）。

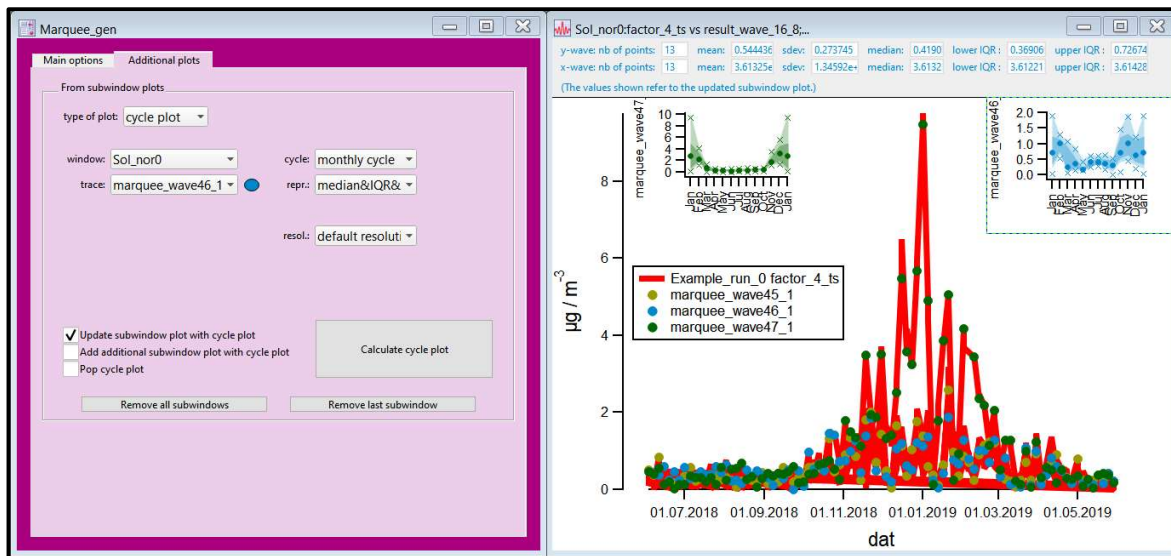


图 130

- **散点图 Scatter plots.** 将您突出显示的数据与任何其他数据进行绘图。用户再次需要定义窗口并选择感兴趣的 IGOR wave。如果主图是散点图，用户还必须进一步指定在子窗口中应使用 x 轴还是 y 轴进行散点图绘制。对于散点图，可以选择任何时间序列：首先指定类型（例如，外部 external），然后指定变量（例如，氯化物 chloride）。

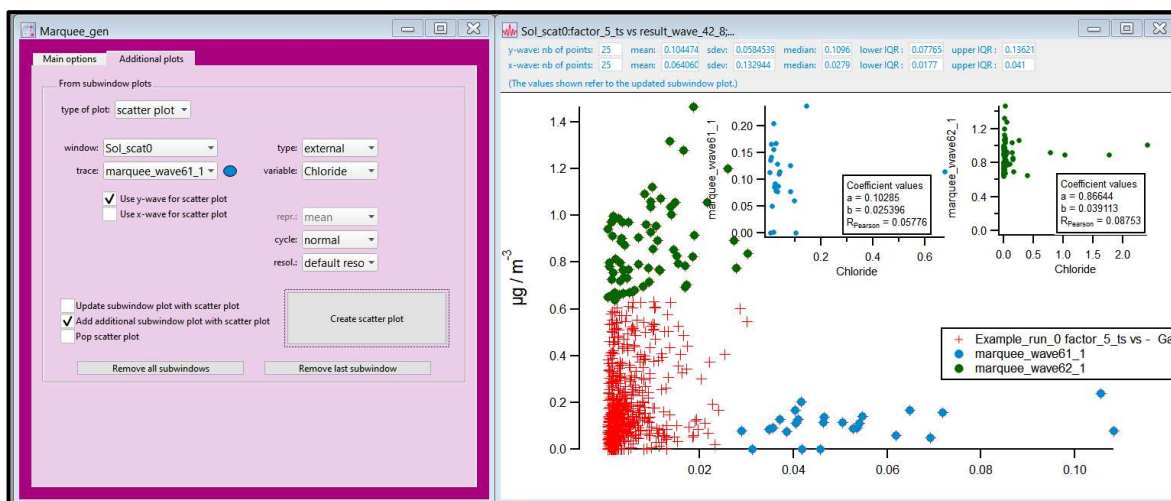


图 131

- **饼图 Pie chart.** 突出显示期间因子贡献的饼图（有或没有未解释的变异）。用户可以选择因子的源贡献和包括未解释变化程度的因子贡献。此外，用户可以控制标签选项。

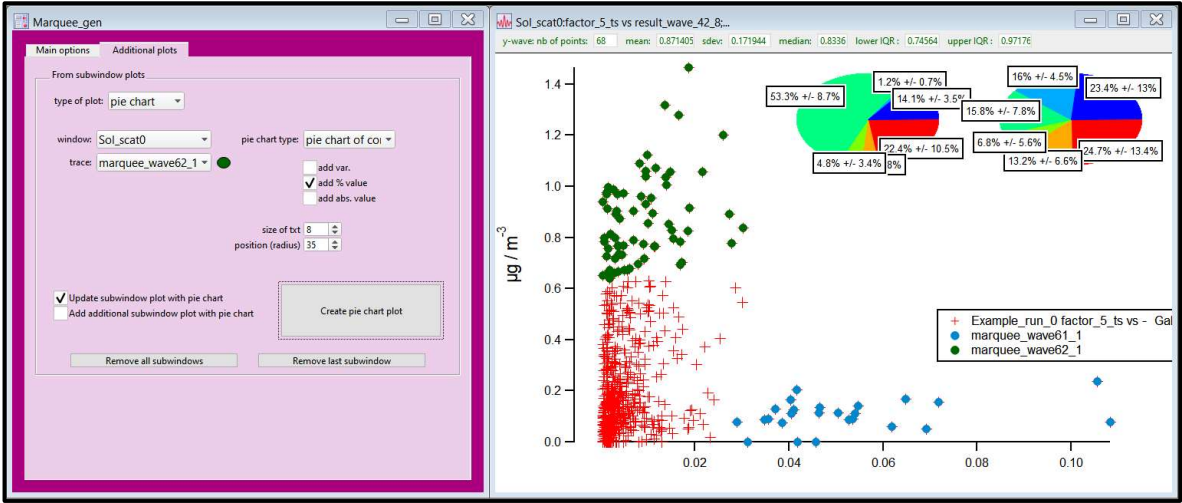


图 132 两个选中时段对应的因子贡献饼图。

9.4. 聚类面板 Cluster panel

聚类面板允许通过 K 均值 K-Means clustering 或最远点 (FP) 聚类 farthest point (FP) clustering 来聚类数据。有关这两种计算的聚类算法的更多详细信息，请参见 Igor 帮助菜单。

9.4.1. 聚类选项 Cluster options

在 SoFi 中，可以聚类不同的变量（包括结果、外部变量、解释性变化等），PMF 输入矩阵，以及如果有多个 PMF runs，可以跨 PMF runs 的变量。在跨 PMF runs 的变量的情况下，还有一个选项可以将聚类作为标准面板中的选择的一部分。在第一步中，必须定义输入数据：

- **custom data:** 变量需要由用户定义。时间序列和源谱文件不能混合，除此之外，组合数据的数量没有限制。单个变量可以从下拉菜单中选择，类似于可以在红色分析面板中选择变量的菜单，或者可以使用与蓝色标准面板 criteria panel 中相同的样式输入。选择/输入变量名称后，按“写入 write”，变量将出现在右侧的列表中。
- **PMF input matrix:** 不需要进一步定义。黑名单输入矩阵将被聚类。
- **Variable across runs:** 仅在所选的 hdf5 文件中有多个 PMF runs 可用时才可用。在完全访问此功能之前，用户首先需要定义至少一个图像条件并进行检查。例如，要对所有 PMF runs 中一个因子的所有因子源谱进行聚类，用户首先必须为该因子定义图像条件（源谱、解、因子 x），评估它并进行检查，否则会弹出提醒。下拉菜单中列出了用户可以选择的所有图像条件。此外，如果所选的图像条件基于源谱，则可以通过勾选“缩放因子源谱 **scale factor profiles**”的复选框来缩放源谱。如果勾选，则每个变量将乘以缩放因子 $(m/z)_{sf}$ ，以逐渐增加质谱末端模式的权重。更多细节可以在 (Äijälä et al., 2017) 中找到。由于聚类针对 NaN 时是不稳定的，因此此选项仅适用于源谱和周期图，因为 bootstrap 法正常时间序列包含太多 NaNs。

对于自定义数据和 PMF 输入矩阵选项，基于类别/时间的选择 **select based on class/time** 选项可用，可以选择应考虑用于聚类的不同类别或时间点。在不同 runs 间变量的情况下，此选项在聚类面板中不可用，而应在条件定义期间完成时间或聚类的筛选。

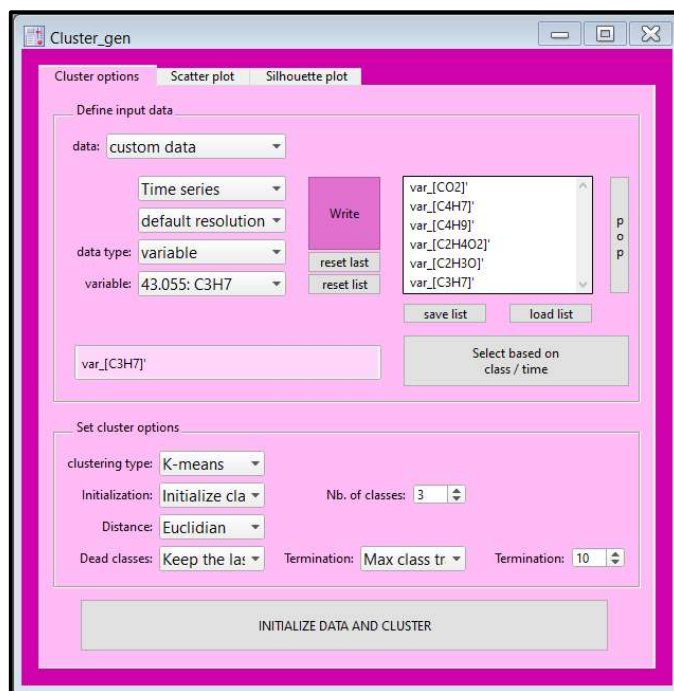


图 133

接下来，需要设置聚类选项。SoFi 提供两种不同的聚类选项：**K-Means** 聚类和最远点 (FP) 聚类。由于 Igor 的限制，在跨 PMF runs 之间聚类变量时，仅提供 FP 聚类选项。

FP 聚类 FP clustering:

- **-Termination:** 确定迭代何时停止
 - **Numbers of clusters** (默认): 当聚类数量达到指定值时终止计算
 - **Intercluster radius:** 当最大距离小于或等于定义的距离时终止计算
- **Number of classes/Intercluster radius:** 指定终止数量。该数字的含义由终止方法决定。如果选择了聚类数量，当聚类数量达到指定值时计算将终止。请注意，这个终止条件是充分但不必要的，即，即使 cluster 数量尚未达到指定值，如果某个元素到其 hub 的最远距离已经小于平均距离，计算也可能提前结束。

K-Means 聚类:

- **Initialization:** 指定初始化方法
 - 从一个类别 class 中随机选择数据点的数值
 - 从总体中随机选择数据点进行初始化 (默认)
- **Distance:**
 - 欧几里得 Euclidian: 距离被评估为欧几里得距离 (默认)

- 曼哈顿 **Manhattan**: 距离被评估为绝对值的总和（也称为曼哈顿距离）。
- **Dead classes**: 指定算法应如何处理“死”类，即在给定迭代中失去所有成员的类
 - 如果聚类失去所有成员，则移除该类
 - 保留该 **cluster** 上一次的随机均值向量 **mean vector**，以便后续迭代中可能重新获得成员（默认）
 - 为该类分配一个随机均值向量
- **Termination**: 确定迭代何时停止
 - 用户指定的迭代次数
 - 持续迭代，直到在一次迭代中改变类的元素不超过固定数量（默认）
- **Termination number**: 指定终止数。该数字的含义由终止方法决定
- **Number of classes**: 设置数据中的聚类数量

要对数据进行聚类，请点击“初始化数据和聚类 **Initialize data and cluster**”按钮。注意：聚类对 **NaN** 是不稳定的。因此，如果输入矩阵中有任何 **NaN**，整行将被删除，即在聚类自定义数据时，如果选择的某些外部时间序列在某些时间点有 **NaN**，则这些时间点将完全不被考虑。

估计聚类数量 **Estimate number of clusters**

此选项应有助于根据两种不同的方法确定最佳聚类数量：肘部法和源谱法。

肘部法 **elbow method**

SoFi 对选择的聚类算法进行 1 到 10 个聚类的 **runs**。因此，绘制了每个聚类数量的平方距离和 (SSD)。结果图中的“肘部”表示 SSD 下降速率减缓的点，暗示最佳聚类数量。

轮廓法 **Silhouette method**

SoFi 对选择的聚类算法进行 1 到 10 个聚类的 **runs**，并计算每个聚类数量的轮廓分数 **silhouette score**（可能会耗费相当多的时间）。最佳聚类数量通常被识别为最大化所有数据点的平均轮廓分数的数量。

警告：此估算可能需要一些时间，具体取决于需要聚类的数据量。

9.4.2. 标准图 **Standard plot**

在跨 **PMF runs** 聚类变量时，标准图是一个选项，可以探索例如在各个 **PMF runs** 中源谱的样子以及聚类中心 **cluster center** 的位置。输入数据以交叉点显示，每个变量的聚类中心以粗体符号表示。还可以根据各 **PMF** 运行所属 **cluster** 对输入数据进行颜色编码。

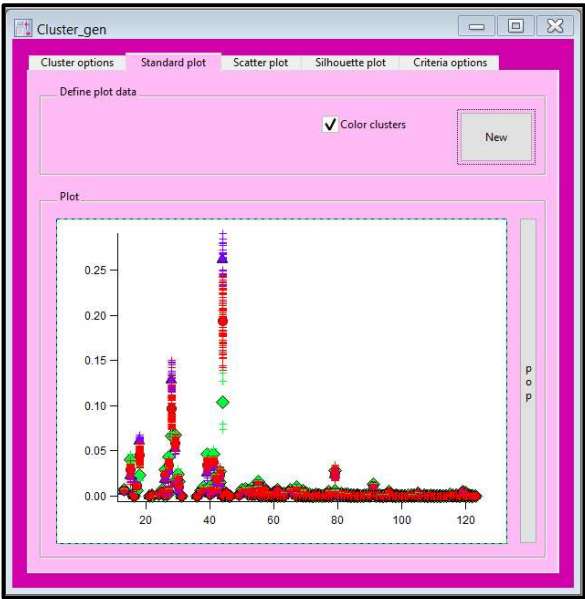


图 134

9.4.3. 散点图 Scatter plot

在此标签中，用户可以通过散点图了解不同的聚类。从下拉菜单中，可以分别选择在 x 轴和 y 轴上显示的变量。点击“新建 **New**”按钮将创建/更新图表。输入的数据点以 + 显示，聚类中心以 o 显示。为了使聚类更加明显，可以勾选“颜色聚类 **color clusters**”框，以便点按聚类着色。

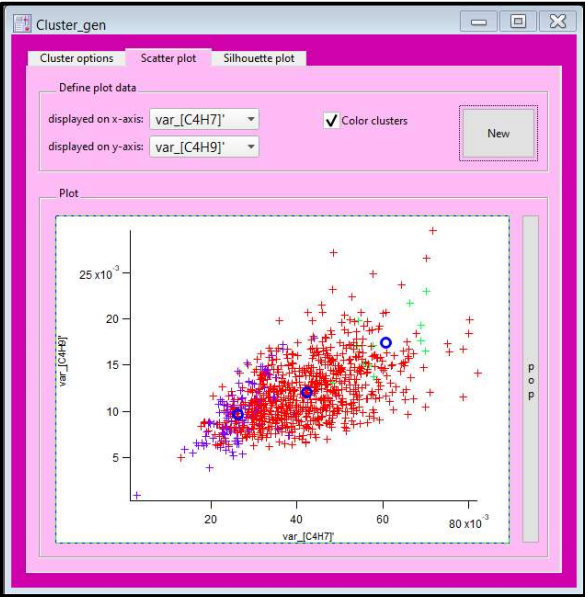


图 135

9.4.4. 轮廓图 Silhouette plot

Silhouette plot 用于评价每个数据点与自身 cluster 以及相邻 cluster 的接近程度。每个点的轮廓值定义为

$$s_i = f(x) = \begin{cases} 1 - \frac{a_i}{b_i}, & a_i < b_i \\ 0, & a_i = b_i \\ \frac{b_i}{a_i} - 1, & a_i > b_i \end{cases} \quad (32)$$

其中 a_i 是与同一聚类 A 中所有其他对象的平均距离， b_i 是与最近邻聚类 B 中所有其他对象的平均距离。接近 1 的轮廓值意味着聚类内的距离远小于聚类间的距离，因此该点可能被正确分组，而接近 -1 的轮廓值则意味着该点在平均上更类似于邻近聚类的点，可能被错误分类。

轮廓图通过点击“新建 **new**”创建/更新。由于必须计算每个点与同一簇内以及所有其他簇内所有邻居的距离，因此对于大型数据集，这可能会耗时。计算完成后，图形和平均轮廓值会自动更新。每个组用不同的颜色表示，这些颜色与标准图和散点图中显示的颜色/组相对应。有关更多信息，即每个簇的平均轮廓值和点数，用户可以从下拉菜单中选择一个簇，所选簇的值将被更新。

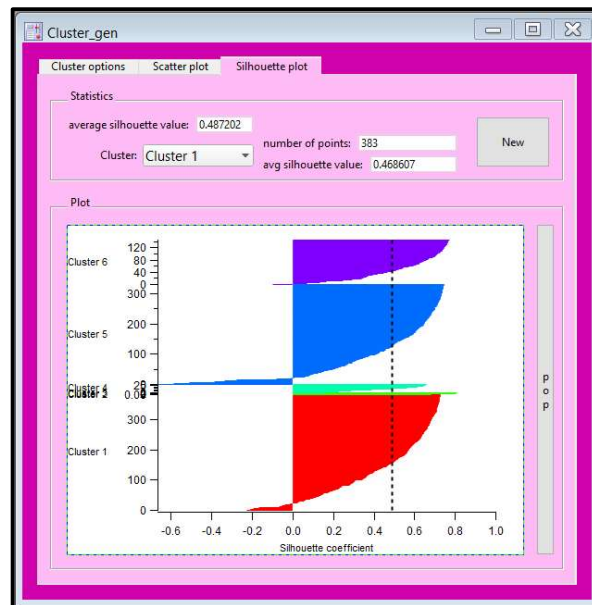


图 136

9.4.5. 标准选项 Criteria option

如果聚类对象是不同 PMF 运行之间的变量，即来自 criteria 方法中某个 image criterion 的数据，则可以根据聚类结果进一步选择或取消选择 PMF 运行。SoFi 会显示所有 cluster 的列表。。需要检查（绘制）轮廓图以激活簇列表。用户可以选择哪些簇被视为“差 **bad**”的解并应在标准面板中排除，即不被选择。通过按“排除 **exclude**”，所选簇将在标准面板中被取消选择（图 137： 137

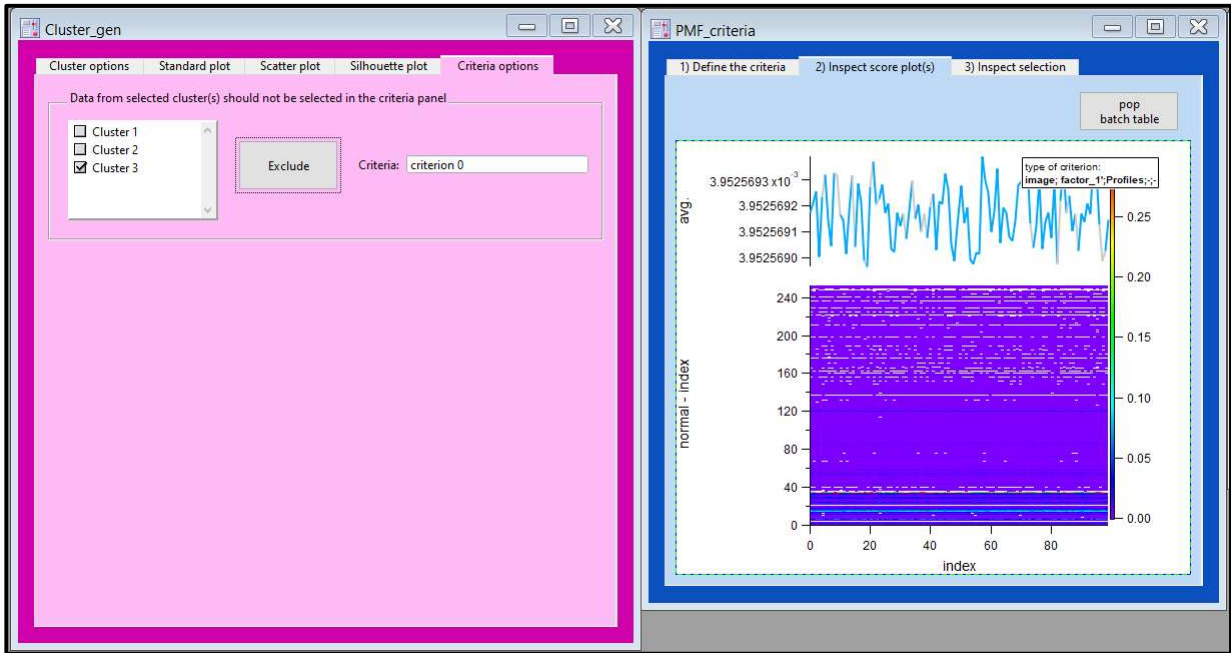


图 137： 137

10. SoFi 中的实时模块（SoFi RT）

10.1. 关于

10.1.1. SoFi RT 是什么

本节描述了如何操作 SoFi 实时模块，简称为 SoFi RT。SoFi RT 执行自动化和实时源分配分析。SoFi RT 可用于(Q 或 ToF)-ACSM 和 Xact 的单独和组合源分配运行。此外，存在一个用于从 Aethalometer 仪器导入 BC 数据的自动导入通道，在此进行源解析。(Sandradewi et al., 2008) 被自动应用，用户将能够访问分配到生物质燃烧源(BC_{bb}) 和交通源(BC_{tr}) 的黑碳数据。SoFi RT 仅在 SoFi 9 中工作，并且不需要在采集数据的 PC / 笔记本电脑上。用户负责从云服务提供商（例如，dropbox、onedrive、icloud 等）设置云账户，以确保采集数据的 PC / 笔记本电脑与进行实时源解析的 PC / 笔记本电脑之间的数据文件夹同步（见图 138）

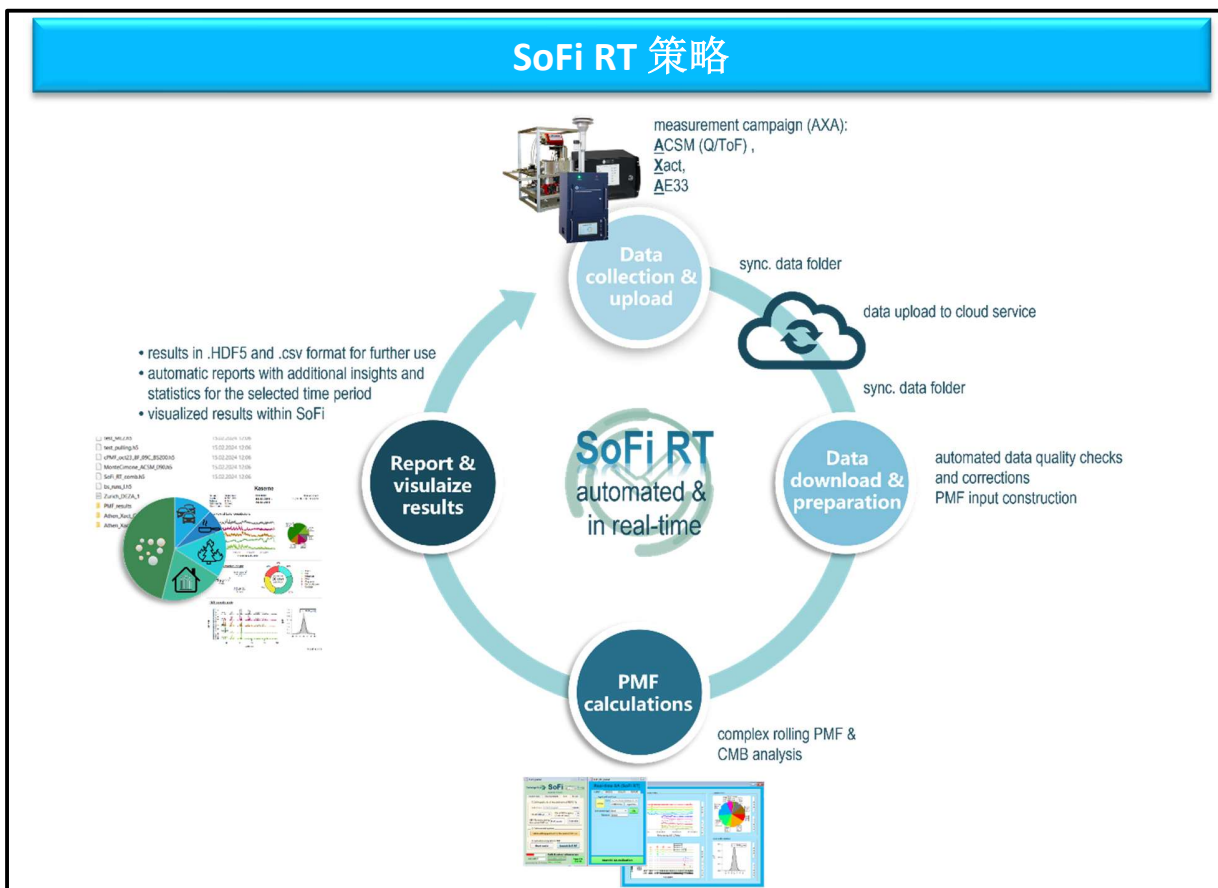


图 138: SoFi RT 数据采集的工作原理总体工作流程示意图。

10.1.2. 工作原理

由 ACSM、Xact 或 Aethalometer 收集的数据会被仪器制造商自动处理（校正和标定），并为 PMF 做好准备（PMF input 文件）。对于 ACSM 和 Xact 数据（这两种都是商业仪器），这一步目前已经可以自动实时生成 PMF input 文件。当这些 PMF 输入文件存在于一个同步文件夹中时，相同的数据也可用于 SoFi RT，以及源解析（见章节 2 可以实时进行。因此，SoFi RT 执行完全自动化的 rolling PMF，用户可以完全控制多种校正和模型设置，后面将详细描述。SoFi RT 执行两种类型的自动源解析，即一种更复杂的（rolling PMF），通常进行多个 PMF runs 以评估基于统计和旋转不确定性的 PMF 误差，以及一种更简单的使用化学质量平衡（CMB）以实时获取最新实时数据的源解析结果。rolling PMF 在一个更小的、用户可扩展的输入窗口上进行，长度例如为 7-14 天，并在测量进行时例如每天移动。在此期间，CMB 在每次新的测量扫描进入同步文件夹时执行源分配的更新。这使得用户/监管部门能够实时了解各种源的浓度变化（见图 139 Chen et al（2022））。

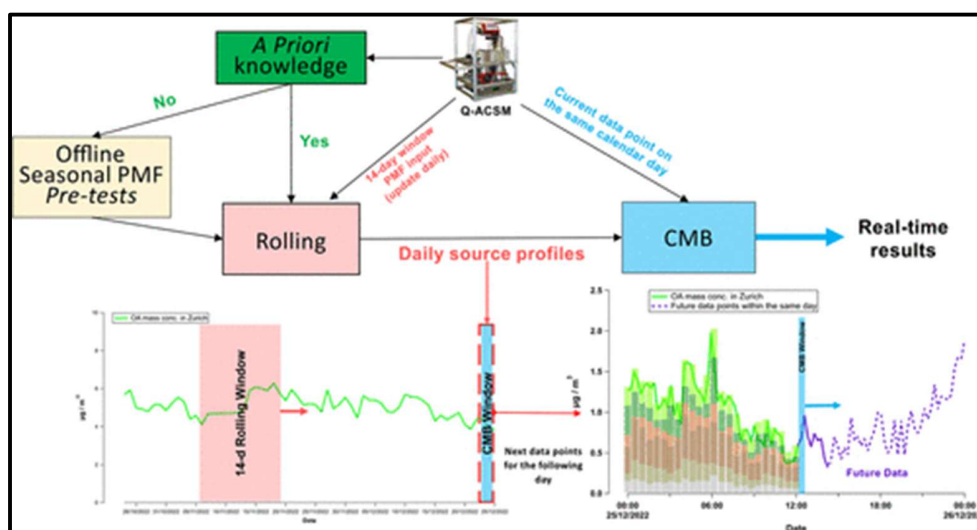


图 139

为了使 SoFi RT 正常运行，这涉及到了 SoFi Pro 中的全部高级功能。因此，SoFi RT license 同时包含 SoFi Pro license。SoFi RT 会在专门的独立面板中显示实时源解析结果。不过，无论是 rolling PMF 结果还是 CMB 结果，都可以随时使用 SoFi Pro 的全部分析功能进行详细检查。因此，SoFi RT 虽然高度自动化，但模型设置和最终结果仍然完全由用户或建模人员掌控。接下来的几行描述了成功启动 SoFi RT 并使用 SoFi RT 进行半自动源归属所需的所有技术步骤。

10.2. 配置文件 Configuration file

对于实时源归属，可以使用一个源谱文件来存储设置 SoFi RT 所需的所有信息。配置文件以 RT_config.txt 文件的形式提供，可以从我们的存储库网站下载。

以下是使用配置文件运行 SoFi RT 的简要步骤:

- 1) 在新的 IGOR 实验中启动 SoFi。
- 2) 转到第三个标签“运行 Run”，点击蓝色按钮“调用 Call SoFi RT”以启动 RT 模块。
- 3) 点击按钮“配置文件 CONFIG file”以选择配置文件。
- 4) 名称可以更改，因为系统会提示用户在硬盘上选择配置文件。
- 5) 点击按钮“开始 RT-SA 评估 Start RT-SA evaluation”

请勿从配置文件中删除任何行，而是将其在 : 后留空。

此过程避免了从 10.3 10.5 10.5

它遵循配置文件的所有行:

[0-4] 文件信息 **File Information**

- | | | |
|------|---|---|
| [5] | 路径 PATHS | 空 |
| [6] | 可执行文件路径
path of executable | 例如: “C:\ME2 engine” (见 10.3.2) |
| [7] | PMF input 文件路径
path of input files | 例如: “D:\OneDrive\Desktop\Input\ACSM”
当有多个不同的 PMF input 时，需要用分号隔开
(更多关于 SoFi RT 第一个 panel 的信息见 10.4.1) |
| [8] | 逐 CMB 结果扫描路径
path for scan-wise results (CMB) | 例如: “D:\OneDrive\Desktop\Input\CMB_results”
主 SoFi 面板第三个标签的信息 (见 10.3.2) |
| [9] | 因子源谱文件的路径
path of factor profiles from file | 例如: “D:\OneDrive\Desktop\profiles.itx”
来自文件如果这一行为空，SoFi 将在我们的服务器数据库中搜索要约束的因子配置文件的名称 (见下方[23] 行) |
| [10] | BC input 文件的路径
path of BC input file | 例如: “D:\OneDrive\Desktop\Input\AE33”
此文件夹必须包含通过我们的批处理文件生成的每日 .csv 文件 (见章节 10.8) |
| [11] | 其他外部路径数据
path of other external data | 待定 |
| [12] | 空 | 空 |
| [13] | 名称 NAMES | 空 |
| [14] | HDF 结果文件名
(大文件) | “test_config”
HDF 文件的名称 (见 10.3.2) |

name of HDF result files (large files)		
[15]	文本 txt 结果文件的名称 (单次扫描)	“CMB_result” 包含每次扫描结果的文本文件前缀 (见 10.4.3)
name of txt result file (single scans)		
[16]	input 文件前缀 prefix 名称	例如 “RTdata” 如果是组合的, 必须用分号分隔
name of prefix for input filename		
[17]	BC input 文件的前缀名称	通常为空白, 但可以在这里更改
Name of prefix for BC input filename		
[18]	其他外部数据前缀名称	待定
name of prefix for other external data		
[19]	空	空
[20]	数据类型 DATA TYPE	空
[21]	数据类型 data type	“AMS UMR / Q-ACSM (不包括 m/z 19 和 20)” 主 SoFi 面板第一个标签上的信息 (见 10.3.1)
[22]	AMS 特定 AMS specific	“排除 CO2 相关变量 exclude CO2 related variables” 主 SoFi 面板第一个标签上的信息 (见 10.3.1)
[23]	仪器类型 Instrument type	“ACSM” 如果是组合的, 必须用分号分隔 SoFi RT 面板第一个标签上的信息 (见 10.4.1) 包含输入数据的文件名前缀 (见 10.4.1)
[24]	空	空
[25]	模型选项 MODEL OPTIONS	空
[26]	重启 RT restart RT	“0.1” SoFi RT 面板第二个标签的信息 (见 10.4.2)
[27]	初始因子数 factor start	“4” SoFi 面板的主要设置信息 (见 10.3.2)
[28]	最终因子数 factor end	“4” SoFi 面板的主要设置信息 (见 10.3.2)
[29]	因子名称 factor name	“HOA,COA,OOA1,OOA2” SoFi 面板的主要设置信息 (见 10.3.2)
[30]	迭代次数 nb. of iterations	“2” SoFi 面板的主要设置信息 (见 10.3.2)
[31]	重采样 resampling	“启用 enabled” SoFi 面板的主要设置信息 (见 10.3.2)
[32]	滚动	“启用 enabled”

	rolling	SoFi 面板的主要设置信息（见 10.3.2
[33]	窗口长度	“5”
	window length	SoFi 面板的主要设置信息（见 10.3.2
[34]	窗口移动	“1”
	window shift	SoFi 面板的主要设置信息（见 10.3.2
[35]	在 PMF 之间	“每次扫描后的 CMB CMB after each scan ”
	between PMF	SoFi RT 面板第二个标签的信息（见 10.4.2
[36]	统计测试（相关性）	“禁用 disabled ”
	stats tests (corr)	SoFi RT 面板第二个标签的信息（见 10.4.2
[37]	统计测试（比较）	“禁用 disabled ”
	stats tests (comp)	SoFi RT 面板第二个标签的信息（见 10.4.2
[38]	a 值约束	“启用 enabled ”
	a value constraints	SoFi 面板的主要设置信息（见 10.3.2
[39]	约束名称	“HOA,Paris,0.1;BBOA,avg,0.3”
	name of constraints over pr	这些是存储在我们仓库网站上的指纹名称。

注意结构：

[名称], [站点的附加信息], a 值上限; [名称], [站点的附加信息], 等。

如果提供了站点的附加信息，则 SoFi 会从我们的数据库中搜索该参考谱文件。否则，如果提供了源谱文件路径，则 SoFi 将在该路径中搜索。

[40]	时间序列约束的名称	“ ”
	name of constraints over ts	与[23]相同，但针对时间序列约束。

请注意，这些位置是硬编码的，因此请勿删除、交换或对这些行进行任何其他操作。此外，结构：

“[txt]: [空格] [txt]”是硬编码的。

除每行末尾的 txt 信息（[txt]）外，请勿更改任何内容或删除任何空格。

10.3. 手动操作的第一步

10.3.1. 启动 SoFi 并准备输入数据

- 1) 在新的 IGOR 实验中启动 SoFi。

目前，SoFi RT 需要一个与用于 PMF 的变量大小相同的虚拟输入文件 **dummy input file**。该文件直接为 Q-ACSM 数据生成，存储在同步的 PMF 输出文件夹中，文件名为：“RTdata_template”。在不久的将来，ACSM/AMS 和 Xact 源谱文件的数据库将会在 Datalystica 服务器上提供，此步骤将实现完全自动化。对于 Xact 数据，目前请直接联系我们以获取一个合理的模板，您可以进一步修改。或者如果您习惯使用 IGOR，可以自行生成四个 IGOR waves，即数据矩阵、误差矩阵、时间序列和变量的数值 IGOR waves。确保数据矩阵和误差矩阵的列维度与数值变量的 IGOR waves 匹配。数据矩阵和误差矩阵必须仅包含一行，该行在数据矩阵、误差矩阵以及时间序列中对应的条目均为 NaN。最后，将此文件保存为 **igor txt** 文件，然后通过 PMF 导入面板导入回 SoFi（图 1404.1.1

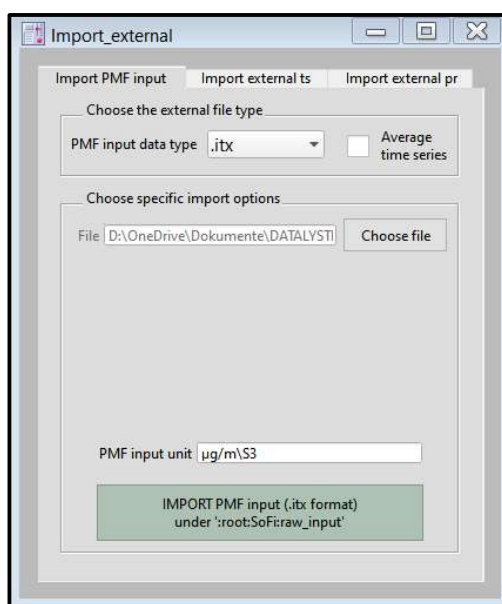


图 140

- 2) 通过点击“**PMF input**”将内部输入字符串，即“数据矩阵”、“误差矩阵”、“时间序列编号”、“变量编号”、“变量 txt”链接到适当的 IGOR waves。同样，使用导入部分导入应受约束的潜在指纹 **fingerprint**（图 1414.1.3

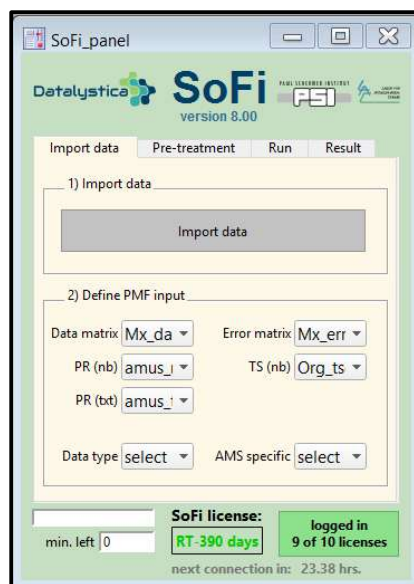


图 141

- 3) 点击“初始化 PMF 输入并检查缺失数据 **initialize PMF input & check for missing data**”按钮，以准备 PMF 输入结构，然后传递给 ME-2（图 142）

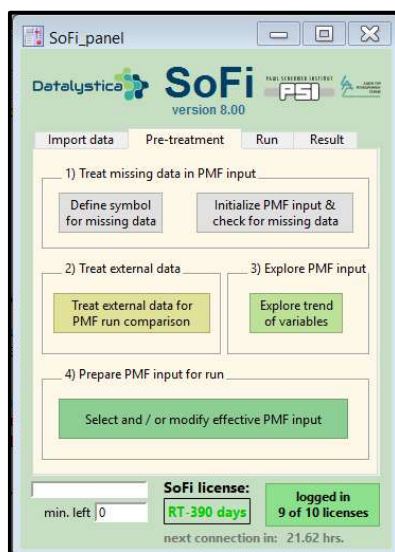


图 142

- 4) 点击“处理 PMF runs/结果比较的外部数据 **Treat external data for the PMF run/result comparison**”按钮，并从您的外部源谱文件中提取因子源谱文件约束的变量（图 1435.2.1）

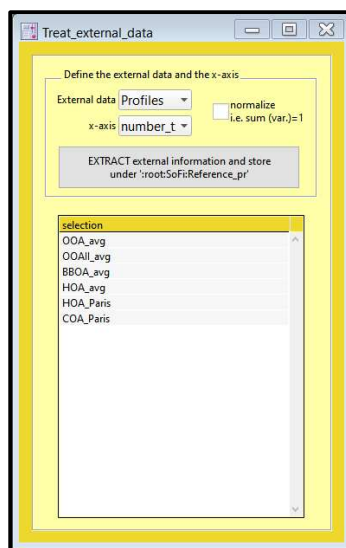


图 143: 参考源谱 reference profile 文件的提取。

注意!!!在导入 AMS 或 ACSM 数据的质谱时，请确保这些外部数据与 PMF 输入的离子碎片表匹配。否则，SoFi 中的提取程序将失败!!!通过 SoFi 导入参考源谱文件将自动处理此事。

10.3.2. 定义主要模型设置

10.3.2.1. 路径 Path、可执行 executables 文件数量和 HDF 文件名

- 5) 通过在 PMF runs 标签页上按“可执行文件文件夹 **folder of executable**”按钮来定义 ME2.exe 文件所在的位置。
- 6) 将可执行 ME2 文件的数量 **number of executables** 增加到最大值。这将提高计算期间的性能。
- 7) 在“存储当前 PMF runs 的 HDF5 文件名 **HDF5 file name storing the current PMF run**”下为 PMF runs 定义一个合适的名称。如果该名称已存在，SoFi 会提示用户尝试另一个名称（图 1446.1

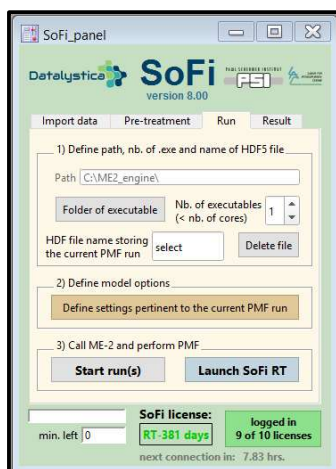


图 144

注意!!! “删除文件 Delete file” 按钮会从操作系统中删除指定的 HDF5。尽管有最后的警告提示，用户应注意不要删除重要信息。SoFi 不会维护 HDF5 文件的隐藏备份!!!

10.3.2.2. 定义因子数量和重复次数

点击橙色按钮“定义与当前 PMF runs 相关的设置 **Define settings pertinent to the current PMF run**”。这将弹出控制所有 PMF 设置的面板（图 1456.2

定义因子数量和重复次数。

“**starting factor**” 定义了模型所研究的最小因子数量（p）。

“**ending factor**” 定义模型所研究的最大因子数量（p）。

“**nb. of PMF calls**” 定义重复的次数。如果启用了 bootstrap 重采样，此选项每次都会生成一个新的重采样 PMF 输入。

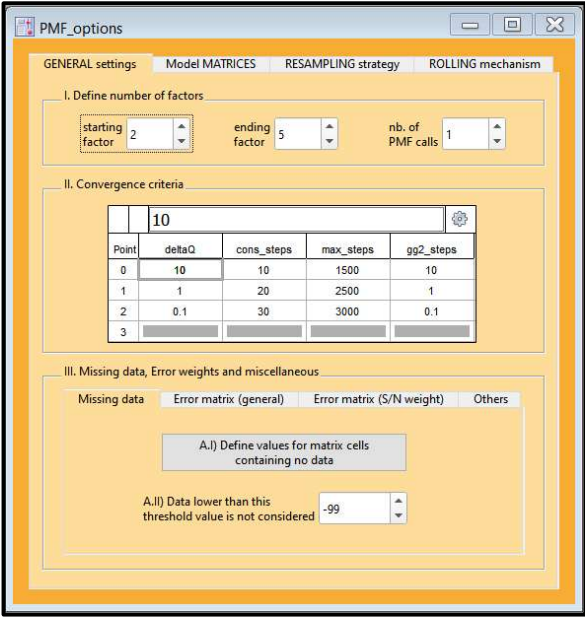


图 145

10.3.2.3.为主要因子配置定义种子 seed

转到橙色面板的第二个标签，以根据之前导入的外部源谱定义已知因子的源谱。使用以下选项（图 480

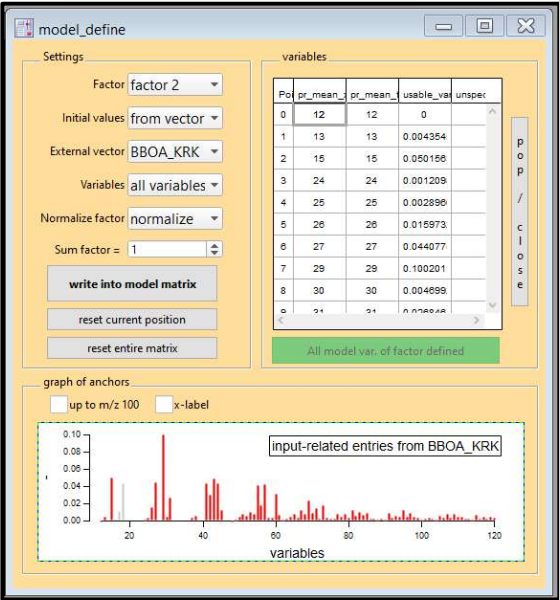


图 146

10.3.2.4.为一次源因子源谱约束一个值

使用 a 值技术约束定义的因子源谱。

要启用 a 值约束，请使用表格下方的下拉菜单。只有当按钮“启用 a 值约束 **A-value constraints are enabled**”变为绿色时，ME-2 求解器才会使用 a 值（图 147

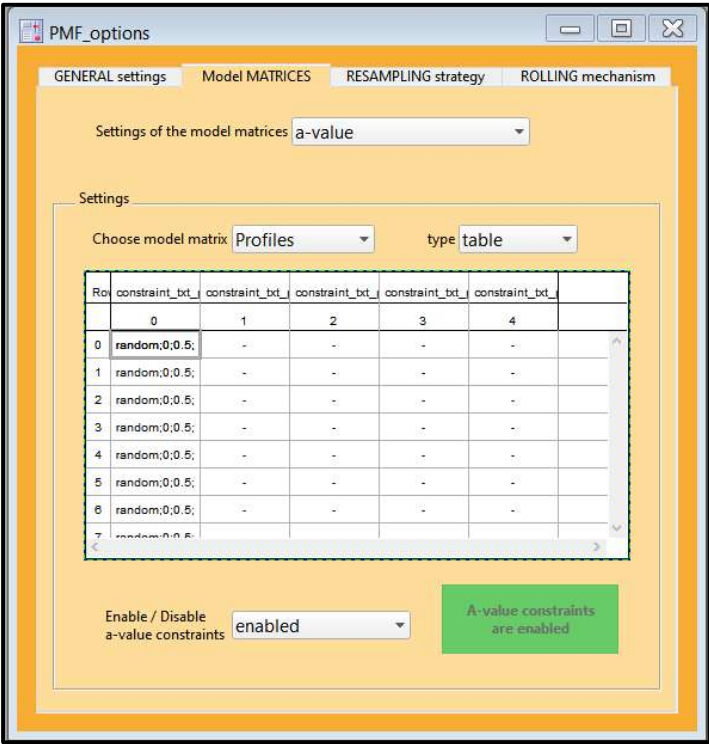


图 147

10.3.2.5.启用随机重采样

转到第三个标签并通过选择“启用 **enable**”选项来启用随机重采样（bootstrap 策略），并使用未阻止 **unblocked** 的选项（图 1486.4

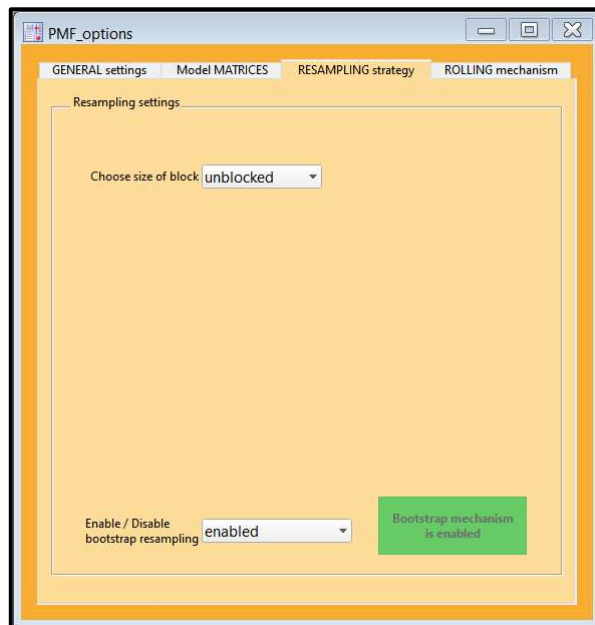


图 148

10.3.2.6. 启用滚动策略

转到第四个标签页并选择“启用 **enable**”选项，以及 **window width (days)** 和 **window shift (days)**（图 1496.5

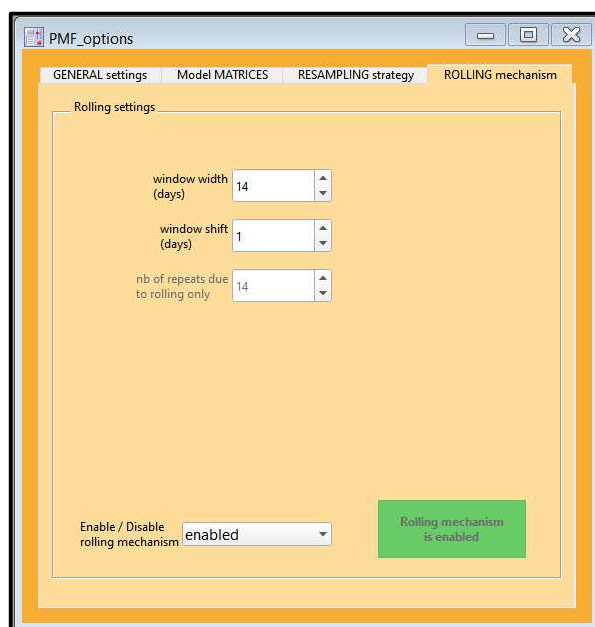


图 149

10.3.3. 调用 SoFi RT

点击蓝色按钮“调用 **Call SoFi RT**”以启动 RT 模块。

10.4. SoFi RT 模块

10.4.1. 定义输入 INPUT 选项。

图 150

Input path and type

“offline / online button” SoFi RT 可以在线 online 使用，即数据实时导入到 IGOR 中，或者离线 offline 使用，数据已经作为完整的 PMF 输入存在于 IGOR 中。然后，SoFi RT 执行 SoFi RT 的离线模拟。

“Path” 路径必须指向包含输入数据的同步文件夹。

“type of instrument” ACSM、Xact、Aethalometer。用户可以独立使用“开/关”“on”/“off”按钮将它们打开。

Specific instrument options

“instrument options” 对于 ACSM 数据，支持基于压力 pressure 和检测限 detection limit 的过滤选项。文件名已经硬编码，并直接来自采集软件的附加组件，不需要用户进一步关注。

对于 Xact 数据，存在两个商业仪器，其输出文件名不同。用户必须传递样本文件的基本名称，例如“data”。对于 Xact 625，元素存储在三个不同的文件中，“data1_data&time”、“data2_data&time”和“data3_data&time”。用户只需输入“data”。

如果启用了 BC 并且 BC 数据在 SoFi 中被自动导入和处理，黑碳数据将被分配到生物质燃烧（BCbb）和交通（BCtr）。用户可以选择启用或禁用可能的交通或生物质燃烧源与 BCtr 和 BCbb 之间的自动相关，分别适用于单一和/或组合仪器设置。默认情况下，仅对于 ACSM 数据，Sofi 会评估假设的 HOA/BBOA 与 BCtr/BCbb 之间的相关性。

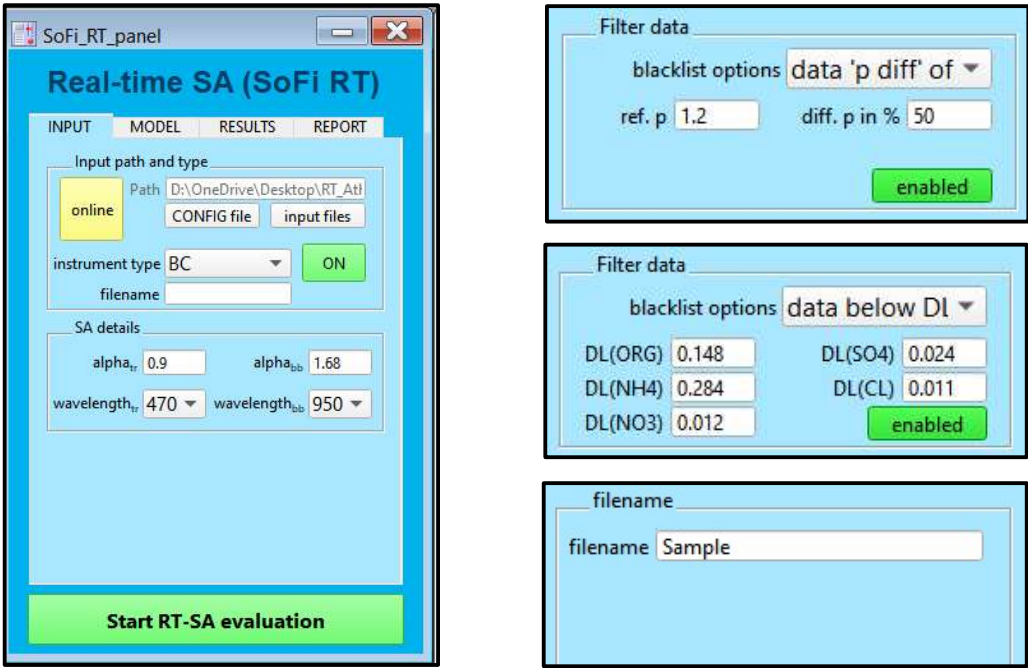


图 150: SoFi RT 面板的第一个标签页（左侧），以及基于压力 p 的过滤机制选项（右上角）和 ACSM 的检测限 DL（右中）以及 Xact 仪器的文件名 filename 选项（右下）。

10.4.2. 定义模型选项。

图 151

Rolling settings

“PMF initialize PMF after...” 在此天数之后，自动滚动 PMF 分析将开始。滚动参数下定义的窗口宽度（见 10.3.2.6

“PMF update every...” 在此天数之后，自动 rolling PMF 分析将被更新。滚动参数下定义的窗口移动（见 10.3.2.6

“SoFi RT restart rate...” 指示 SoFi RT 重启的频率。这通常是仪器获取最新扫描并更新 CMB 结果的扫描速率（结合以下选项）

“between PMF updates” 可选择没有更新 No update（SoFi RT 更新仅在顶部定义的自动 rolling PMF）。

CMB 在每次扫描后 CMB after each scan（CMB 在每次新扫描后额外执行）。

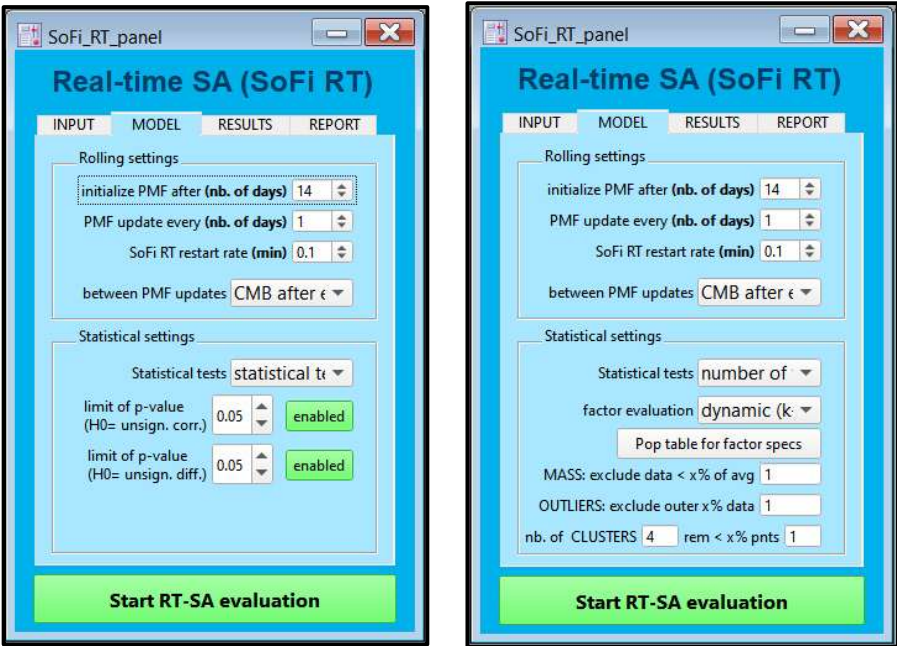


图 151: SoFi RT 面板的第二个标签页显示下面的统计 T 检验选项（左）和下面的动态因子评估选项（右）。

Statistical settings

- “Statistical tests”

对于与其他因子无显著相关性或差异的 p 值 T 检验限制可以启用或禁用
- “factor evaluation”

默认 default（来自橙色面板），用户应保持静态因子数量；动态 dynamic（每个窗口的因子评估）基于聚类分析（见下文以获取更多详细信息）。
- “Pop table for factor specs”

打开表格，用户可以将动态区分不同因子数量的相关信息传递给 SoFi RT。
- “MASS: exclude data...”

在滚动窗口中，低于 x % 的数据将从聚类分析的平均值中排除
- “OUTLIERS: exclude...”

在滚动窗口中，最外部 x 百分位的数据被排除用于聚类分析。
- “nb. of Clusters”

用于后续线性分析的聚类质心 centroid 数量
- “REMOVE Clusters...”

聚类质心中少于 x % 数据点的将从最终聚类质心列表中移除。

10.4.2.1. 动态因子评估

如果用户选择“静态 **static**”选项，则因子数量应保持不变，因为在 SoFi 中执行不同因子数量的 PMF runs 时目前除了一个较简单的平均控制机制外，没有自动分析机制。该控制机制确保在不同因子数量的 PMF runs 之间没有平均，仅允许在时间点逐一的基础上对最频繁因子数量的 PMF runs 进行平均。

相反，当选择“动态 **dynamic**”选项时，会在滚动窗口基础上进行额外的统计聚类分析，允许因子数量的变化。这具有避免扫描不同因子数量的优势，从而导致的 PMF runs 数量远少于传统手动 PMF 分析，在传统分析中，用户必须首先执行不同因子数量的 PMF runs，然后在后期分析中评估哪些因子解在测量活动中是最合适的。点击“因子规格弹出表 **Pop table of factor specs**”按钮（图 151

ACSM 数据的 m/z 44 和 m/z 43 示例

如果测量的数据在这个空间中呈三角形分布，至少需要两个因子来捕捉数据的变化程度。在它们的指纹中，这两个因子的 m/z 44 与 m/z 43 的比率很可能位于这个三角形的边缘（红点在图 152 图 152 图 152

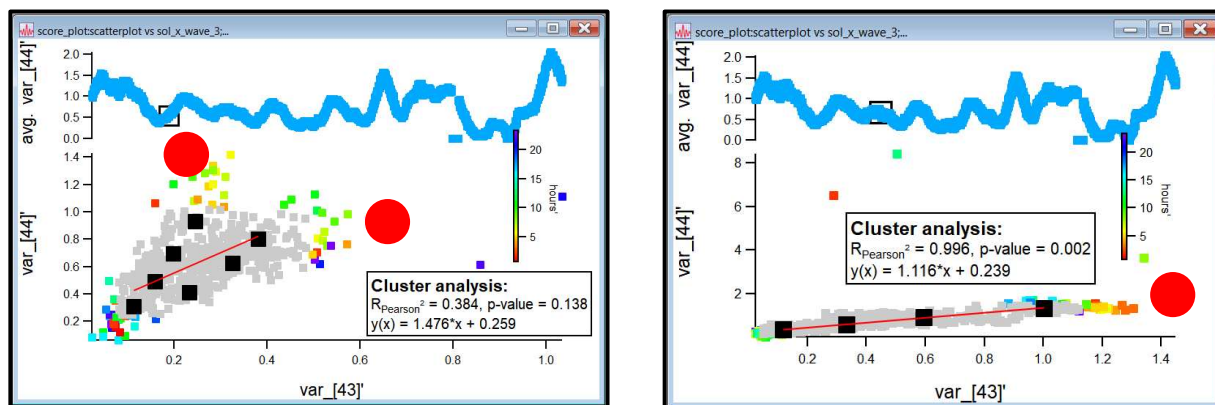
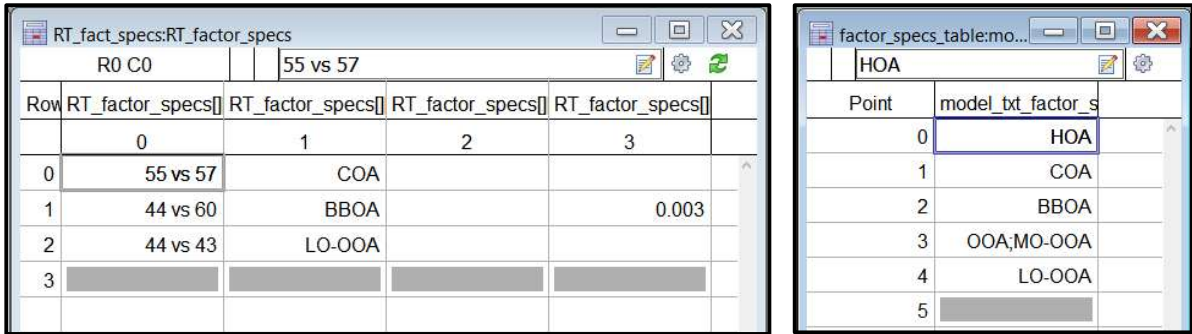


图 152: f44 与 f43 空间中的测量点（灰色）和聚类质心（黑色）。集群质心未对齐，因此需要两个因子来捕捉数据的变化程度（左），而在另一个示例中，它们是对齐的，仅需要一个因子（右）。

对于上述示例，用户将在弹出表的第一列中输入信息 44 与 43 “44 vs 43”（44 和 43 指的是变量 txt wave 中的 txt 信息）（表 106.3.17.5.3

表 10: 聚类算法的选项（左侧）和因子的完整列表（右侧）来自 PMF 选项面板



Row	RT_factor_specs	RT_factor_specs	RT_factor_specs	RT_factor_specs
	0	1	2	3
0	55 vs 57	COA		
1	44 vs 60	BBOA		0.003
2	44 vs 43	LO-OOA		
3				

Point	model_txt_factor_s
0	HOA
1	COA
2	BBOA
3	OOA,MO-OOA
4	LO-OOA
5	

表格的第三列和第四列可以用于传递关键变量的分数下限。低于此阈值的数据不用于聚类分析。第三列保留给第一个变量的下限，以防已知合理的阈值来自先验信息或文献数据。例如，0.003（Cubison et al., 2011）在第四列中表 10

来自表 10 表 11

表 11

关键变量空间	聚类中心	有效因子
	对齐	
m/z 55 与 m/z 57	是	没有 COA，只有 HOA
	否	使用了 HOA 和 COA
m/z 44 与 m/z 60	是	没有 BBOA，只有 OOA
	否	BBOA 和 OOA
m/z 44 与 m/z 43	是	仅有 OOA
	否	MO-OOA 和 LO-OOA

有效因子的各种组合来自表 11

- 2 因子解： HOA+OOA
- 3 因子解： HOA+MO-OOA+LO-OOA
 - HOA+BBOA+OOA
 - HOA+COA+OOA
- 4 因子解： HOA+BBOA+MO-OOA+LO-OOA
 - HOA+COA+BBOA+OOA
 - HOA+COA+MO-OOA+LO-OOA

➤ 5 因子解: HOA+COA+BBOA+MO-OOA+LO-OOA

请注意，在这些设置中，HOA 始终存在，OOA 要么分为 MO-OOA 和 LO-OOA，要么保持为 OOA。用户可以完全控制哪些因子以及可以评估多少因子及其组合，通过定义哪些因子可以根据聚类算法选项下的关键变量图进行禁用（表 10）

10.4.3. 定义 RESULT 选项。

图 153

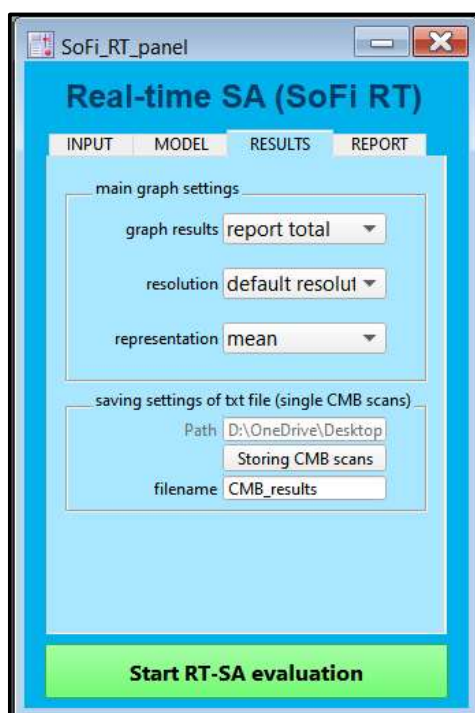


图 153

Main settings

- “graph results” 报告总计 report total（整个 PMF 解，包括最新的 CMB 更新），仅报告 PMF 调用 Report PMF calls only（仅 PMF 解），仅报告最后一次 PMF 调用 report last PMF call only（最后一次 PMF 窗口），仅报告 CMB 扫描 report CMB scans only（仅当天的 CMB runs）
- “resolution” 默认、每小时、每日、每周、每月和每年分辨率。
- “representation” 可以使用几种类型的表示，即均值、均值&标准差、中位数、中位数&IQR、中位数&IQR&10_90%、中位数&IQR&10_90%&最小_最大

Saving settings of txt file (single CMB scans)

- “Path” 显示用于存储每日 CMB txt 文件的选择路径。

“Storing CMB scans” 从操作系统中选择路径来保存 CMB 结果。

“filename” 每日 CMB txt 文件的前缀。

图 154

date&time	HOA	COA	BBOA	MO-OOA	LO-OOA	traffic	dust	coal_industry	biomass	chloride	sulfates	BC_tr	BC_bb
2024/03/07 00:00:00	nan	nan	nan	nan	nan	0.0733	0.0102	0.157	0	0.0025	0.516	0.0425	0.144
2024/03/07 01:00:00	nan	nan	nan	nan	nan	0.055	0.00286	0.159	0	0.00201	0.467	0.0909	0.121
2024/03/07 00:01:22	0.0292	0.693	0.364	0.83	0.825	nan	nan	nan	nan	nan	nan	0.0929	0.12
2024/03/07 00:11:28	1.23e-110	0.288	0.0917	0.422	0.72	nan	nan	nan	nan	nan	nan	0.106	0.122
2024/03/07 00:21:28	0.0256	0.225	0.103	0.355	0.704	nan	nan	nan	nan	nan	nan	0.163	0.103
2024/03/07 00:31:28	0.0373	0.281	0.04	0.35	0.746	nan	nan	nan	nan	nan	nan	0.0679	0.13
2024/03/07 00:41:28	0.0539	0.203	0.127	0.336	0.791	nan	nan	nan	nan	nan	nan	0.0471	0.131
2024/03/07 00:51:29	0.022	0.257	0.114	0.321	0.789	nan	nan	nan	nan	nan	nan	0.0567	0.124
2024/03/07 01:01:28	0.0284	0.276	0.0584	0.395	0.717	nan	nan	nan	nan	nan	nan	0.114	0.103
2024/03/07 01:11:28	0.057	0.25	0.0161	0.338	0.673	nan	nan	nan	nan	nan	nan	0.109	0.0951
2024/03/07 01:21:28	0.0776	0.353	0.0753	0.384	0.81	nan	nan	nan	nan	nan	nan	0.078	0.105
2024/03/07 01:31:28	0.046	0.389	0.121	0.424	0.752	nan	nan	nan	nan	nan	nan	0.0601	0.112

图 154

10.4.4. 报告 Report

图 155

图 155

Report content

“Station name” 输入采样点的名称

“Station information” 用户可以输入站点信息，例如国家、坐标、高度、站点代码和分类（城市、乡村等），这些信息将生成报告文件。

“Visualized results” 报告中可以可视化以下结果：贡献 contribution（默认饼图）、时间序列、概况、残差

“air quality limits” SoFi 允许将平均每日浓度汇总为分组的甜甜圈图。不同类别的限值以及颜色可以自定义。

File properties

图 156

“Report path” 报告存储文件夹的路径。

“Report file prefix” 报告的文件名。SoFi 将自动在定义的名称后添加日期和时间

“Write report NOW” 生成当前已导入并由 SoFi RT 评估的 PMF 结果的报告。

“AUTOMATED report” 每周或每月自动生成并存储在操作系统上的报告（即将推出）。

生成的报告将以.pdf 格式保存在定义的文件路径下。此外，主要图表可以存储为.png 文件。如果选中复选框，则 SoFi 将在定义的文件位置下生成一个以当前日期命名的新文件夹，图表将存储在该文件夹中。

报告第一页的顶部包含了由仪器测量的空气质量的一般概述。它包括当前的污染水平以及过去几天的概述。首先，列出了每个仪器测量的平均浓度及其数据覆盖率。其次，提供了一个同步平均值，所有仪器同时测量的结果（平均为一小时的平均值）。这包括按污染水平分类的每日平均值概述。

报告第一页的底部包含了 SoFi 解决的来源概述。左侧的饼图显示了当前来源贡献，右侧显示了过去几天的贡献（所有仪器同时测量的数据点）。

在第二页，提供了关于来源分配的更多细节，如因子时间序列（一般和日变化）、因子源谱以及缩放残差。

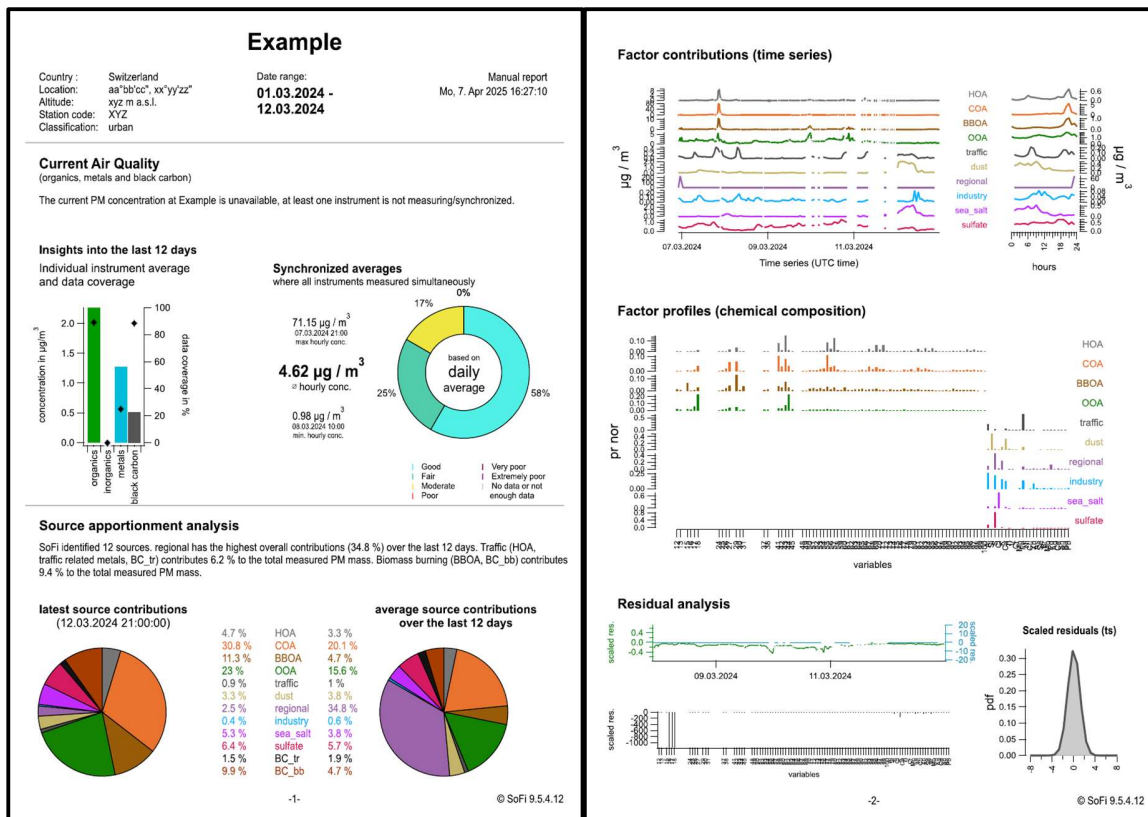


图 156

10.5. 启动 SoFi RT

10.5.1. 开始/停止 SoFi RT

可以在 SoFi RT 面板的任何标签页上按下“开始 RT-SA 评估 **Start RT-SA evaluation**”按钮以启动 SoFi RT（图 153 图 157

SoFi RT 有一个主要的无限循环，其中包含定义在“SoFi RT 重启率（分钟） **SoFi RT restart rate (min)**”下的空闲 idle 时间（10.4.2

Start RT-SA evaluation **绿色按钮**（SoFi RT 可以启动），**红色按钮粗体文本**（SoFi RT 正在运行，但可以中断），**红色按钮灰色文本**（SoFi RT 正在运行，无法中断）

中断 SoFi RT 并重新启动绝对没有任何害处。每当用户想要更详细地检查 SoFi RT 图表面板上的概览图，或者用户想要修改 SoFi RT 设置面板上的图表选项时，这是必要的。

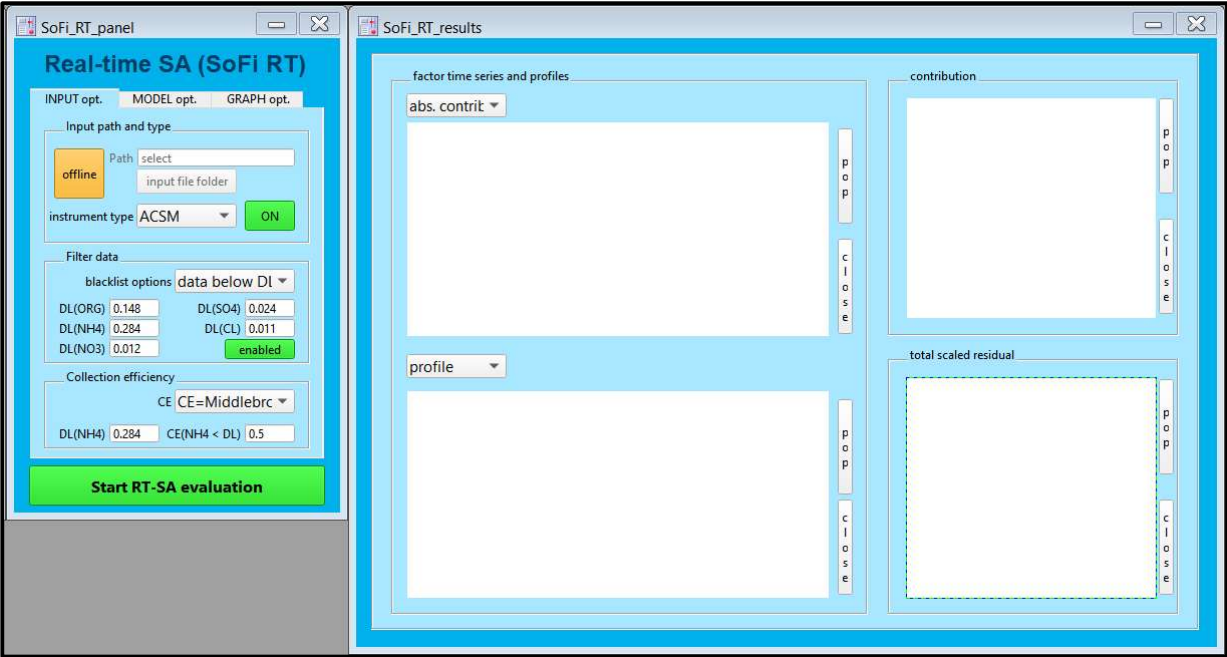


图 157

10.5.2. SoFi RT 概览结果

用户可以在 SoFi RT 结果面板上检查主要概览图（图 159

Graph options

- “factor time series” 绝对贡献、相对贡献、残差、日变化
- “factor profiles” 源谱、残差
- “contribution” 因子贡献的饼图
- “total scaled residual” 是所有点的缩放残差。它应该是单峰的，围绕 0，可能呈高斯形状。

首次启动 SoFi RT 时，会在“硬盘（例如，C）:\ME2 engine\”下创建一个名为 SoFi_RT 的文件夹，其中包含一个子文件夹 Variables。文件夹 SoFi_RT\Variables 包含一个硬编码的 IGOR 文本文件，名为“SoFi_RT_criteria”。该文件将由 SoFi 自动生成，并包含在 SoFi RT 运行期间提取适当 PMF runs 所需的用户定义标准列表。在第一次 SoFi RT 周期结束时，即当第一次自动滚动窗口的所有重复完成后，在 SoFi 提取用于平均的适当 PMF runs 之前，如果不存在 SoFi_RT_criteria 文件，SoFi 会自动生成一个包含虚拟标准的文件，该标准监控因子 1 的平均贡献，并考虑所有 PMF runs 以进行 PMF 平均。用户应在下一个空闲期间中断 SoFi RT，并使用标准面板更改标准（见 7.3

要获取空闲时间开始时的音调警告，用户可以在 SoFi 的主设置菜单中启用 IGOR 中的“哔声 **beep**”选项（见图 1583.4.6

请注意，每个 SoFi_RT_criteria 文件应专用于特定仪器集的单独 SoFi RT 运行。例如，如果从同一台 PC 执行多个 SoFi RT 运行且仪器套件有所不同，则用户应更改条件。这可以通过删除或将旧文件移动到另一个文件夹，或在 SoFi RT 的空闲阶段修改标准来完成。在重新初始化 SoFi RT 时，系统会询问用户是否将新的本地标准列表保存到 Windows 中。

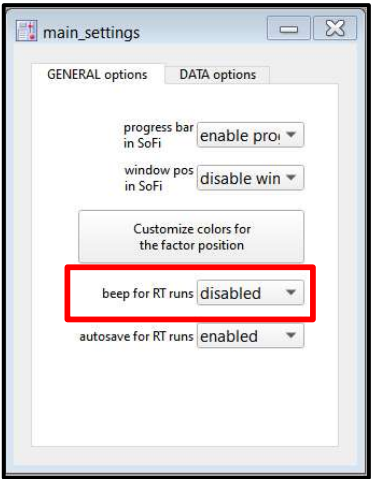


图 158

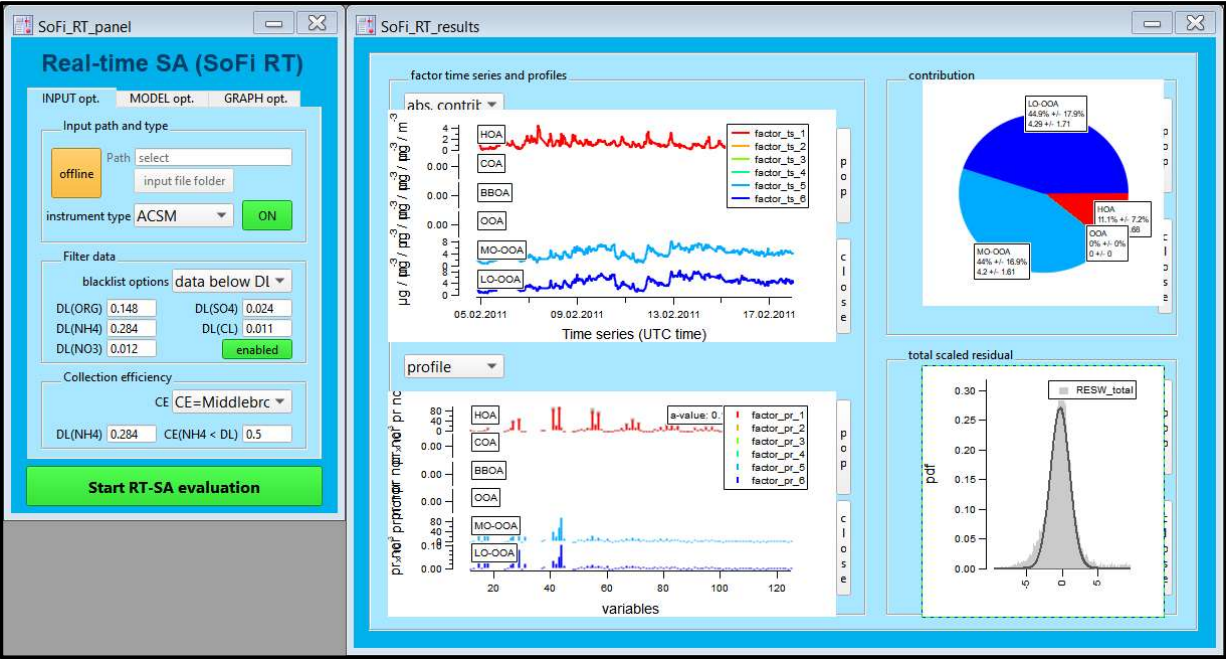


图 159

10.6. 使用 SoFi Pro 检查 SoFi RT 结果

SoFi RT 允许用户完全控制自动选择的 PMF runs。用户可以在停止 SoFi RT 后检查通过 SoFi RT 执行的所有 PMF 和 CMB runs。如果 SoFi RT 在 SoFi RT 结果面板上显示了一些结果，这意味着 HDF 文件已成功加载，并且标准面板在 SoFi 中已处理。因此，用户只需点击“**III.a Select PMF run...**”按钮即可查看已选择的 runs，并通过点击“**III.b Analyze chosen PMF run(s)**”直接检查所选的 PMF runs，无需延迟（图 160）

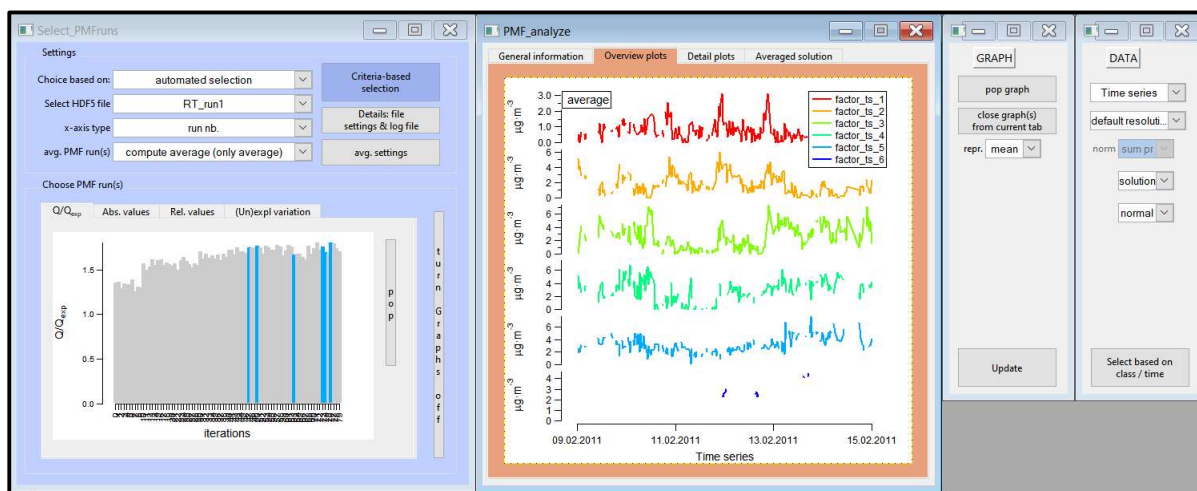


图 160

10.7. ACSM 数据的建议

根据(Crippa et al., 2014) 对离线 AMS/ACSM 数据分析的建议，在进行自动化和实时 rolling PMF 分析时，策略还包括限制已知存在于测量位置的一次有机气溶胶源（POA）的因子源谱，并留出一个到两个不受限制的因子用于评估氧化有机气溶胶（OOA）。这些因子，POA 和 OOA 因子，并不一定需要始终存在。当使用动态因子评估时，SoFi RT 会自动启用或禁用用户定义的因子（表 1010.4.2.1）

10.8. 从 AE33 生成每日 BC 文件

SoFi RT 需要每日 .csv 文件，其中包含由 Aethalometer 测量的光吸收数据，目前适用于 AE33。为了生成这些每日文件，必须在 AE33 的采集软件上定义一些设置，并生成批处理文件和一个任务调度程序，以触发这些文件的正确生成。读者可以参考由保罗·谢尔研究所 (PSI) 的 Dr. Manousos I. Manousakas 于 2022 年撰写的手册 AE33_data_SoFi_instructions，以及可以在我们主页上找到的批处理文件和示例文件 (<https://datalystica.com/downloads/>)。如果实验者决定使用与此处提议的不同软件包，请在从 AE33 采集软件中提取信息时注意以下事项：

- BC 文件名包含当天的信息，格式如下：
“前缀或空”+年份+_+月份+_+日期.csv，例如，2024_03_21.csv。
- 文件中的列表示以下内容：
 - 仪器序列号 serial number（最后 4 位数字）
 - 日期（“dd-mm-yy”）
 - 时间（“hh:mm”）
 - BC1, BC2, BC3, BC4, BC5, BC6, BC7
 - 流量 FlowC (lpm)

示例：

```
0804,"01-Mar-24","00:00", 939, 900, 840, 823, 781, 758, 796, 5.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0,
0.0, 0.0
```

11. 定义和缩略语列表

OA	有机气溶胶
BBOA	生物质燃烧有机气溶胶
COA	烹饪有机气溶胶
HOA	碳氢化合物类 有机气溶胶
ME-2	多线性引擎第二版
OOA	被氧化有机气溶胶
OS	操作系统
PMF	正矩阵因子分解
PMF2	双线性正矩阵因子分解算法的求解器
PMF3	三线性正矩阵分解算法的 PMF3 求解器
PMF call	ME-2 来解决 PMF 算法的特定设置，例如，改变因子数量、探索 a 值等。
PMF run	指的是从一次 PMF 调用中进行的单次 PMF run
PMF result/ solution	结果/解表示来自一个或多个 PMF 调用的所有具有合理环境意义的 PMF runs ，必须在彻底的旋转多义性探索和统计不确定性传播后考虑。
POA	一次有机气溶胶
SA	源解析
SOA	二次级有机气溶胶
SoFi	Source Finder 软件
SoFi Pro	启用 SoFi 中高级实用工具的密钥
SoFi RT	启用完全自动化 SoFi RT 软件的密钥

12. 参考文献

Äijälä, M., Heikkinen, L., Fröhlich, R., Canonaco, F., Prévôt, A. S. H., Junninen, H., Petäjä, T., Kulmala, M., Worsnop, D., and Ehn, M.: Resolving anthropogenic aerosol pollution types – deconvolution and exploratory classification of pollution events, *Atmos. Chem. Phys.*, 17, 3165–3197, <http://10.5194/acp-17-3165-2017>, 2017.

Aiken, A. C., DeCarlo, P. F., Kroll, J. H., Worsnop, D. R., Huffman, J. A., Docherty, K. S., Ulbrich, I. M., Mohr, C., Kimmel, J. R., Sueper, D., Sun, Y., Zhang, Q., Trimborn, A., Northway, M., Ziemann, P. J., Canagaratna, M. R., Onasch, T. B., Alfarra, M. R., Prevot, A. S. H., Dommen, J., Duplissy, J., Metzger, A., Baltensperger, U., and Jimenez, J. L.: O/C and OM/OC Ratios of Primary, Secondary, and Ambient Organic Aerosols with High-Resolution Time-of-Flight Aerosol Mass Spectrometry, *Environ Sci Technol*, 42, 4478–4485, <http://doi.org/10.1021/es703009q>, 2008.

Alton, M. W., Stark, H. J., Canagaratna, M. R., and Browne, E. C.: Generalized Kendrick analysis for improved visualization of atmospheric mass spectral data, *Atmos. Meas. Tech.*, 16, 3273–3282, <https://10.5194/amt-16-3273-2023>, 2023.

Brown, S. G., Eberly, S., Paatero, P., and Norris, G. A.: Methods for estimating uncertainty in PMF solutions: Examples with ambient air and water quality data and guidance on reporting PMF results, *Science of The Total Environment*, 518–519, 626–635, <https://doi.org/10.1016/j.scitotenv.2015.01.022>, 2015.

Canagaratna, M. R., Jimenez, J. L., Kroll, J. H., Chen, Q., Kessler, S. H., Massoli, P., Hildebrandt Ruiz, L., Fortner, E., Williams, L. R., Wilson, K. R., Surratt, J. D., Donahue, N. M., Jayne, J. T., and Worsnop, D. R.: Elemental ratio measurements of organic compounds using aerosol mass spectrometry: characterization, improved calibration, and implications, *Atmos Chem Phys*, 15, 253–272, <https://doi.org/10.5194/acp-15-253-2015>, 2015.

Canonaco, F., Crippa, M., Slowik, J. G., Baltensperger, U., and Prevot, A. S. H.: SoFi, an IGOR-based interface for the efficient use of the generalized multilinear engine (ME-2) for the source apportionment: ME-2 application to aerosol mass spectrometer data, *Atmos Meas Tech*, 6, 3649–3661, <https://doi.org/10.5194/amt-6-3649-2013>, 2013.

Canonaco, F., Slowik, J. G., Baltensperger, U., and Prevot, A. S. H.: Seasonal differences in oxygenated organic aerosol composition: implications for emissions sources and factor analysis, *Atmos Chem Phys*, 15, 6993–7002, <https://doi.org/10.5194/acp-15-6993-2015>, 2015.

Canonaco, F., Tobler, A., Chen, G., Sosedova, Y., Slowik, J. G., Bozzetti, C., Daellenbach, K. R., El Haddad, I., Crippa, M., Huang, R.-J., Furger, M., Baltensperger, U., and Prévôt, A. S. H.: A new method for long-term source apportionment with time-dependent factor profiles and uncertainty

assessment using SoFi Pro: application to 1 year of organic aerosol data, *Atmos Meas Tech*, 14, 923-943, <https://doi.org/10.5194/amt-14-923-2021>, 2021.

Chen, G., Canonaco, F., Tobler, A., Aas, W., Alastuey, A., Allan, J., Atabakhsh, S., Aurela, M., Baltensperger, U., Bougiatioti, A., De Brito, J. F., Ceburnis, D., Chazeau, B., Chebaicheb, H., Daellenbach, K. R., Ehn, M., El Haddad, I., Eleftheriadis, K., Favez, O., Flentje, H., Font, A., Fossum, K., Freney, E., Gini, M., Green, D. C., Heikkinen, L., Herrmann, H., Kalogridis, A.-C., Keernik, H., Lhotka, R., Lin, C., Lunder, C., Maasikmets, M., Manousakas, M. I., Marchand, N., Marin, C., Marmureanu, L., Mihalopoulos, N., Močnik, G., Nęcki, J., O'Dowd, C., Ovadnevaite, J., Peter, T., Petit, J.-E., Pikridas, M., Matthew Platt, S., Pokorná, P., Poulain, L., Priestman, M., Riffault, V., Rinaldi, M., Róžański, K., Schwarz, J., Sciare, J., Simon, L., Skiba, A., Slowik, J. G., Sosedova, Y., Stavroulas, I., Styszko, K., Teinmaa, E., Timonen, H., Tremper, A., Vasilescu, J., Via, M., Vodička, P., Wiedensohler, A., Zografou, O., Cruz Minguillón, M., and Prévôt, A. S. H.: European aerosol phenomenology – 8: Harmonised source apportionment of organic aerosol using 22 Year-long ACSM/AMS datasets, *Environment International*, 166, 107325, <https://doi.org/10.1016/j.envint.2022.107325>, 2022.

Crippa, M., Canonaco, F., Slowik, J. G., El Haddad, I., DeCarlo, P. F., Mohr, C., Heringa, M. F., Chirico, R., Marchand, N., Temime-Roussel, B., Abidi, E., Poulain, L., Wiedensohler, A., Baltensperger, U., and Prévôt, A. S. H.: Primary and secondary organic aerosol origin by combined gas-particle phase source apportionment, *Atmos Chem Phys*, 13, 1-47, <https://doi.org/10.5194/acp-13-8411-2013>, 2013.

Crippa, M., Canonaco, F., Lanz, V. A., Aijala, M., Allan, J. D., Carbone, S., Capes, G., Ceburnis, D., Dall'Osto, M., Day, D. A., DeCarlo, P. F., Ehn, M., Eriksson, A., Freney, E., Hildebrandt Ruiz, L., Hillamo, R., Jimenez, J. L., Junninen, H., Kiendler-Scharr, A., Kortelainen, A. M., Kulmala, M., Laaksonen, A., Mensah, A., Mohr, C., Nemitz, E., O'Dowd, C., Ovadnevaite, J., Pandis, S. N., Petaja, T., Poulain, L., Saarikoski, S., Sellegri, K., Swietlicki, E., Tiitta, P., Worsnop, D. R., Baltensperger, U., and Prevot, A. S. H.: Organic aerosol components derived from 25 AMS data sets across Europe using a consistent ME-2 based source apportionment approach, *Atmos Chem Phys*, 14, 6159-6176, <https://doi.org/10.5194/acp-14-6159-2014>, 2014.

Cubison, M. J., Ortega, A. M., Hayes, P. L., Farmer, D. K., Day, D., Lechner, M. J., Brune, W. H., Apel, E., Diskin, G. S., Fisher, J. A., Fuelberg, H. E., Hecobian, A., Knapp, D. J., Mikoviny, T., Riemer, D., Sachse, G. W., Sessions, W., Weber, R. J., Weinheimer, A. J., Wisthaler, A., and Jimenez, J. L.: Effects of aging on organic aerosol from open biomass burning smoke in aircraft and laboratory studies, *Atmos. Chem. Phys.*, 11, 12049-12064, <https://doi.org/10.5194/acp-11-12049-2011>, 2011.

Davison, A. C., and Hinkley, D. V.: *Bootstrap Methods and Their Application*, Cambridge University Press, 1997.

Efron, B.: 1977 Rietz Lecture - Bootstrap Methods - Another Look at the Jackknife, *Ann Stat*, 7, 1-26, <https://doi.org/10.1214/aos/1176344552>, 1979.

Henry, R., Norris, G. A., Vedantham, R., and Turner, J. R.: Source Region Identification Using Kernel Smoothing, *Environ Sci Technol*, 43, 4090-4097, 10.1021/es8011723, 2009.

Kendrick, E.: A Mass Scale Based on $\text{CH}_2 = 14.0000$ for High Resolution Mass Spectrometry of Organic Compounds, *Anal. Chem.*, 35, 2146-2154, <https://doi.org/10.1021/ac60206a048>, 1963.

Miller, M. S., Hidy, G. M., and Friedlander, S. K.: Chemical element balance for Pasadena aerosol, *J. Colloid Interface Sci.*, 39, 165-+, [https://doi.org/10.1016/0021-9797\(72\)90152-x](https://doi.org/10.1016/0021-9797(72)90152-x), 1972.

Ng, N. L., Canagaratna, M. R., Jimenez, J. L., Chhabra, P. S., Seinfeld, J. H., and Worsnop, D. R.: Changes in organic aerosol composition with aging inferred from aerosol mass spectra, *Atmos Chem Phys*, 11, 6465-6474, <https://doi.org/10.5194/acp-11-6465-2011>, 2011a.

Ng, N. L., Herndon, S. C., Trimborn, A., Canagaratna, M. R., Croteau, P. L., Onasch, T. B., Sueper, D., Worsnop, D. R., Zhang, Q., Sun, Y. L., and Jayne, J. T.: An Aerosol Chemical Speciation Monitor (ACSM) for routine monitoring of the composition and mass concentrations of ambient aerosol, *Aerosol Science and Technology*, 45, 780-794, <https://doi.org/10.1080/02786826.2011.560211>, 2011b.

Norris, G., Duvall, R. M., Brown, S. S., and S., B.: EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide, U.S. Environmental Protection Agency, 2014.

Ogulei, D., Hopke, P. K., Zhou, L., Paatero, P., Park, S. S., and Ondov, J. M.: Receptor modeling for multiple time resolved species: The Baltimore supersite, *Atmos Environ*, 39, 3751-3762, <https://doi.org/10.1016/j.atmosenv.2005.03.012>, 2005.

Paatero, P., and Tapper, U.: Positive Matrix Factorization - a nonnegative factor model with optimal utilization of error-estimates of data values, *Environmetrics*, 5, 111-126, <http://10.1002/env.3170050203>, 1994.

Paatero, P.: Least squares formulation of robust non-negative factor analysis, *Chemometr Intell Lab*, 37, 23-35, [http://10.1016/S0169-7439\(96\)00044-5](http://10.1016/S0169-7439(96)00044-5), 1997.

Paatero, P.: The multilinear engine - A table-driven, least squares program for solving multilinear problems, including the n-way parallel factor analysis model, *J Comput Graph Stat*, 8, 854-888, <http://10.2307/1390831>, 1999.

Paatero, P., Hopke, P. K., Song, X. H., and Ramadan, Z.: Understanding and controlling rotations in factor analytic models, *Chemometr Intell Lab*, 60, 253-264, Doi 10.1016/S0169-7439(01)00200-3, 2002.

Paatero, P., and Hopke, P. K.: Discarding or downweighting high-noise variables in factor analytic models, *Anal. Chim. Acta*, 490, 277-289, [https://doi.org/10.1016/S0003-2670\(02\)01643-4](https://doi.org/10.1016/S0003-2670(02)01643-4), 2003.

Paatero, P.: User's guide for positive matrix factorization programs PMF2 and PMF3, University of Helsinki, Helsinki, Finland, 2004.

Paatero, P., and Hopke, P. K.: Rotational tools for factor analytic models, *J. Chemom.*, 23, 91-100, <https://doi.org/10.1002/cem.1197>, 2009.

Paatero, P.: User's guide for positive matrix factorization programs PMF2 and PMF3. In: University of Helsinki, Helsinki, Finland, 2010.

Paatero, P., Eberly, S., Brown, S. G., and Norris, G. A.: Methods for estimating uncertainty in factor analytic solutions, *Atmos. Meas. Tech.*, 7, 781-797, <https://doi.org/10.5194/amt-7-781-2014>, 2014.

Parworth, C., Fast, J., Mei, F., Shippert, T., Sivaraman, C., Tilp, A., Watson, T., and Zhang, Q.: Long-term measurements of submicrometer aerosol chemistry at the Southern Great Plains (SGP) using an Aerosol Chemical Speciation Monitor (ACSM), *Atmos Environ*, 106, 43-55, 10.1016/j.atmosenv.2015.01.060, 2015.

Petit, J. E., Favez, O., Albinet, A., and Canonaco, F.: A user-friendly tool for comprehensive evaluation of the geographical origins of atmospheric pollution: Wind and trajectory analyses, *Environmental Modelling & Software*, 88, 183-187, <https://doi.org/10.1016/j.envsoft.2016.11.022>, 2017.

Polissar, A. V., Hopke, P. K., Paatero, P., Malm, W. C., and Sisler, J. F.: Atmospheric aerosol over Alaska: 2. Elemental composition and sources, *Journal of Geophysical Research: Atmospheres*, 103, 19045-19057, <https://doi.org/10.1029/98JD01212>, 1998.

Sandradewi, J., Prévôt, A. S. H., Weingartner, E., Schmidhauser, R., Gysel, M., and Baltensperger, U.: A study of wood burning and traffic aerosols in an Alpine valley using a multi-wavelength Aethalometer, *Atmos Environ*, 42, 101-112, <http://dx.doi.org/10.1016/j.atmosenv.2007.09.034>, 2008.

Silverman, B. W.: *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall/CRC, 1986.

Slowik, J. G., Vlasenko, A., McGuire, M., Evans, G. J., and Abbatt, J. P. D.: Simultaneous factor analysis of organic particle and gas mass spectra: AMS and PTR-MS measurements at an urban site, *Atmos Chem Phys*, 10, 1969-1988, <https://doi.org/10.5194/acp-10-1969-2010>, 2010.

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., and Jimenez, J. L.: Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data, *Atmos Chem Phys*, 9, 2891-2918, <http://10.5194/acp-9-2891-2009>, 2009.

Van Krevelen, D. W.: Graphical-statistical method for the study of structure and reaction processes of coal, *Fuel*, 24, 269-284, 1950.

Visser, S., Slowik, J. G., Furger, M., Zotter, P., Bukowiecki, N., Canonaco, F., Flechsig, U., Appel, K., Green, D. C., Tremper, A. H., Young, D. E., Williams, P. I., Allan, J. D., Coe, H., Williams, L. R., Mohr, C., Xu, L., Ng, N. L., Nemitz, E., Barlow, J. F., Halios, C. H., Fleming, Z. L., Baltensperger, U., and Prévôt, A. S. H.: Advanced source apportionment of size-resolved trace elements at multiple sites in London during winter, *Atmos Chem Phys*, 15, 11291-11309, <https://doi.org/10.5194/acp-15-11291-2015>, 2015.

Wold, S., Esbensen, K., and Geladi, P.: Principal component analysis, *Chemometr Intell Lab*, 2, 37-52, [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9), 1987.

Zotter, P., Herich, H., Gysel, M., El-Haddad, I., Zhang, Y., Močnik, G., Hüglin, C., Baltensperger, U., Szidat, S., and Prévôt, A. S. H.: Evaluation of the absorption Ångström exponents for traffic and wood burning in the Aethalometer-based source apportionment using radiocarbon measurements of ambient aerosol, *Atmos. Chem. Phys.*, 17, 4229-4249, <https://doi.org/10.5194/acp-17-4229-2017>, 2017.