

Note technique

Traitement des données ACSM pour les
études de sources de particules fines :
Recommandations pour l'application de la
méthode PMF avec le logiciel SoFi Pro

Septembre 2025



Groupement d'intérêt scientifique



«L'expertise au service de la qualité de l'air»



Travaux réalisés par l'Ineris et IMT Nord Europe
dans le cadre du

Laboratoire Central de Surveillance de la Qualité de l'Air

Traitement des données ACSM pour les études de sources de particules fines :

Recommandations pour l'application de la méthode PMF avec le
logiciel SoFi Pro

Hasna CHEBAICHEB - Ineris

Vérification : FAVEZ Olivier – Ineris

Approbation : Ineris : Document approuvé le 24/07/2026 par MARCHAND-MARQUIS
CAROLINE

Liste des personnes ayant participé à l'étude : Véronique Riffault – IMT Nord Europe

Table des matières

1. RAPPEL DES PRINCIPES DE LA PMF.....	9
2. PRÉPARATION DES INPUTS POUR LA PMF.....	11
2.1 Import des données ACSM	11
2.2 Vérification des paramètres d'étalonnage.....	11
2.3 Validation des données.....	11
2.3.1 Vérification des diagnostics basiques	11
2.3.2 Vérification de quelques ions spécifiques	12
2.3.3 Application des corrections	12
2.3.4 Vérification de la balance ionique	12
2.3.5 LOD	12
2.3.6 Vérification de la fermeture chimique.....	13
2.3.7 Vérification et correction de la date ACSM Local Time	13
2.4 Export PMF	13
2.4.1 Possible problème lié à la « minError »	14
2.4.2 Possible problème lié à un « DwellTime » égal à zéro	15
2.4.3 Problème lié au nb. of sets.....	15
3. APPLICATION DE LA PMF SUR SOFI PRO.....	17
3.1 Installation et initialisation de SoFi Pro	17
3.2 Import des données et pré-traitement.....	17
3.2.1 Import des données	17
3.2.2 Définition des données d'entrée.....	20
3.2.3 Pré-traitement des données.....	21
3.3 Exécution de la PMF	25
3.3.1 Définir le chemin et le nom du fichier HDF5.....	26
3.3.2 Définir les options du modèle.....	26
3.3.3 Appeler ME2 et exécuter le PMF	29
3.4 Visualisation des résultats PMF	30
3.4.1 Définir le chemin du fichier ME2.exe.....	30
3.4.2 Sélectionner le fichier HDF5	30
3.4.3 Sélection et analyse de l'exécution PMF	30
3.5 Exécuter la « PMF avec contraintes »	38
3.5.1 Seed.....	39
3.5.2 Approche par a-value.....	40
3.5.3 fpeak	43
3.5.4 « Pulling equations »	44
3.6 Utiliser la stratégie de rééchantillonnage « Bootstrap ».....	45
3.7 Exécuter la « PMF glissante »	46
3.8 Sélection par critères	47
3.9 Solution moyenne 'Averaged solution'	50
4. CAS CONCRETS D'APPLICATION.....	52
4.1 PMF classique.....	52

4.2	PMF glissante	52
4.2.1	Étapes de définition des paramètres d'exécution PMF	53
4.2.2	Étapes de définition des critères	53
4.3	« Multi-time PMF »	54
5.	RÉFÉRENCES.....	56
6.	ANNEXE	57

Liste des figures

Figure 1 :	Vérification de l'ACSM_LocalTime	13
Figure 2 :	Panel « Export »	14
Figure 3 :	Exemple d'incertitudes très faibles	14
Figure 4 :	Procédure ACSM Local	15
Figure 5 :	Waves de la matrice DAQ	15
Figure 6 :	Exemple de problème lié au nb. of sets	16
Figure 7 :	Panneau SoFi et Data Browser	17
Figure 8 :	Import des données d'entrées PMF	18
Figure 9 :	Import des données externes	19
Figure 10 :	Import des profils externes	19
Figure 11 :	Définir les inputs PMF	20
Figure 12 :	Traitement des données manquantes dans les données d'entrée PMF	21
Figure 13 :	Treat external data 'profiles'	22
Figure 14 :	Treat external data 'Time Series'	22
Figure 15 :	Explore PMF input data ; exemple de profil journalier de m/z 55	23
Figure 16 :	Exemples pour ' Explore PMF input data'	23
Figure 17 :	Sélection des données d'entrées	24
Figure 18 :	PMF input panel to control user-specified PMF	24
Figure 19 :	Panneau de saisie du PMF pour préparer l'exécution	25
Figure 20 :	Onglet PMF run	25
Figure 21 :	Définition du chemin et du nom du fichier HDF	26
Figure 22 :	paramètres PMF	27
Figure 23 :	Onglet « Matrice d'erreurs (pondération) »	28
Figure 24 :	Paramètres principaux de la matrice d'erreur	28
Figure 25 :	S2N and weighted errors for the averaged data	28
Figure 26 :	Onglet « Autres paramètres »	29
Figure 27 :	Exécution PMF	29
Figure 28 :	Panneau de résultats	30
Figure 29 :	Sélection de quelques solutions dans le fichier HDF	31
Figure 30 :	Graphique de la variation inexplicée	31
Figure 31 :	Onglet « Informations générales »	32

Figure 32 : Onglet « Overview plots »	33
Figure 33 : Différents tracés à réaliser dans le panneau 'Overview'	34
Figure 34 : Différents tracés à réaliser dans le panneau « Detail plots » ; onglet « Fraction »	35
Figure 35 : Différents tracés à réaliser dans le panneau « Detail plots » ; Onglet Standard	35
Figure 36 : Exemples de graphiques possibles de l'onglet « Scatter »	36
Figure 37 : Résultat de l'option de soustraction de modèle pour f44 vs f43	36
Figure 38 : Histogramme des résidus 'scaled'	37
Figure 39 : Onglet Corrélation ; Visualisation graphique et tableau de corrélations	38
Figure 40 : Panneau Matrices du modèle	39
Figure 41 : Onglet « factor specs », noms par défaut (à gauche)	39
Figure 42 : Exemple 'seed'	40
Figure 43 : Onglet « a-value »	41
Figure 44 : « Enabled » a-value	42
Figure 45 : Paramètres de valeur a	42
Figure 46 : Analyse de sensibilité	43
Figure 47 : Onglet fpeak	44
Figure 48 : Onglet 'pulling equations'	44
Figure 49 : Exemple de contrainte BBOA par équation d'extraction	45
Figure 50 : Exemple de contrainte BBOA par équation d'extraction avec analyse de sensibilité	45
Figure 51 : Panneau de stratégie de rééchantillonnage	46
Figure 52 : Onglet PMF glissante	46
Figure 53 : Q/Qexp et pourcentage de points temporels non modélisés sur l'ensemble des données en fonction de la taille de la fenêtre (swin), de la valeur a maximale (amax) et du nombre de répétitions PMF par fenêtre (xPMF) (Canonaco et al. 2021)	47
Figure 54 : Sélection des exécutions PMF selon le panneau de critères	47
Figure 55 : Panneau de critères PMF	48
Figure 56 : premier critère « Corrélation entre le facteur 1 et BCtr »	48
Figure 57 : Définition des critères ; type image ou vidéo dans SoFi 8	49
Figure 58 : panneau 'Inspect the selection'	50
Figure 59 : Panneau « Averaged solution »	50
Figure 60 : Quelques graphiques de l'onglet « Averaged solution »	51
Figure 61 : Étapes de définition des paramètres d'exécution PMF	53
Figure 62 : Étapes de définition des critères	53
Figure 63 : Liste de quelques critères de sélection	54
Figure 64 : Panneau de prétraitement ; Préparer l'entrée PMF	54
Figure 65 : Panneau d'exécution PMF ; Définir les options du modèle	54

Figure 66 : Plot général	55
Figure 67 : Panneau des tracés détaillés ; onglet Résidus.....	55

Résumé

Ce document définit un protocole opérationnel pour l'étude de la répartition des sources des aérosols organiques à partir des données ACSM (Aerosol Chemical Speciation Monitor) à l'aide d'un outil statistique de type PMF (Positive Matrix Factorization). Il détaille toutes les étapes de la préparation des données dans Igor au sein du logiciel d'acquisition de données ACSM (ACSM Local, Aerodyne Research Inc.) pour l'export des matrices de concentrations organiques et leurs incertitudes, puis celles permettant de réaliser l'analyse PMF à l'aide du logiciel SoFi Pro (Source Finder Datalystica Ltd., Villigen, Suisse).

De plus amples informations sont également disponibles dans le document « User guide for ACSM data processing and PMF application using SoFi Pro. 2025 » [\(hal-04935126\)](#) préalablement établi avec les AASQAs dans le cadre des travaux de thèse LCSQA d'Hasna Chebaicheb.

Abstract

This document defines the procedures for organic aerosol source apportionment from ACSM (Aerosol Chemical Speciation Monitor) data using the PMF (Positive Matrix Factorization) method on SoFi Pro software. It details all the steps involved in preparing data in Igor within the ACSM data acquisition software (ACSM Local, Aerodyne Research Inc.) for exporting organic concentration matrices and their uncertainties, and then the steps involved in performing PMF with SoFi Pro (Source Finder Datalystica Ltd., Villigen, Switzerland).

Further information is also available in the document "User guide for ACSM data processing and PMF application using SoFi Pro. 2025" ([hal-04935126](#)) previously prepared with the AASQAs as part of Hasna Chebaicheb's LCSQA thesis work.

1. Rappel des principes de la PMF

La méthode Positive Matrix Factorisation (PMF) est une méthode source-récepteur, couramment utilisée dans les études de répartition des sources pour identifier les sources des polluants. Elle permet de décomposer la matrice des observations (x_{ij}) en deux matrices non négatives de profils de facteurs de sources statiques (f_{kj}) et de leurs contributions en fonction du temps (g_{ik}). Dans sa forme standard, la PMF résout l'équation du bilan massique :

$$x_{ij} = \sum_{k=1}^p g_{ik} f_{kj} + e_{ij} \quad (Eq. 1)$$

Où e_{ij} est la matrice des résidus, et p le nombre de facteurs.

La PMF se base sur un algorithme interactif des moindres carrés qui vise à minimiser itérativement une quantité Q qui représente la somme totale des carrés des résidus e_{ij} du modèle, pondérés par les incertitudes de mesure σ_{ij} pour tous les points ij .

$$Q = \sqrt{\sum_i^m \sum_j^n \frac{e_{ij}^2}{\sigma_{ij}^2}} \quad (Eq. 2)$$

L'application de la PMF peut entraîner des difficultés qui sont liées à l'ambiguïté rotationnelle, car différentes combinaisons de G et F peuvent donner des valeurs similaires de Q , dont certaines peuvent contenir des facteurs mixtes et/ou des solutions incohérentes du point de vue environnemental. Pour remédier à ce problème, et si des sources sont connues a priori pour le site étudié, l'approche par "*a-value*" permet de contraindre les profils de facteurs prévus ou les séries temporelles à l'aide du solveur ME-2 dans le logiciel SoFi Pro. Cette approche permet de faire varier chaque variable d'un spectre de masse d'un certain pourcentage selon la valeur de "*a*" (qui est comprise entre 0 - aucune variabilité - et 1 - 100% de variabilité autorisée) par rapport à un spectre de référence.

De plus, la méthode de rééchantillonnage "*Bootstrap*" permet de déterminer des solutions PMF robustes et stables et d'évaluer les incertitudes statistiques du modèle. Cette méthode implique la création de sous-échantillons aléatoirement à partir des données d'entrée, permettant ainsi de vérifier si ces sous-échantillons conduisent aux mêmes solutions et contributions de facteurs que celles identifiées précédemment.

Il existe plusieurs types de PMF :

- **PMF saisonnières**

Pour effectuer une analyse pluriannuelle, une PMF standard peut être réalisée par saison dans une année de référence pour (i) déterminer le nombre de facteurs attendus dans l'ensemble de la base de données et (ii) récupérer les spectres de référence des facteurs primaires.

- **PMF glissante**

Elle consiste à diviser la base de données en périodes plus courtes (par exemple des fenêtres d'un mois, décalées de sept jours), ce qui présente l'avantage de permettre une certaine variabilité dans les profils des facteurs tout au long de la période d'observation, à travers les saisons et les années.

- **PMF ‘multi-time’**

La PMF multi-time (MTR-PMF) permet d’appliquer la PMF en tenant compte des mesures recueillies à différentes résolutions temporelles (TR), comme le montre l’équation suivante :

$$x_{sj} = \frac{1}{ts_2 - ts_1 + 1} \sum_{k=1}^p \left(f_{jk} \sum_{t=ts_1}^{ts_2} g_{ik} + \eta_{jm} \right) + e_{sj} \quad (Eq. 3)$$

Où ts_1 et ts_2 sont les instants de début et de fin de l’échantillon i , et η_{jm} est un facteur d’ajustement pour l’espèce j si celle-ci est répliquée dans différents TR ou mesurée à l’aide de méthodes analytiques différentes (représentées par l’indice m).

Ces différentes méthodes de PMF sont mises en œuvre dans le cadre du dispositif national de surveillance de la qualité de l’air afin d’identifier et de quantifier les sources de particules. Elles s’appuient sur des jeux de données issus des différents instruments déployés au sein de ce dispositif (notamment ACSM, AE33, ainsi que les prélèvements sur filtre) par les AASQAs (Associations agréées pour la surveillance de la qualité de l’air).

Par ailleurs, des sessions de formation ont été dispensées aux AASQAs (AirParif, Atmo Occitanie, Atmo Nouvelle-Aquitaine, Atmo Auvergne-Rhône-Alpes, Atmo Normandie, Atmo Grand-Est et Air Breizh), portant à la fois sur ces méthodologies et sur les outils associés de traitement de données.

Le présent manuel a pour objectif de décrire de manière structurée les différentes étapes nécessaires à la préparation des données d’entrée pour la PMF, ainsi que les procédures d’exécution de la PMF à l’aide du logiciel SoFi Pro.

2. Préparation des inputs pour la PMF

Ce premier chapitre propose une brève description des différentes étapes à suivre afin de préparer et valider les données ACSM avant d'exporter les matrices organiques « concentrations » et « erreurs ». Pour compléments d'information, ces étapes sont détaillées dans le guide méthodologique ACSM LCSQA¹.

2.1 Import des données ACSM

L'import des données brutes ACSM.itx se fait sur le logiciel ACSM Local dans l'onglet « Dataload » du logiciel ainsi que toutes les opérations de QA/QC et validation des données et l'export PMF.

2.2 Vérification des paramètres d'étalonnage

La vérification des paramètres de calibration RF et RIE doit être faite dans le « review batch ». Ces paramètres correspondent aux paramètres choisis lors de l'étalonnage de l'ACSM sur la période concernée. Si les valeurs RIE et RF de deux étalonnages consécutifs sont jugées cohérentes, il est possible d'utiliser leur moyenne ; sinon, les valeurs RIE et RF glissantes au cours du temps doivent être utilisées, notamment après ouverture de la chambre à vide et maintenance de la zone d'ionisation de l'instrument (e.g. changement de filament, ...).

2.3 Validation des données

Cette étape est essentielle pour la validation des espèces chimiques non-réfractaires PM₁ (NR-PM₁) dont l'aérosol organique (AO). Elle est principalement faite dans l'onglet « Analysis » de l'ACSM local et contient plusieurs phases :

2.3.1 Vérification des diagnostics basiques

Dans le menu « Basic Diagnostics », deux graphiques (« *tuning plots* ») permettent d'afficher plusieurs paramètres qui montrent le bon fonctionnement de l'instrument dont le Heater Bias, la tension du détecteur SEM et l'Emission Curent, et le graphique des « diagnostics » qui assure la validité des données dont l'Inlet pressure et l'Airbeam qui doivent être dans leur gamme de variation (Tableau 1) sinon une invalidation des données aberrantes peut être effectuée.

Tableau 1 : Validation des données

Paramètres	Pression d'entrée	Airbeam	Humidité relative	Température du vaporiseur
Validation	1,3 ± 0,2 Torr	10 ⁷ ± 30 % d'ions s ⁻¹	<60%	Régulée à partir de la courbe d'étalonnage de tension initialement définie par le fabricant

¹<https://www.lcsqa.org/fr/rapport/guide-methodologique-mesure-de-la-composition-chimique-des-particules-submicroniques-non-réfractaires-par-ACSM>

2.3.2 Vérification de quelques ions spécifiques

En utilisant la fonction « Select Ion List », il est possible de vérifier la série temporelle de quelques ions spécifiques. Par exemple les pics des ions correspondant à l'eau (m/z 18) et à l'air (m/z 28, 32 et 40) qui semblent anormaux/aberrants peuvent être supprimés via une procédure de nettoyage exécutée dans le logiciel Igor Pro. D'autres ions à savoir les ions associés aux chlorures (m/z 35 et 36) ou à certains composés organiques (m/z 60, 43, 44 et 55) peuvent être consultés pour vérifier l'influence de certaines sources (sels marins, combustion de la biomasse, AOS, trafic ...).

2.3.3 Application des corrections

Cette étape peut être exécutée à l'aide de la fonction « Corrections » du logiciel. Nous avons essentiellement deux corrections :

2.3.3.1 Time Series Correction

Cette correction permet la vérification de la variabilité de la sensibilité du détecteur. Elle inclut l'ajustement des paramètres techniques de l'instrument tels que « référence » N₂, O₂, Ar, P, T, Naph., et RF, tels qu'ils ont été mesurés lors de l'étalonnage et qui influencent les concentrations mesurées.

2.3.3.2 Ion Transmission correction

Cette correction permet la vérification de la transmission relative du spectromètre de masse basée sur les pics des ions de naphthalène (51, 62, 76, 102 et 128 (normalisés à 1 en dessous de 51). La courbe moyenne de RIT doit être comparable à la courbe par défaut, mais une vérification des séries temporelles des ions de RIT (m/z 62, 76, 102, et 128) doit être examinée, en particulier lorsque les écart types sont très larges.

2.3.3.3 Correction CDCE

La correction de l'efficacité de collection (CE) est basée sur l'algorithme de Middlebrook (CE = 0.5, la valeur minimale de CE est 0,5). Cette correction est appliquée en exécutant la procédure Igor correspondante, permettant de calculer CE_{dry} et de corriger les concentrations massiques des espèces chimiques de l'instrument (AO, SO₄, NO₃, NH₄ et Cl).

2.3.4 Vérification de la balance ionique

Cette vérification peut se faire grâce à la fonction « Frag Data Diagnoxe », en évaluant la corrélation entre les concentrations de NH₄ mesurées et prévues, ceci permet la validation des valeurs RIE de NH₄ et SO₄. En effet, la concentration d'ammonium mesurée par l'instrument doit être proche des concentrations d'ammonium « prédites » et la pente soit linéaire et proche de 1+25%, à part dans certains cas d'épisodes de pollution avec la présence d'aérosols acides ou de SO₄ ou NO₃ liés à la matière organique (organo-nitrate, organo-sulfate) par exemple. Dans ces cas, l'ammonium prédit serait surestimé et le rapport serait réduit.

2.3.5 LOD

Une analyse des limites de détection (LoD) pour les différentes espèces chimiques est recommandée. Elles peuvent être calculés comme suit : 3*Ecart-type (des mesures du blancs effectuées sur site). Les données inférieures à -3*LoD doivent être invalidées.

2.3.6 Vérification de la fermeture chimique

La somme des espèces chimiques NR-PM₁ et l'eBC de l'AE33 doit être comparée aux mesures continues des PM₁ ou PM_{2.5} obtenues par FIDAS, TEOM-FDMS ou autre. La pente doit être linéaire et proche de 1 avec une erreur inférieure à ±20%. Cette corrélation entre les deux mesures renseigne sur la justesse des concentrations absolues mesurées par l'ACSM dont les valeurs dépendent linéairement de l'efficacité d'ionisation (IE) et du coefficient de collection (CE).

Il est admis que les PM₁ reconstitués (NR-PM₁ + eBC) peuvent être légèrement supérieures aux mesures de PM_{2.5}, principalement en raison des incertitudes de mesure respectives de chaque technique/instrument.

2.3.7 Vérification et correction de la date ACSM Local Time

L'ACSM Local Time peut ne pas être correcte. En effet, ce paramètre est modifié uniquement lorsque le Prisma est arrêté/redémarré après le changement d'heure. Elle doit donc être vérifiée et corrigée avant l'export PMF car elle conditionne les profils journaliers calculés des espèces et des fragments organiques.

- Changement d'heure au printemps : dernier dimanche de mars ; à 2 h du matin, il est 3 h.
- Changement d'heure à l'automne : dernier dimanche d'octobre ; à 3 h du matin, il est 2 h.

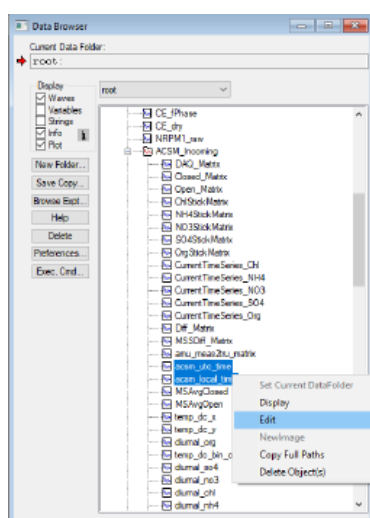


Figure 1 : Vérification de l'ACSM_LocalTime

2.4 Export PMF

Après la validation des espèces chimiques, vient l'étape de l'export PMF. Elle peut être réalisée dans l'onglet « Export » du logiciel.

Il y a trois étapes à suivre. D'abord, initialisation des matrices « Initialize Matrices », puis vérification des incertitudes calculées « Check Error Sanity » et enfin l'export des matrices « Export Matrices ».

Le « downweight » peut se faire dans SoFi Pro.

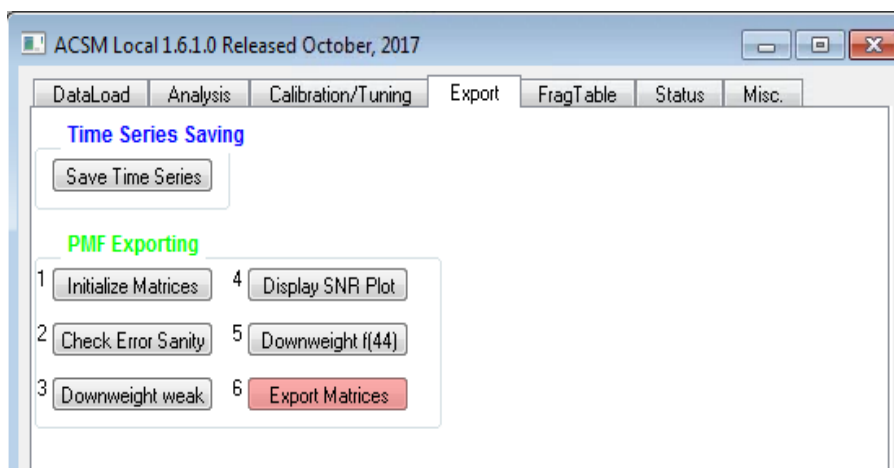


Figure 2 : Panel « Export »

La vérification des ions constitutifs de l'AO doit être faite en examinant tous les fragments m/z afin de pouvoir détecter d'éventuels erreurs liées au calcul des variables « minError », « DwellTime », et « no of sets ».

2.4.1 Possible problème lié à la « minError »

Des incertitudes très faibles sont observées pour certains m/z, comme le montre la figure 3 :

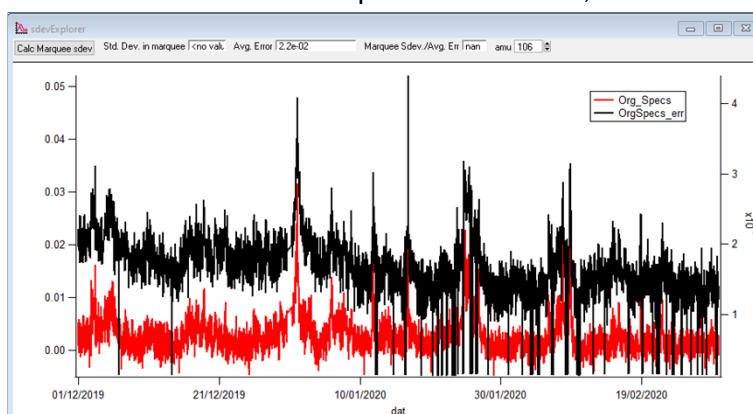


Figure 3 : Exemple d'incertitudes très faibles

Les erreurs correspondant au signal « Closed » pour certains points n'ont pas été calculées car elles étaient négatives (inférieures à la baseline MS), ce qui a donné lieu à des NaN. Les NaN sont remplacés par des zéros lors du processus de génération de la matrice d'erreur, puis remplacés par les valeurs minError. La wave minError semble beaucoup trop faible (facteur 10) par rapport aux incertitudes des autres points de données.

$$e_{Closed} = 1.2 \sqrt{closed} \text{ and } e_{Open} = 1.2 \sqrt{Open}$$

$$e_{DiffMat} = \sqrt{e_{Open}^2 + e_{Closed}^2}$$

Dans certaines versions d'ACSM Local, cela peut être résolu en changeant la variable « ka_baseline » à 1 dans la procédure ACSM Local, ou en appliquant la fonction « ACSM_ReprocessAll() », mais ce n'est pas toujours le cas. Cela peut corriger certaines erreurs, mais pas toutes.

D'autre part, dans ACSM local, la minError est déterminée comme suit :

```

minErrorW = 1.2*sqrt(2*GainW[p]/(6.24e18*DwellITW[p]))
GainW = (gain / smCorr_w[p])*(daq_matrix[p][85]/daq_matrix[p][1])
DwellITW = 2*ka_amu_window * daq_matrix[p][54] * 0.001*daq_matrix[p][75]

```

Par conséquent, l'autre possibilité consiste à calculer ces erreurs et/ou minError différemment. Dans la figure ci-dessous, nous modifions la procédure ACSM local afin de calculer l'erreur en tenant compte de la valeur absolue de ces signaux. Cela résout le problème des faibles incertitudes.

```

acsm_local_1616.ipf

//Calculate difference and its error in cts
//MatrixOp/O eOpenMatCts = a*powr(OpenMatCts,0.5)
//MatrixOp/O eClosedMatCts = a*powr(ClosedMatCts,0.5)
//////////Test for negative values in open and closed signals
MatrixOp/O eOpenMatCts = a*powr(abs(OpenMatCts),0.5)
MatrixOp/O eClosedMatCts = a*powr(abs(ClosedMatCts),0.5)

```

Figure 4 : Procédure ACSM Local

2.4.2 Possible problème lié à un « DwellTime » égal à zéro

Le DwellTime est parfois nul pour certaines périodes :

Table15:DAQ_Matrix				
R0 C54	28			
Row	DAQ_Matrix[[74]	DAQ_Matrix[[75]	DAQ_Matrix[[76]	
	74	75	76	
14885	140204	200	3.485	
14886	140204	200	3.46	
14887	140204	200	3.43	
14888	140204	0	3.475	
14889	140204	0	3.475	
14890	140204	0	3.465	
14891	140204	0	3.47	
14892	140204	0	3.465	
14893	140204	0	3.46	
14894	140204	0	3.47	
14895	140204	0	3.47	
14896	140204	0	3.47	
14897	140204	0	3.455	
14898	140204	0	3.465	
14899	140204	0	3.47	
14900	140204	0	3.465	
14901	140204	0	3.475	
14902	140204	0	3.44	
14903	140204	0	3.47	
14904	140204	0	3.475	
14905	140204	0	3.465	
14906	140204	0	3.465	
14907	140204	0	3.475	
14908	140204	0	3.47	

Figure 5 : Waves de la matrice DAQ

Ce problème correspond à un bug de la procédure ACSM Local, et il peut être résolu manuellement en corrigeant les valeurs nulles par 200.

2.4.3 Problème lié au nb. of sets

L'utilisateur doit vérifier la wave « nb. of sets » car parfois, cette wave n'est pas toujours égale à 28. Cela entraîne donc des incertitudes élevées. L'utilisateur doit donc supprimer les périodes où le nombre d'ensembles est différent de 28, comme le montre cette figure.

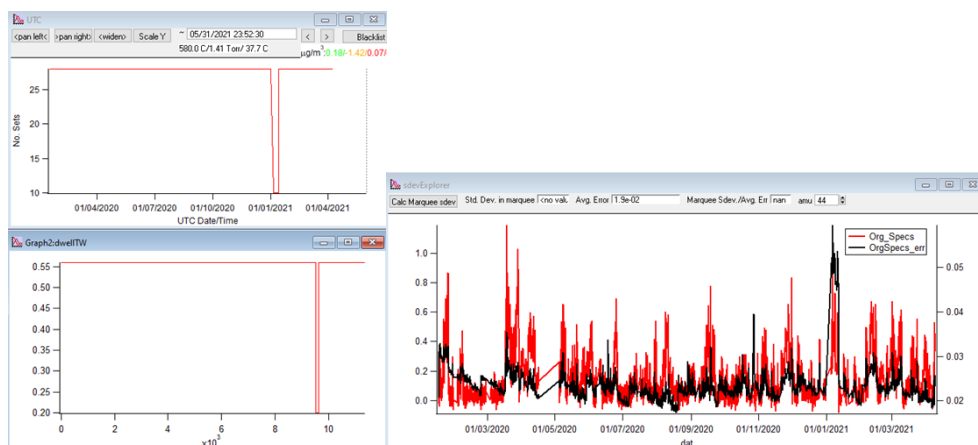


Figure 6 : Exemple de problème lié au nb. of sets

L'export des matrices peut être réalisé une fois ces vérifications effectuées.

Ces matrices doivent ensuite être corrigée par la correction de Middelbrook en utilisant CE_dry, comme suit :

Correction CE

Avant d'exécuter l'analyse PMF, l'utilisateur doit importer la wave CDCE (CE dépendant du temps) au sein de l'expérience Igor, puis diviser les matrices Org et Error par le CE (ou CDCE) manuellement avant d'exécuter PMF. Pour ce faire, tapez la commande suivante (ici, avec CE=1 dans le Review batch) :

- Make /O/N = (DimSize(Org_Specs,0),DimSize(Org_Specs,1))
Org_Specs_corrected=Org_Specs/CE for CE Or
Org_Specs_corrected=Org_Specs [p][q]/CDCE[p] for CDCE
- Make /O/N = (DimSize(Org_Specs,0),DimSize(Org_Specs,1))
OrgSpecs_Err_corrected=OrgSpecs_Err/CE for CE Or
OrgSpecs_Err_corrected=OrgSpecs_Err [p][q]/CDCE[p] for CDCE

Les matrices organiques suivante exportées peuvent enfin être intégrées dans SoFi Pro pour appliquer la PMF :

- Org_Specs => concentrations massiques (en $\mu\text{g m}^{-3}$) des fragments organiques
- OrgSpecs_err => erreurs (en $\mu\text{g m}^{-3}$) associés avec ces concentrations
- amus => liste des m/z (jusqu'au 100 par exemple pour Q-ACSM)
- acsm_local_time => date locale
- acsm_utc_time => date UTC

3. Application de la PMF sur SoFi pro

3.1 Installation et initialisation de SoFi Pro

L'installation du logiciel Pro n'est pas compliquée mais demande l'acquisition de deux licences commerciales : la licence du solveur ME2_engine (valable pour toujours) et la licence SoFi Pro (Datalystica, à durées déterminées).

Le téléchargement des fichiers SoFi peut se faire depuis le site internet de Datalystica (https://www.datalystica.com/userdata/filemanager/data/ME2_engine.zip). Ces étapes sont expliquées dans le manuel d'utilisation d SoFi.

Le logiciel est installé sur Igor Pro, les versions recommandées à utiliser sont SoFi 8 ou 9 sur Igor 8 ou 9. Après son installation, il faut l'initialiser et ouvrir le logiciel.

Les différents onglets de SoFi sont synchronisés avec Data Browser.

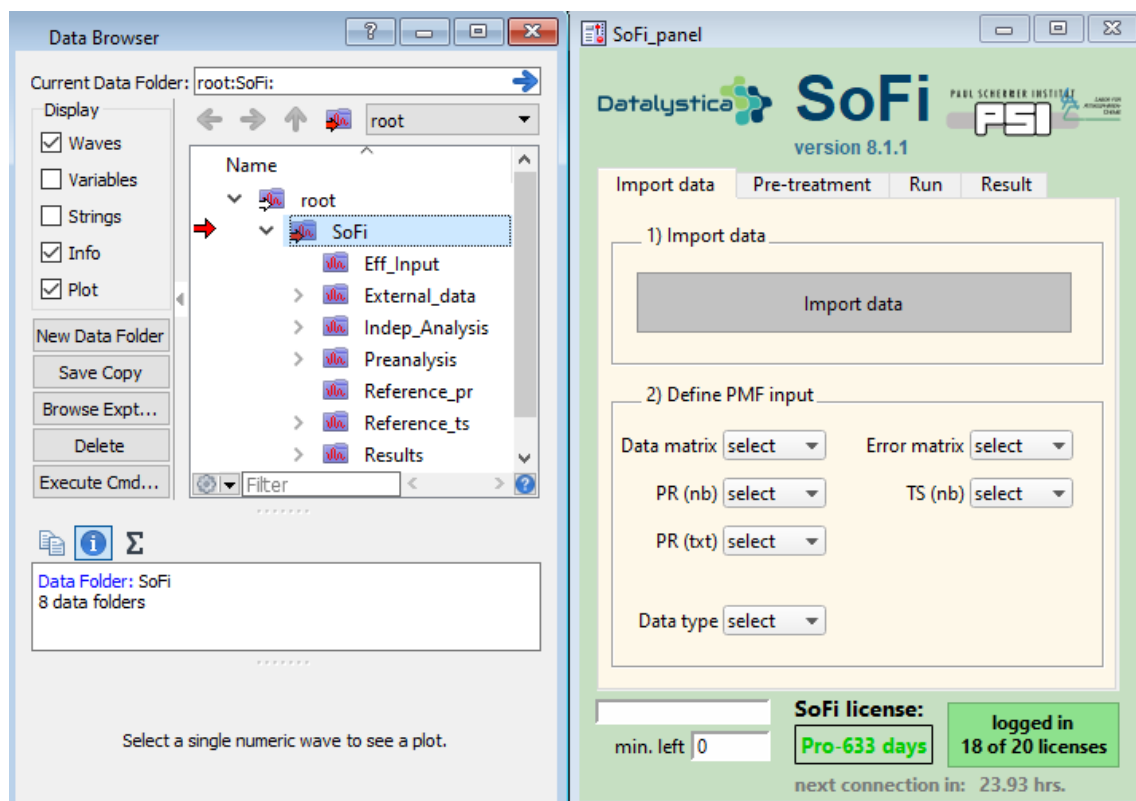


Figure 7 : Panneau SoFi et Data Browser

3.2 Import des données et pré-traitement

3.2.1 Import des données

Dans cette première étape, les données entrées de la PMF peuvent être importées dans différents formats de fichiers (.xls, .csv, .dat, .itx, etc.). L'export effectué avec ACSM Local est en format igor (.itx) qui sera défini par la suite. Ensuite, le chemin du fichier doit être défini ainsi que l'unité d'entrée PMF qui est par défaut $\mu\text{g m}^{-3}$.

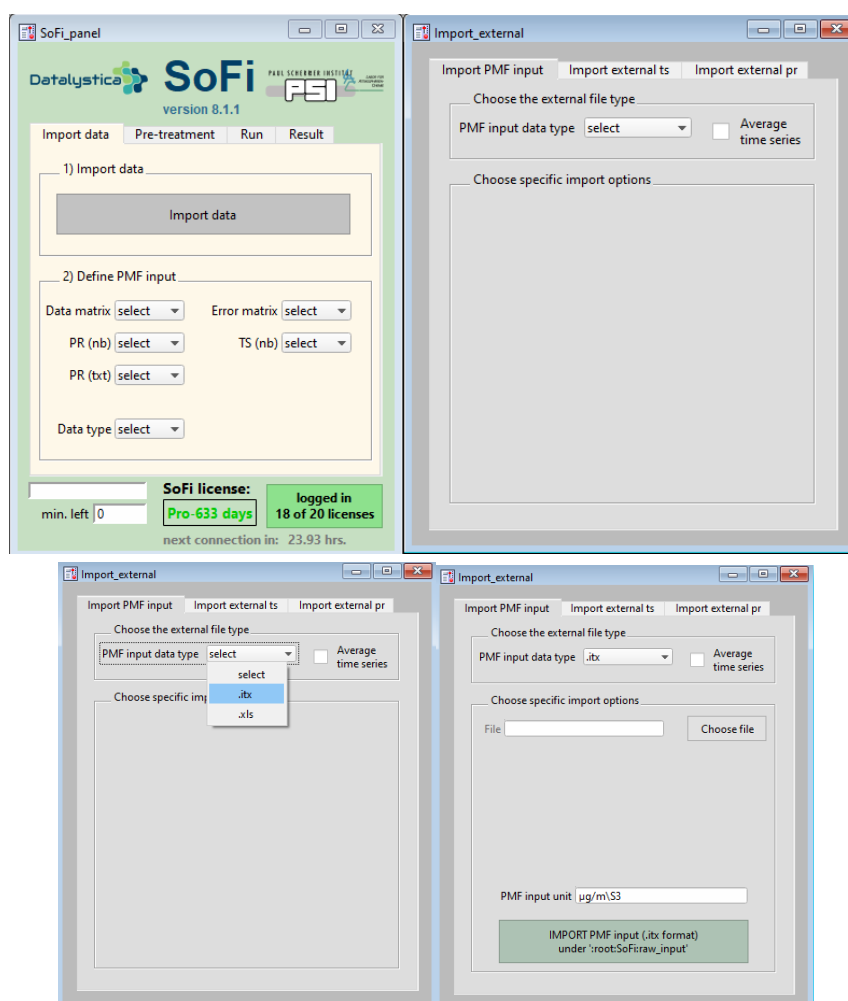


Figure 8 : Import des données d'entrées PMF

Il est également possible d'importer des séries temporelles externes ; cette étape peut être réalisée à tout moment de l'analyse. Il s'agit généralement de toutes les données externes disponibles sur le site, c'est-à-dire les espèces inorganiques (SO_4 , NO_3 , NH_4 et Cl), les concentrations de $\text{PM}_{2.5}$, les données sur le carbone suie BC, les données sur les NO_x ou les données météorologiques, afin d'étayer l'analyse PMF. Le fichier peut être dans n'importe quel format, et SoFi permet d'utiliser n'importe quel format. Ainsi, l'utilisateur choisit le format de fichier (.itx, .xls, .csv, .dat, .txt, ou même les données AE33), puis définit le chemin d'accès au fichier, et enfin clique sur « Importer des fichiers ».

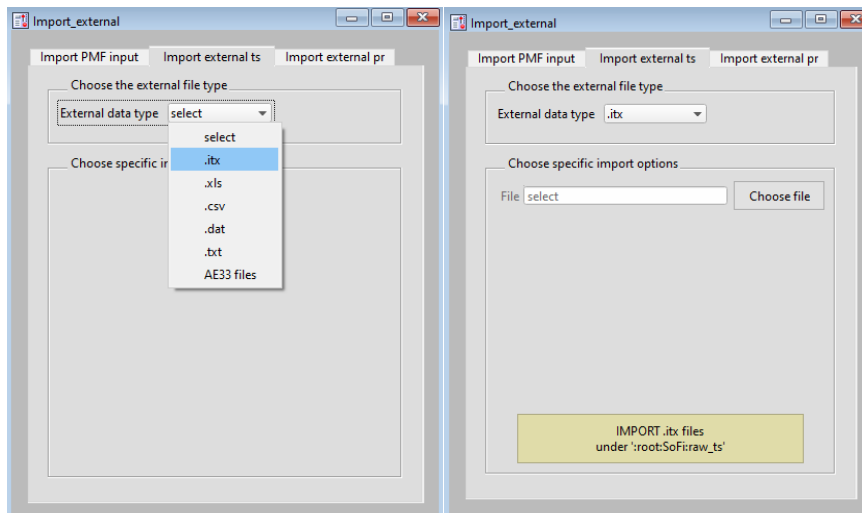


Figure 9 : Import des données externes

Il peut également être intéressant d'importer des profils externes (en tant que spectres de masse de référence), que ce soit à des fins de comparaison ou de contrainte. Une base de données de profils externes est disponible à l'adresse <http://cires1.colorado.edu/jimenez-group/AMSsd/>. Là encore, l'utilisateur doit sélectionner le type de fichier, puis choisir son chemin d'accès et cliquer sur « Importer des fichiers ».

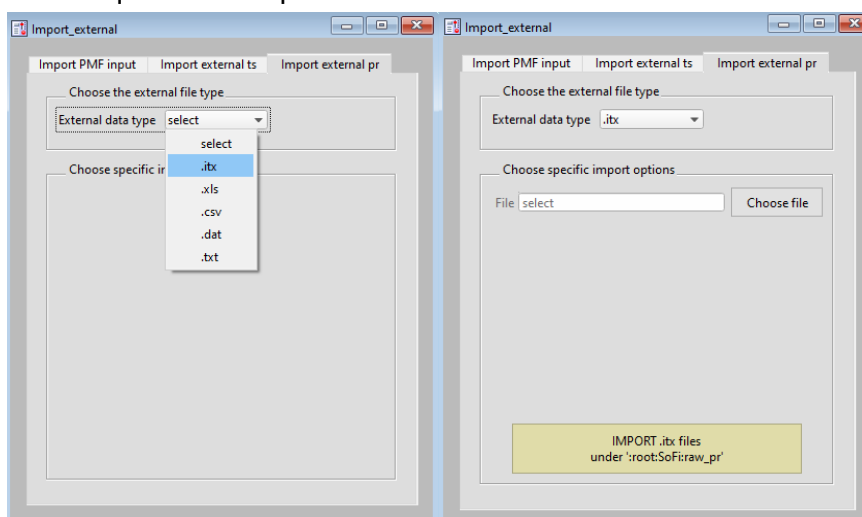


Figure 10: Import des profils externes

Les données importées sont stockées sous le chemin root : External_data dans « Data_Browser ».

3.2.2 Définition des données d'entrée

A la suite de l'import des données, vient l'étape de définir les données d'entrée :

Figure 11: Définir les inputs PMF

- **Data matrix** est la matrice organique 'concentrations' (e.g., Org_Specs);
- **Error matrix** est la matrice organique 'erreurs' (e.g., OrgSpecs_err), si les données PMF sont importées à partir d'un fichier Excel, SoFi peut calculer la matrice d'erreurs sur la base de deux projets différents (Polissar et al. 1998) ou (Norris et al. 2014);
- **PR (nb)**: PR est les profils (e.g., amus), Ils présentent les variables (i.e., les fragments m/z pour ACSM), qui sont ici importées au format numérique. SoFi demande ensuite à l'utilisateur s'il dispose du fichier au format texte. Si ce n'est pas le cas, SoFi le crée automatiquement ;
- **PR (txt)**: Il s'agit des informations en texte relatives aux variables. Ces informations sont importantes pour définir les étiquettes des graphiques ;
- **TS (nb)**: TS est la série temporelle des données importées (e.g., acsm_local_time ou acsm_utc_time) ;
- **Data type**: plusieurs options de type de données (AMS-UMR/Q-ACSM, Tof-ACSM, HR-AMS, données filtrées ou autres), il s'agit essentiellement de l'endroit où est définie, dans le cas des instruments Aerodyne, la table de fragmentation ; dans ce cas, l'utilisateur doit sélectionner Q-ACSM ;
- Ensuite '**AMS specific**' apparaît et l'utilisateur dispose de deux options : « prendre en compte toutes les variables » ou « exclure les variables liées au CO2 ». En fait, dans les instruments Aerodyne, certaines variables sont liées au CO2 (m/z 44) et ne sont pas directement mesurées, elles sont calculées dans la table de fragmentation à partir du CO2. Ainsi, si l'utilisateur prend en compte toutes les variables, le m/z 44 est surestimé. Il est donc plus intéressant de sélectionner l'option « exclure les variables liées au CO2 ». Généralement, elles sont exclues du calcul du PMF mais elles seront réinsérées dans l'étape des résultats.

3.2.3 Pré-traitement des données

Une fois l'import terminé, le pré-traitement des données inclut les étapes suivantes.

3.2.3.1 Définir le symbole pour les données manquantes

L'utilisateur définit un symbole pour les données manquantes. Ce symbole est inséré dans la matrice de données s'il n'y a pas de mesure pour ce point. SoFi vérifie simplement toutes les lignes et colonnes qui ne contiennent aucune information afin de les remplir avec ce symbole. Dans le cas de Q-ACSM, l'utilisateur utilise NaN (ou 0 pour les données AMS par exemple), puis clique sur « Continue ».

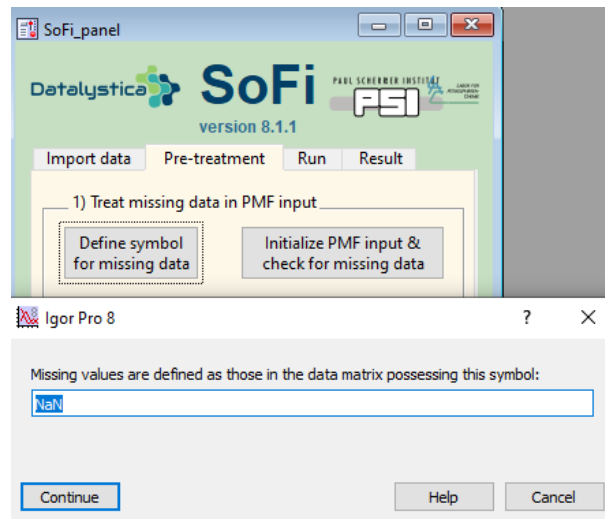


Figure 12 : Traitement des données manquantes dans les données d'entrée PMF

3.2.3.2 Initialiser l'entrée PMF et vérifier les données manquantes

Une fois les données manquantes définies, l'utilisateur doit cliquer sur « initialize PMF input & check for missing data », afin que les colonnes et les lignes entières soient supprimées. Lorsque l'initialisation est terminée, l'entrée PMF se trouve dans le dossier « Eff_Input » du Data Browser.

3.2.3.3 Traitement des données externes

Dans la deuxième étape, l'utilisateur peut initialiser les données externes afin de les faire correspondre au format des matrices d'entrée de l'analyse PMF. L'utilisateur doit ainsi cliquer sur « Treat external data for PMF run comparison ». Il existe deux options pour le bouton « External data » :

Profils

L'axe des x est l'information texte texte profil externe (par exemple, number_text) et le logiciel SoFi compare l'information textuelle de l'entrée PMF (PR (text)) avec celle du profil externe, puis l'utilisateur sélectionne les profils et clique sur le bouton « Extract information ». Les profils externes seront placés dans le dossier « reference_pr » dans le panneau du Data Browser.

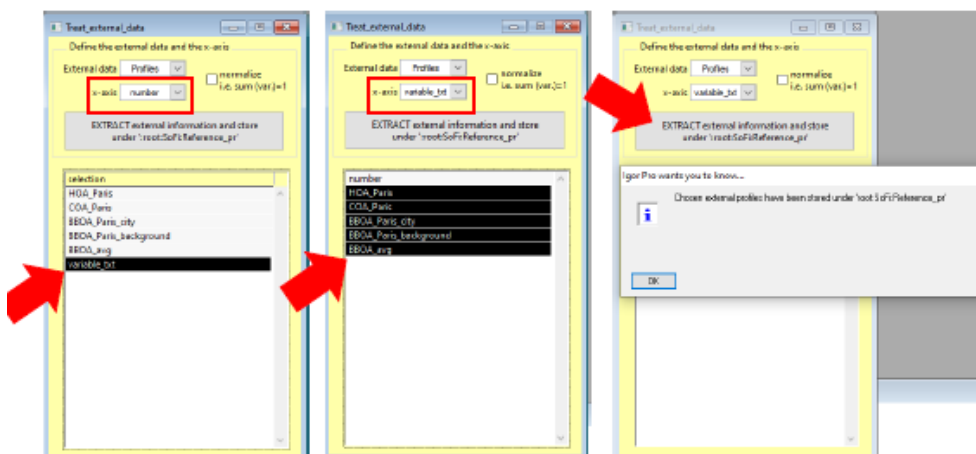


Figure 13 : Treat external data 'profiles'

Séries temporelles

Pour les séries temporelles : l'utilisateur sélectionne l'axe des abscisses qui est la série temporelle des données externes, puis il sélectionne les variables à extraire, et enfin l'option du « type » d'extraction. L'utilisateur peut choisir le type : « EXTRACT » si les données externes ont les mêmes points temporels que l'entrée PMF (par exemple, les espèces ACSM). L'option « INTERPOLATE », si la résolution temporelle des données externes est similaire mais pas égale. Et l'option « MOYENNE » (save start, middle, or end time) si la résolution temporelle des données externes est plus élevée. La différence entre les 3 types d'option « Moyenne » est le temps de sauvegarde des données : le ToF-ACSM, par exemple, crée le fichier avec un horodatage correspondant au début de la période de mesure, alors que le Q-ACSM sauvegarde le fichier de données élémentaire en l'indexant sur la fin de la période de mesure. L'utilisateur choisit ensuite le type de données à considérer et peut cliquer sur « REGRID the data information » afin d'être sauvegardé dans le dossier « reference_ts ». Cette partie générera des graphiques pour ces données avec différentes résolutions afin d'avoir une idée de la distribution de ces données externes.

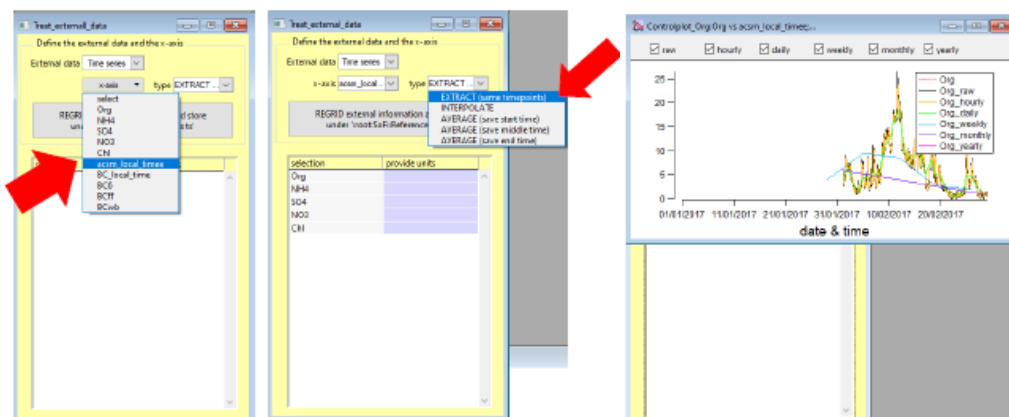


Figure 14 : Treat external data 'Time Series'

3.2.3.4 Explorer les données d'entrée de PMF

Après avoir initialisé les données d'entrée et avant d'exécuter PMF, il est important d'explorer les données pour mieux les comprendre et avoir une idée préliminaire de leurs principales sources. L'utilisateur doit donc cliquer sur « Explore trend of variables ». Cette partie permet d'obtenir des informations intéressantes sur les données. Cela peut se faire en vérifiant par exemple la variation diurne des masses principales (càd m/z 43, 44, 55, 57, 60, 73), en comparant avec les données externes (par exemple, en établissant des corrélations) ou en examinant f44/43 pour voir s'il y a des tendances particulières ou des valeurs aberrantes.

Il est également possible d'ajouter à la fois des données mz et des données externes pour comparer un mz spécifique avec un traceur de source (mz 60 vs BCwb par exemple). Il existe différentes options pour l'ajustement et la représentation graphique dans ce panneau, comme indiqué dans la figure ci-dessous.

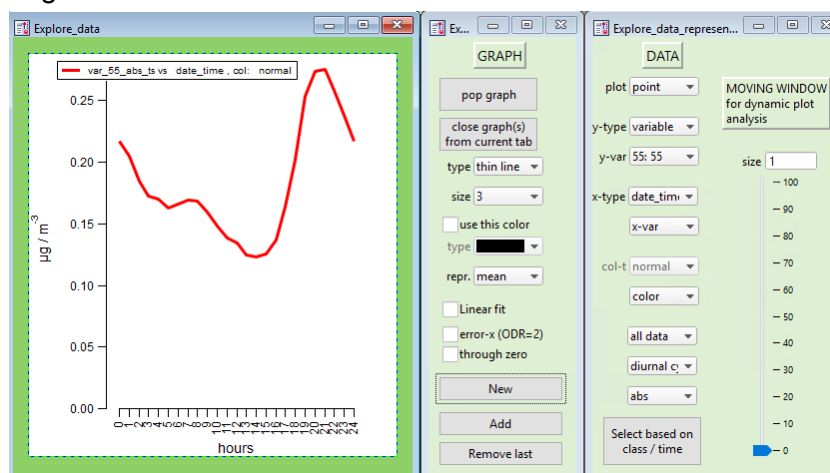


Figure 15 : Explore PMF input data ; exemple de profil journalier de m/z 55

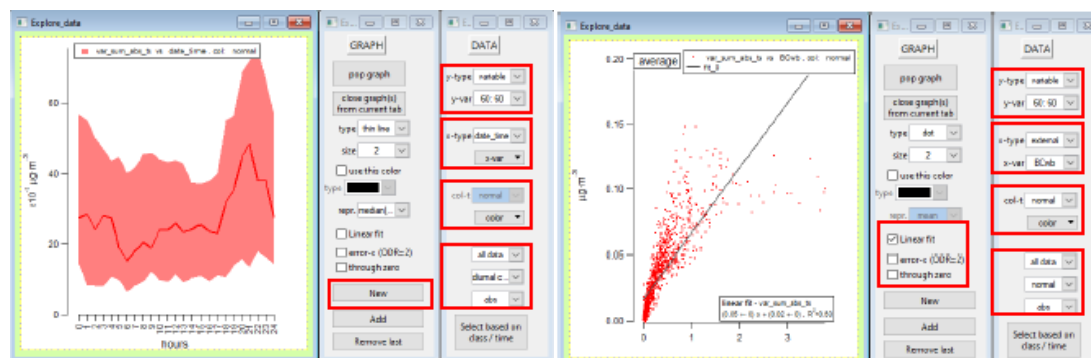


Figure 16 : Exemples pour 'Explore PMF input data'

3.2.3.5 Préparer l'entrée PMF pour l'exécution

Dans cette étape, l'utilisateur peut sélectionner et/ou modifier les données d'entrée prises en compte pour l'analyse PMF. Dans ce panneau, l'utilisateur a la possibilité de supprimer certains éléments et de sélectionner la période et les variables à traiter.

Profils

Dans les profils entrés, l'utilisateur peut vérifier les spectres de masse des profils (e.g., les m/z négatives, le nombre de m/z à analyser). Ainsi, dans le panneau de droite, l'utilisateur peut sélectionner les variables qu'il souhaite conserver pour le PMF, de sorte que les variables indésirables sont désélectionnées, puis l'utilisateur clique sur « APPLY SELECTION » pour les blacklister. Par exemple, le m/z 12 est supprimé dans de nombreux cas en raison de son signal négatif.

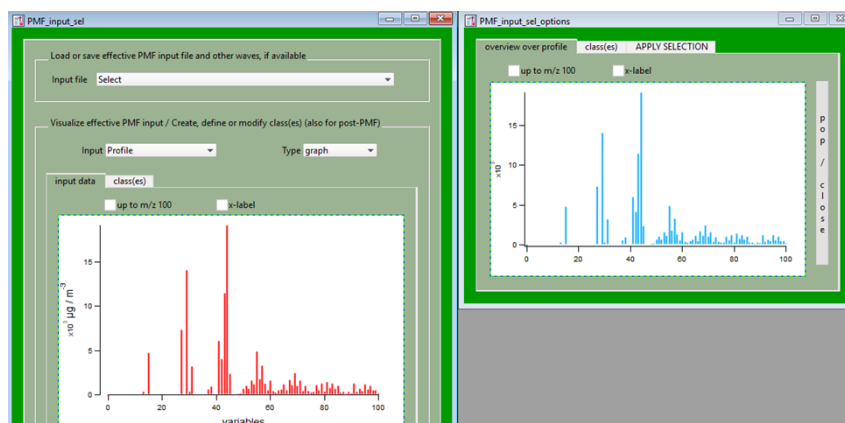


Figure 17 : Sélection des données d'entrées

Il est également possible de créer et d'ajouter des classes en cliquant sur le panneau « classe(s) ». L'utilisateur peut créer ou ajouter des classes en tant que métriques supplémentaires. C'est utile, en particulier dans le cas où l'on dispose de deux types de données, qu'il est possible de séparer en utilisant des classes.

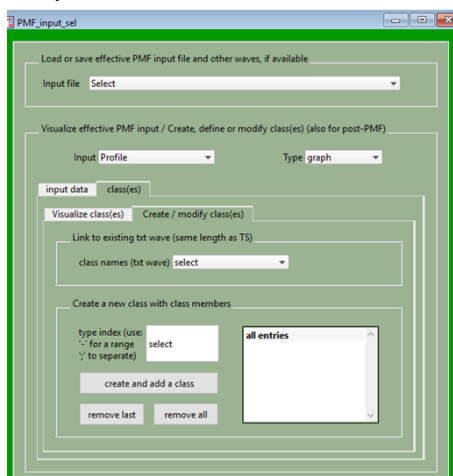


Figure 18 : PMF input panel to control user-specified PMF

Séries temporelles

Idem pour l'entrée « séries temporelles » : l'utilisateur peut sélectionner la période à analyser à l'aide du panneau « PMF_input_sel_options ». Ceci peut être fait en cochant les mois (ou jours, heures, années) dans l'onglet « time-selection » ou en faisant la sélection dans l'onglet « overview over time ». Ensuite, l'utilisateur peut blacklister la partie non sélectionnée en cliquant sur l'onglet « APPLY SELECTION ». Par exemple, l'utilisateur peut faire une sélection basée sur différentes saisons afin d'effectuer le PMF saisonnier.

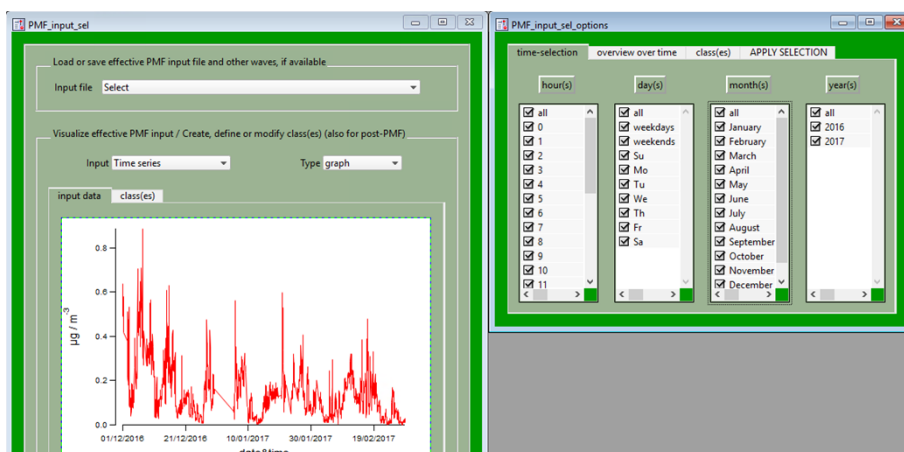


Figure 19 : Panneau de saisie du PMF pour préparer l'exécution

L'utilisateur peut également créer et ajouter des classes (par exemple, créer des classes distinctes de données hors ligne en sélectionnant automatiquement leurs fichiers, ou le faire dans le temps de l'index en créant une classe de l'index p à q).

3.3 Exécution de la PMF

Ce chapitre s'applique à tous les types de PMF.

Après avoir défini les données d'entrée (variables et séries temporelles), l'utilisateur peut passer au 3ème onglet pour exécuter l'analyse PMF et définir ses différents paramètres.

A noter d'ores et déjà que l'utilisateur peut commencer avec une analyse PMF sans contrainte afin d'identifier le nombre de sources possibles d'aérosols organiques (OA) pour le site étudié.

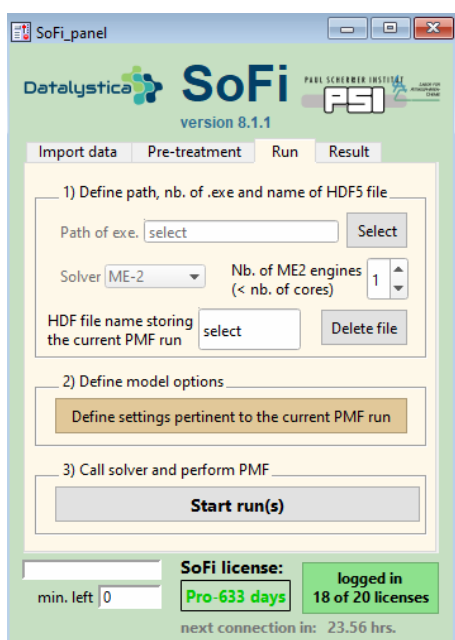


Figure 20 : Onglet PMF run

3.3.1 Définir le chemin et le nom du fichier HDF5

Tout d'abord, l'utilisateur doit définir le chemin du fichier ME2_engine en cliquant sur le bouton « Folder of executable » et en sélectionnant le fichier, puis définir le nom du fichier HDF en écrivant le nom dans la case et en cliquant sur « Enter ». Une fois le fichier HDF créé, tous les résultats seront stockés dans ce fichier HDF.

Ensuite, l'utilisateur sélectionne le nombre de cœurs des processeurs de l'ordinateur à utiliser pour exécuter la PMF. Ce nombre doit évidemment être inférieur au nombre total de cœurs de l'ordinateur. Avec SoFi Pro, il est possible d'utiliser tous les cœurs, mais il est préférable de ne pas le faire car cela peut bloquer d'autres opérations de l'ordinateur (par exemple, s'il y a 4 cœurs dans l'ordinateur le nombre d'exécutables pour l'exécution de l'analyse PMF).

Ensuite, vous devez définir le nom de votre fichier de résultats dans la section « HDF file name storing the current PMF run ». **Attention** : il faut donner un nom spécifique et unique, car si vous utilisez le même nom qu'un fichier existant, vous n'obtiendrez pas de message d'erreur, mais les nouveaux résultats l'écraseront. Faire également attention à la longueur du nom donné au fichier (15 caractères maximum), car vous risquez d'obtenir un message d'erreur.

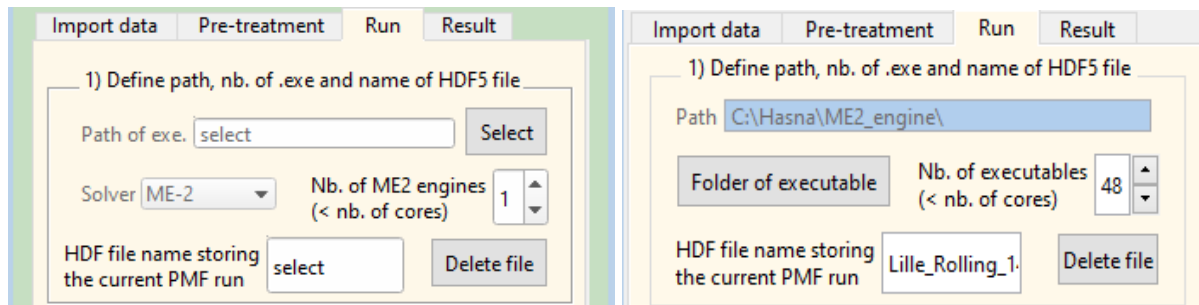


Figure 21 : Définition du chemin et du nom du fichier HDF

3.3.2 Définir les options du modèle

Après avoir défini le fichier HDF pour les résultats PMF, l'utilisateur doit définir les différents paramètres nécessaires à l'exécution du modèle PMF. Cette partie est considérée comme l'une des plus importantes de l'analyse PMF. Elle contient 4 panneaux : le premier contient des paramètres généraux et les autres dépendent de l'analyse que l'utilisateur souhaite effectuer (PMF sans contrainte, PMF avec contrainte, analyse Bootstrap ou rolling PMF). Le rolling PMF par exemple, nécessite de passer par toutes ces étapes en définissant le nombre de facteurs dans le premier panneau, en appliquant des contraintes dans le deuxième panneau, en utilisant le bootstrap dans le troisième panneau et en définissant les paramètres du rolling PMF dans le dernier panneau.

Ainsi, pour l'analyse PMF sans contrainte, l'utilisateur doit définir les paramètres généraux de l'analyse PMF.

3.3.2.1 Paramètres généraux

Dans cette partie, l'utilisateur doit définir les informations générales nécessaires à l'exécution de l'analyse PMF.

PMF_options

GENERAL settings | Model MATRICES | RESAMPLING strategy | ROLLING mechanism

I. Define number of factors

starting factor: 3 | ending factor: 7 | nb. of PMF calls: 10

II. Convergence criteria

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				

III. Missing data, Error weights and miscellaneous

Missing data | Error matrix (general) | Error matrix (S/N weight) | Others

Model type to weight the error of the data: step function | Data type for S2N-based weight: averaged

threshold: | weight: |

weak: 2 | weak: 2 | apply weight factor | inspect signal to noise plot

bad: 0.2 | bad: 10 | undo weight factor | check for weighted data

Figure 22 : paramètres PMF

Définir le nombre de facteurs

L'utilisateur doit définir le nombre de facteurs en spécifiant le nombre de facteurs de départ et de fin, ainsi que le nombre de runs PMF présentant le nombre d'itérations. Par exemple, un facteur de départ avec 3 facteurs et un facteur final avec 8 facteurs pour 10 runs PMF pour une première analyse PMF sans contrainte.

En général, le choix du nombre de facteurs dépend du nombre de sources possibles pour le site.

Critères de convergence

Cette partie des critères de convergence est importante pour contrôler la qualité de l'ajustement lors de l'analyse. Elle est indispensable pour atteindre la convergence. Généralement, l'utilisateur n'a pas besoin de modifier les critères proposés par défaut.

3.3.2.2 Paramètres divers

Cette partie définit de nombreux paramètres concernant les données manquantes, la matrice d'erreur, etc.

Données manquantes

Cette partie contient d'autres paramètres de données manquantes, notamment pour les cellules vides. L'utilisateur ne modifie rien dans cette partie. Le bouton « Define values for matrix cells containing no data » permet de définir les valeurs manquantes. Dans ce cas, il s'agit de NaN. Tous ces NaN seront remplacés par une valeur spécifique, soit -100, et les erreurs correspondantes seront remplacées par l'erreur la plus élevée.

Dans la deuxième étape, une valeur seuil (par exemple -99) signifie que tout ce qui est inférieur à -99 ne doit pas être pris en compte. C'est pourquoi les cellules manquantes ou sans données sont remplacées par une valeur inférieure à -99.

Matrice d'erreur (général)

Cette partie correspond à l'approche par C-Value. Dans de nombreux cas, cette option est désactivée. L'utilisateur peut l'activer si les données proviennent de différents instruments. Il est alors nécessaire de s'assurer que l'erreur des deux instruments est rééchelonnée pour obtenir une PMF efficace. En d'autres termes, cette option est activée en cas d'application d'une PMF combinée avec deux types de données différents.

Matrice d'erreur (pondération)

Dans cette partie, l'utilisateur peut sous-pondérer certaines variables (en particulier, certaines m/z). Il doit d'abord sélectionner le « model type to weight the error of the data » et le « Data type for S2N-based weight ». Deux options sont disponibles : « step function » et « 1/S2N function ». Il est recommandé de choisir la « step function » afin de définir le seuil et la pondération des valeurs faibles et erronées, puis de choisir l'option « moyenne » pour le type de données de la boîte S2N.

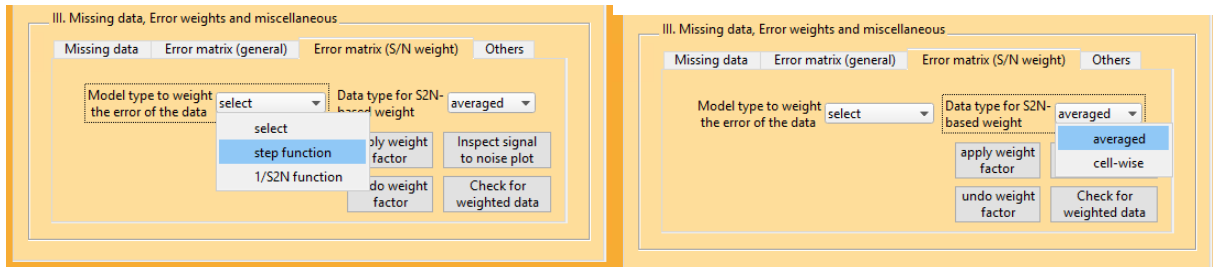


Figure 23 : Onglet « Matrice d'erreurs (pondération) »

La définition des variables de types « weak » and « bad » est effectuée selon les recommandations de (Paatero et Hopke 2003) et (Ulbrich et al. 2009) ; il s'agit de variables faibles avec un rapport signal/bruit de $0,2 < (S/N) < 2$, multiplié par 2, et de variables mauvaises avec un $(S/N) < 0,2$, multiplié par 10.

L'utilisateur peut donc modifier le rapport S2N. Pour les variables dont le S2N est inférieur à 2, la pondération est réduite d'un facteur 2. Si le S2N est inférieur à 0,2, la variable est considérée comme mauvaise et sa pondération est réduite d'un facteur 10.

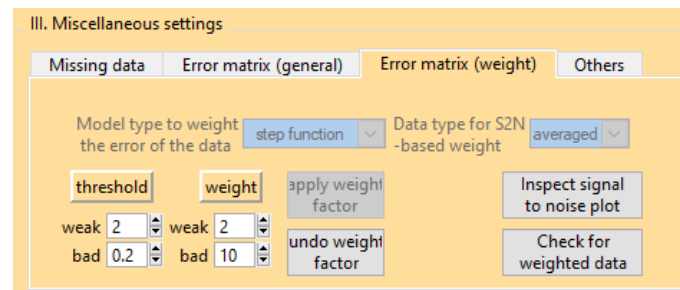


Figure 24 : Paramètres principaux de la matrice d'erreur

L'utilisateur peut ensuite consulter le graphique du S2N, où la moyenne du rapport S2N pour chaque variable est présentée, les variables faibles étant en vert et les variables mauvaises en noir. Ensuite, l'utilisateur clique sur la case « appliquer le facteur de pondération » et vérifie le graphique des données pondérées.

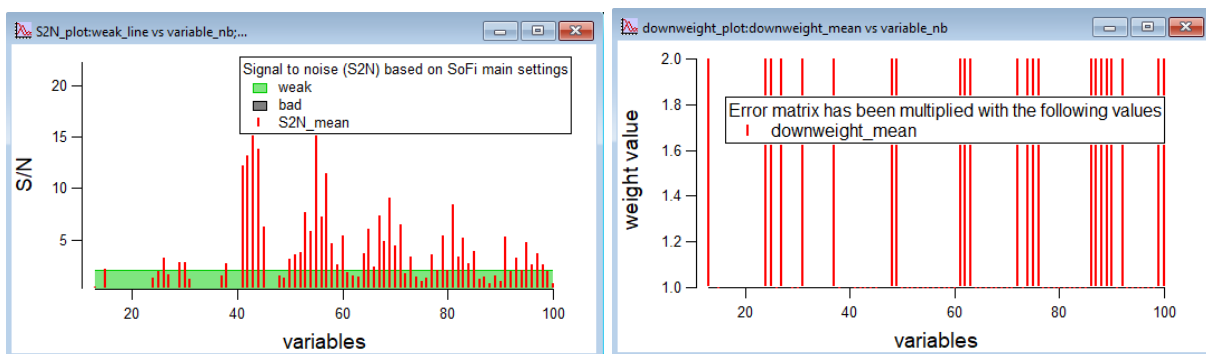


Figure 25 : S2N and weighted errors for the averaged data

3.3.2.3 Autres

Cet onglet regroupe quatre autres paramètres à définir (certaines de ces options n'étant pas modifiables à ce stade de développement du logiciel).

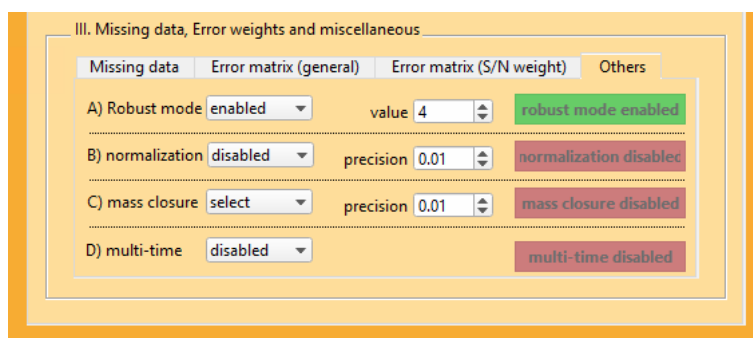


Figure 26 : Onglet « Autres paramètres »

Mode robuste

Le mode robuste est toujours activé. La méthode PMF s'exécute en mode robuste afin d'éviter toute dérive de la PMF due à des valeurs aberrantes ; les valeurs aberrantes importantes sont donc simplement reséquentées à partir de la valeur 4.

Les trois paramètres suivants sont souvent désactivés :

Normalisation

Cette option permet la normalisation lors de l'analyse PMF. Dans le cas par défaut de SoFi, cette normalisation est désactivée ; cela signifie qu'aucune équation de normalisation n'est utilisée pendant l'analyse, car SoFi est généralement utilisé pour contraindre les profils à l'aide de la α -value. Si une équation d'extraction est utilisée pour contraindre les facteurs, cette normalisation doit être activée sur les facteurs, et la valeur 0,01 est considérée comme la pondération de l'équation d'extraction.

Fermeture de masse

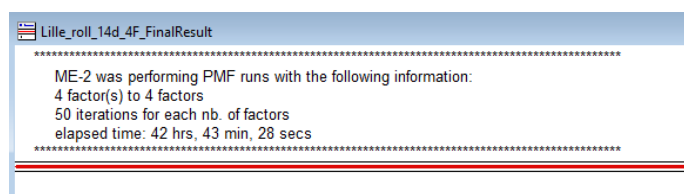
Cette option permet d'ajuster la somme de toutes les contributions des facteurs à une masse donnée (PM1, PM2,5, ...). Cette fermeture de masse est également liée à l'équation d'extraction et à la normalisation.

Multi-time

Cette option est activée lors de la combinaison de données provenant de différents instruments avec différentes résolutions temporelles (voir l'exemple à la fin pour l'utilisation de cette option).

3.3.3 Appeler ME2 et exécuter le PMF

Après avoir défini tous les paramètres principaux, l'utilisateur peut exécuter l'analyse en cliquant sur « Start run(s) ». À la fin de cette PMF, si l'utilisateur obtient le message ci-dessous dans la console (figure ci-dessous), cela signifie que les exécutions du modèle sont terminées. L'utilisateur peut ensuite ouvrir le dossier des résultats et importer les résultats du PMF depuis le fichier HDF.



```
Lille_roll_14d_4F_FinalResult
*****
ME-2 was performing PMF runs with the following information:
4 factor(s) to 4 factors
50 iterations for each nb. of factors
elapsed time: 42 hrs, 43 min, 28 secs
*****
```

Figure 27 : Exécution PMF

3.4 Visualisation des résultats PMF

Ce chapitre s'applique à tous les types de PMF.

L'objectif de cette étape est de sélectionner et d'analyser les principaux résultats obtenus.

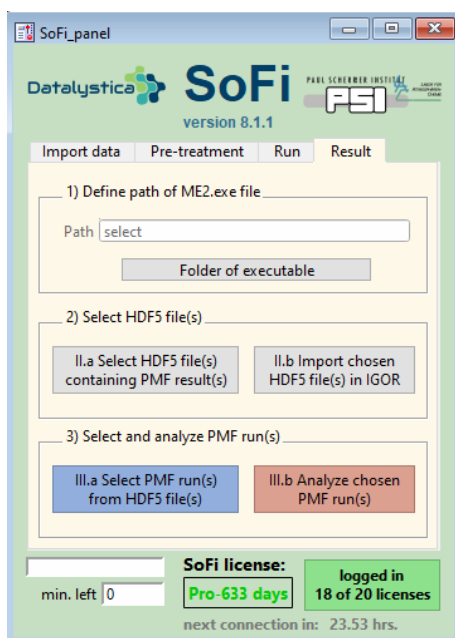


Figure 28 : Panneau de résultats

3.4.1 Définir le chemin du fichier ME2.exe

Tout d'abord, l'utilisateur doit définir le chemin du dossier ME2 (où se trouve ME2.exe). Il s'agit du même chemin défini à l'étape précédente de la section « Exécuter PMF ».

3.4.2 Sélectionner le fichier HDF5

Ensuite, la sélection du fichier HDF contient les résultats. En cliquant sur le bouton « select HDF5 file », l'utilisateur accède à la liste de tous les fichiers HDF présents dans le dossier. Il lui suffit ensuite de sélectionner celui dont il a besoin (ou d'utiliser Ctrl ou Maj pour en sélectionner plusieurs), puis de refermer la boîte de dialogue en cliquant à nouveau sur le bouton correspondant. Cliquez ensuite sur le bouton « Import chosen HDF5 file into IGOR » pour importer le fichier HDF. Cette opération prend un certain temps selon la quantité de données.

3.4.3 Sélection et analyse de l'exécution PMF

L'utilisateur sélectionne ensuite les résultats à analyser. Cette étape se compose de deux étapes importantes :

3.4.3.1 Sélection de l'exécution PMF à partir du ou des fichiers HDF5

Deux options sont disponibles : la sélection manuelle ou automatique, qui conduit à une sélection basée sur des critères statistiques. La sélection automatique est importante, notamment pour sélectionner les meilleures exécutions, différencier les facteurs sans contrainte et obtenir la même position pour chaque facteur, ainsi que pour calculer la solution moyenne.

Sélection manuelle des exécutions PMF

Pour une exécution PMF simple, la sélection manuelle est l'option la plus utilisée pour les exécutions PMF sans contrainte et sans moyenne. L'utilisateur peut ainsi sélectionner le fichier HDF à considérer et visualise les exécutions PMF dans le graphique ci-dessous. Il peut ensuite sélectionner les exécutions à analyser en détail à l'aide d'un cadre. À cette étape, il est préférable de sélectionner toutes les exécutions de PMF afin de vérifier toutes les possibilités de résultats de PMF.

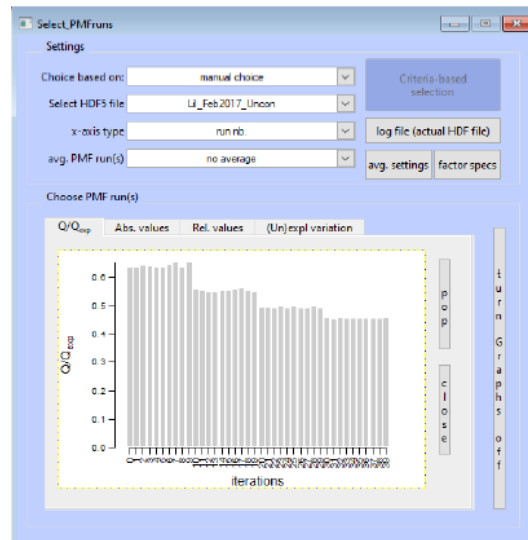


Figure 29 : Sélection de quelques solutions dans le fichier HDF

Le rapport Q/Q_{exp} diminue avec le nombre de facteurs. Les pentes de ce graphique (écarts entre les différents nombres de facteurs) permettent d'identifier le nombre de facteurs de PMF possibles pour ce site.

Pour l'axe des abscisses, plusieurs options sont disponibles (nombre de facteurs, nombre d'exécutions, etc.). La zone des graphiques propose quatre types : l'aperçu des valeurs Q/Q_{exp} , les graphiques des facteurs en termes absolus et relatifs, et la variation inexpliquée (celle-ci devrait normalement être inférieure à 25 %).

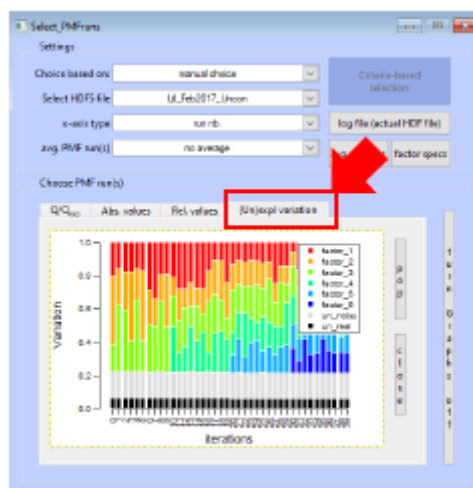


Figure 30 : Graphique de la variation inexpliquée

D'autres zones sont disponibles :

- **Log file (actual HDF file)** indiquant les paramètres choisis ; il s'agit d'un fichier résumant les paramètres de l'expérience igor.
- **Avg. Settings** pour choisir le type de résolution et les données à importer.
- **Factor spec** pour définir les noms des facteurs (par exemple, HOA, BBOA, OOA1, OOA2...) ou ajouter des facteurs comme la somme de OOA1 et OOA2.

3.4.3.2 Analyser les analyses PMF sélectionnées

Une fois les analyses PMF sélectionnées, cliquez sur « Analyze the chosen PMF runs » dans l'encadré rouge et attendez la fin de l'analyse. Vous pouvez ensuite consulter les résultats.

Informations générales

Le premier panneau présente des informations générales sur l'expérience. À gauche, la liste des analyses PMF présente tous les runs précédemment sélectionnés, qu'ils proviennent d'un ou plusieurs fichiers HDF. À droite, des informations moyennes. En dessous se trouve le 'log file' des analyses sélectionnées. L'option « swap factor position » permet de modifier la position des facteurs, mais cette option ne s'applique qu'à l'expérience Igor, et non au fichier HDF.

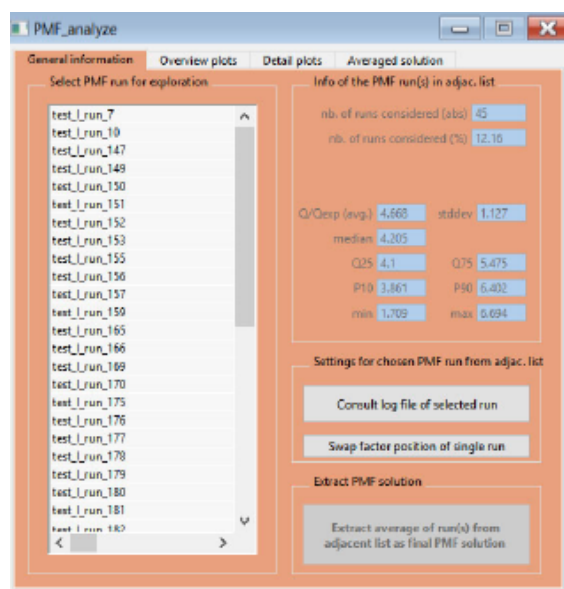


Figure 31 : Onglet « Informations générales »

Après avoir sélectionné les analyses, vous pouvez consulter les différents graphiques pour obtenir un aperçu des analyses PMF. Pour ces résultats PMF sans contrainte, vous devez vérifier chaque analyse afin d'identifier les principaux facteurs d'aérosols organiques du site.

Graphiques d'aperçu 'Overview plots'

Dans ce panneau, l'utilisateur peut consulter les séries temporelles et les spectres de masse des résultats de l'analyse PMF.

Tout d'abord, l'utilisateur peut examiner les profils, en observant le rapport m/z de chaque facteur et en vérifiant leur contribution, afin de déterminer la contribution de chaque variable à chaque facteur. Le graphique ci-dessous illustre le cas des quatre facteurs. Les deux premiers facteurs sont des AO primaires (HOA - Hydrocarbon-like Organic Aerosol, BBOA - Biomass Burning Organic Aerosol), et les deux derniers sont des AO oxygénés, et donc plutôt d'origine secondaire, avec une contribution importante de m/z de 43 et 44. Le facteur HOA est caractérisé par des pics correspondant aux hydrocarbures (m/z 27, 41, 43, 55, 57, 69, 71, etc.), principalement liés au trafic routier, et le facteur BBOA est caractérisé par des fragments de sucres anhydres tels que le lévoglucosan (m/z 29, 60, 73). Le BBOA est fortement influencé par le type de bois, les conditions de combustion, etc., et varie donc d'un site à l'autre.

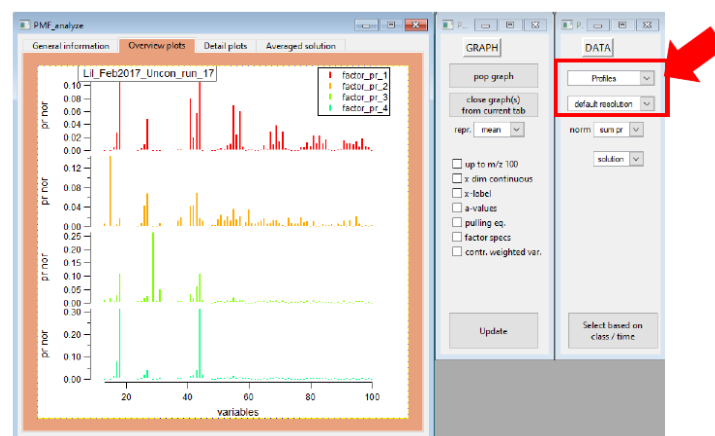


Figure 32 : Onglet « Overview plots »

L'utilisateur peut ensuite sélectionner la série temporelle avec la résolution proposée par défaut et consulter la solution avec la résolution temporelle normale pour tracer la série temporelle des facteurs. Il est également conseillé de vérifier le cycle diurne des facteurs, ainsi que les résidus (abs, scaled, Q_res...). Il est également possible d'effectuer une sélection pour certains mois, week-ends pour avoir une vue d'ensemble, ou pour certaines périodes afin d'observer des éléments spécifiques.

Les principaux facteurs d'OA présentent une tendance diurne prononcée. L'HOA se caractérise par deux pics le matin et le soir, liés aux heures de trafic, et l'BBOA par un pic fort la nuit, correspondant au chauffage résidentiel en période froide. En revanche, les facteurs d'OA oxygénés oxydés (OOA) présentent un cycle journalier plat.

Concernant les résidus (la différence entre les concentrations mesurées et les concentrations reconstruites par le PMF, pondérée par les incertitudes), la structure du cycle journalier des résidus 'scaled' pourrait indiquer des sources manquantes ou mal séparées. De plus, certains pics dans la série temporelle de ce résidu nécessitent encore une attention particulière, suggérant de fortes incertitudes pour ces points ou des problèmes d'instrumentation.

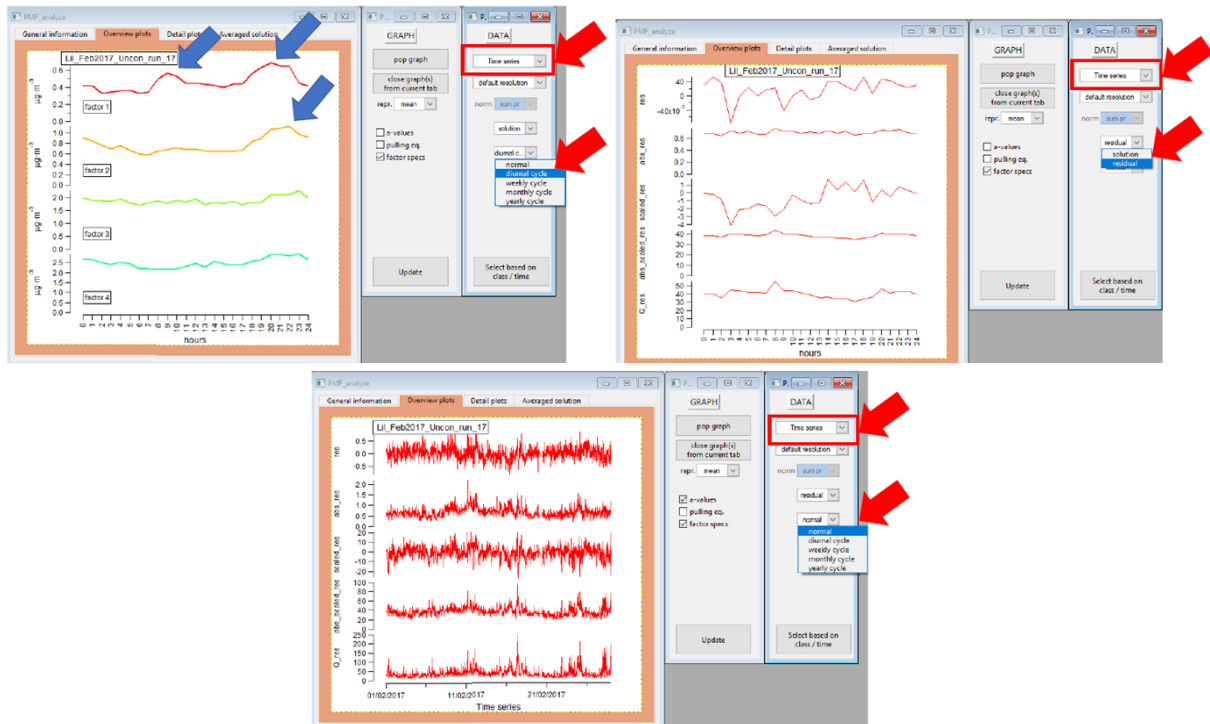


Figure 33 : Différents tracés à réaliser dans le panneau 'Overview'

Tracés détaillés 'Detail plots'

Le panneau des tracés détaillés propose différents ensembles et graphiques : « Fraction », « Standard », « Scatter », « Résidus » et « Corrélation ». Ces tracés fournissent des informations supplémentaires sur les facteurs de la FMP, par exemple leur contribution, leur corrélation avec des variables externes, etc.

Dans le **panneau des fractions**, il est possible de tracer un diagramme circulaire de la contribution, de la variation expliquée et de la concentration. Ces graphiques montrent la contribution de chaque facteur d'OA et de chaque variable. SoFi permet également d'inclure la contribution des espèces ACSM. Il est également possible de tracer un diagramme à barres de la contribution. Le graphique de la variation expliquée présente la partie bruitée inexplicée (la partie grise) liée aux données bruitées, et la partie réelle inexplicée correspond à ce que le modèle n'a pas pu reproduire lors de ce test.

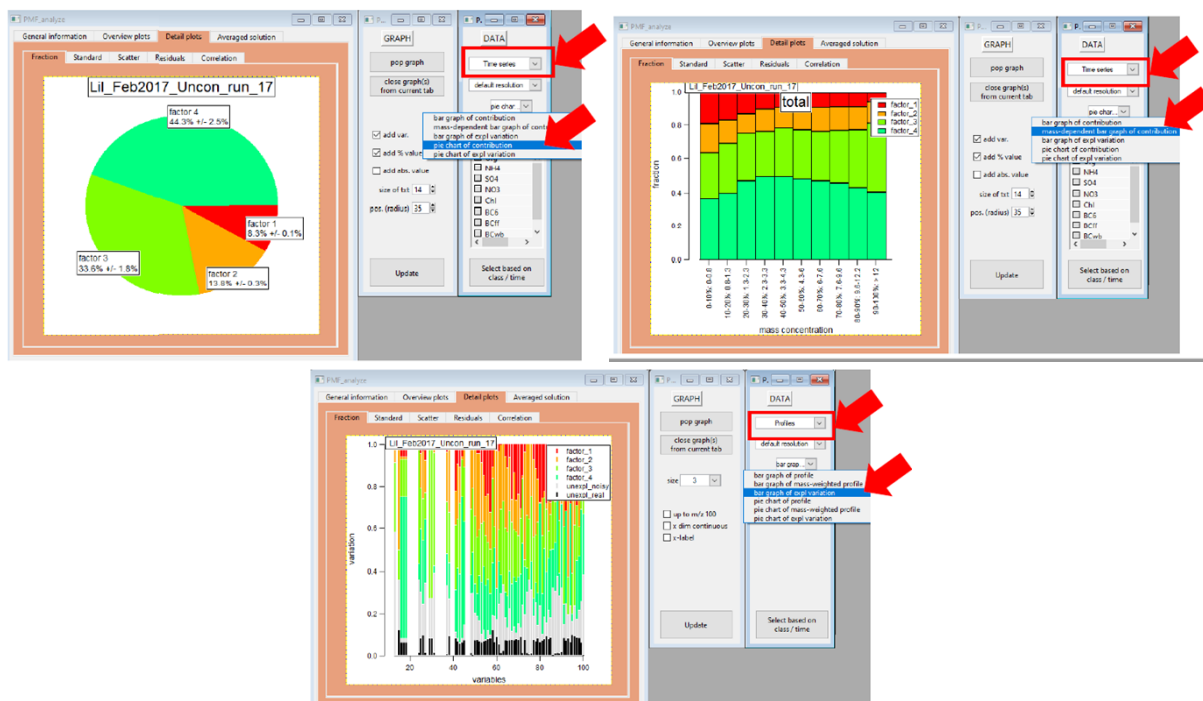


Figure 34 : Différents tracés à réaliser dans le panneau « Detail plots » ; onglet « Fraction »

Dans le **panneau standard**, l'utilisateur peut tracer différents graphiques pour les profils et les séries temporelles. Ce panneau offre davantage d'options que le panneau précédent : il permet de tracer la solution, les résidus, les séries temporelles des facteurs et les données externes. Il est également possible d'observer la variation de chaque variable de la solution.

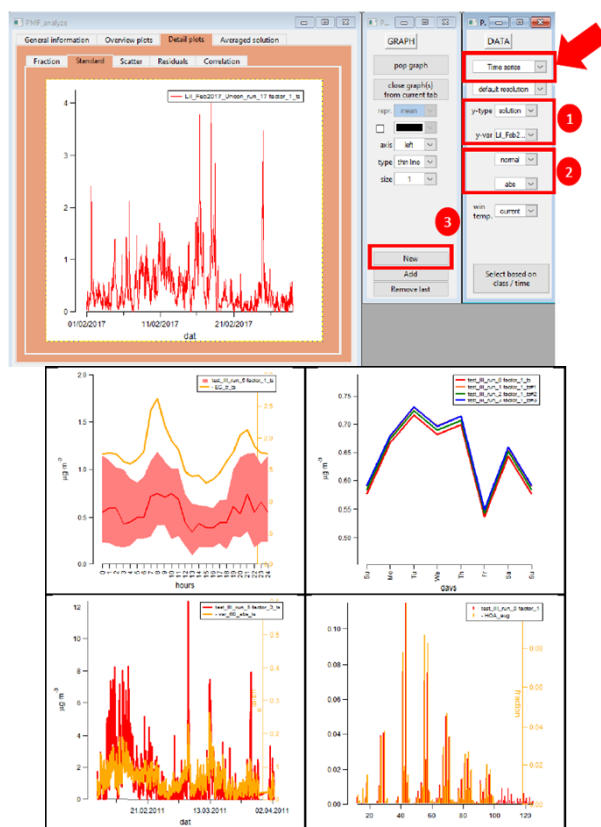


Figure 35 : Différents tracés à réaliser dans le panneau « Detail plots » ; Onglet Standard

Puis, dans le **panneau 'Scatter'**, des options supplémentaires sont disponibles. L'onglet « Scatter » permet de représenter les informations sous forme de nuage de points et de les coder par couleur en fonction d'autres informations au fil du temps ou sur plusieurs variables. Par exemple, tracer scatter plot des facteurs avec les données externes.

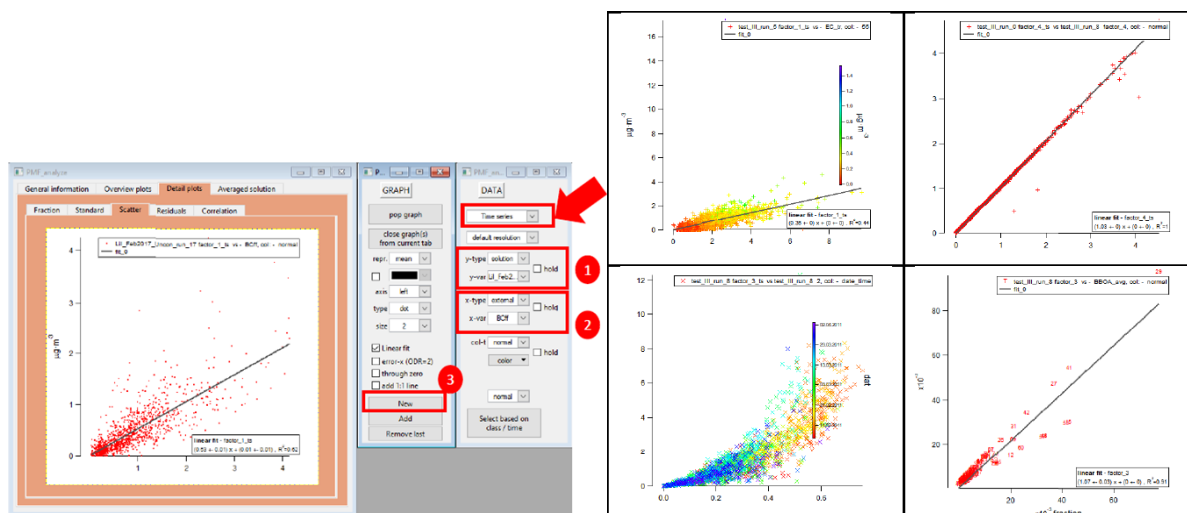


Figure 36 : Exemples de graphiques possibles de l'onglet « Scatter »

Cet onglet est également important pour analyser l'espace f44/43. Pour ce faire, l'utilisateur doit choisir « model subtraction » pour les types x et y en termes de fractions, et sélectionner les facteurs les plus oxydés (i.e., de type OOA), par exemple pour soustraire les facteurs primaires, colorés par date, puis visualiser la différence entre les deux OOA dans le graphique f44/43.

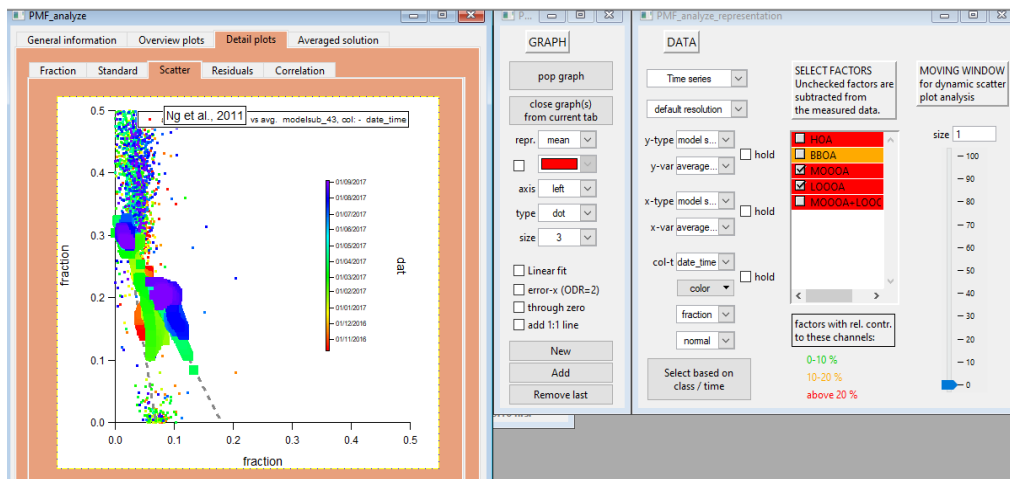


Figure 37 : Résultat de l'option de soustraction de modèle pour f44 vs f43

Des informations supplémentaires sur les résidus sont disponibles dans le **panneau des résidus**. L'utilisateur peut consulter la matrice entière ou les histogrammes et choisir le type de résidu (par exemple, la matrice entière des résidus 'scaled'). Il peut également vérifier les résidus de chaque variable à chaque instant et les présenter sous forme d'histogramme pour visualiser la distribution des résidus et détecter d'éventuels problèmes dans les données.

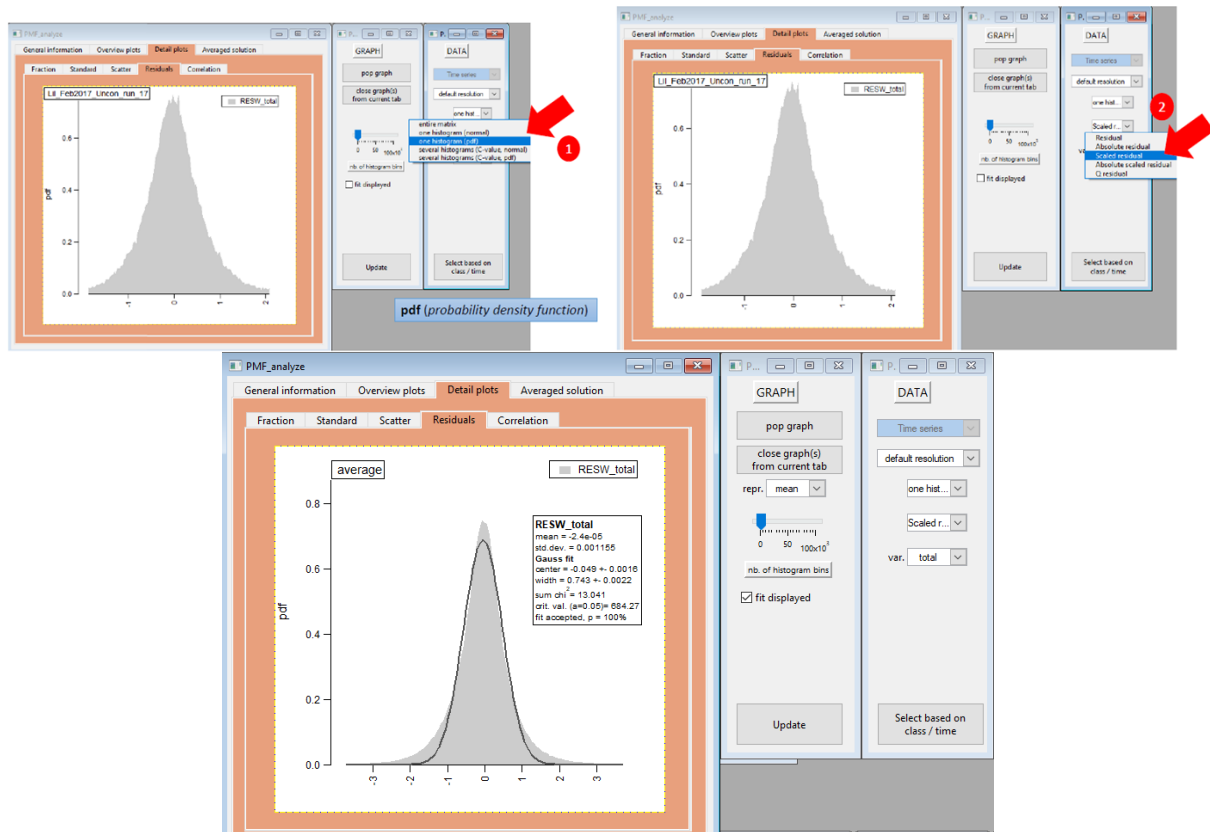


Figure 38 : Histogramme des résidus 'scaled'

Dans le **panneau de corrélation**, l'utilisateur peut vérifier la corrélation (matrice ou tableau) entre les séries temporelles de facteurs (ou les profils de facteurs) et les données externes. Il est possible de représenter graphiquement les valeurs R ou P sous forme de tableau ou d'image, ainsi que le type de corrélation. Cette corrélation est obtenue en vérifiant les listes x et y, puis en les affichant, comme illustré dans les figures ci-après.

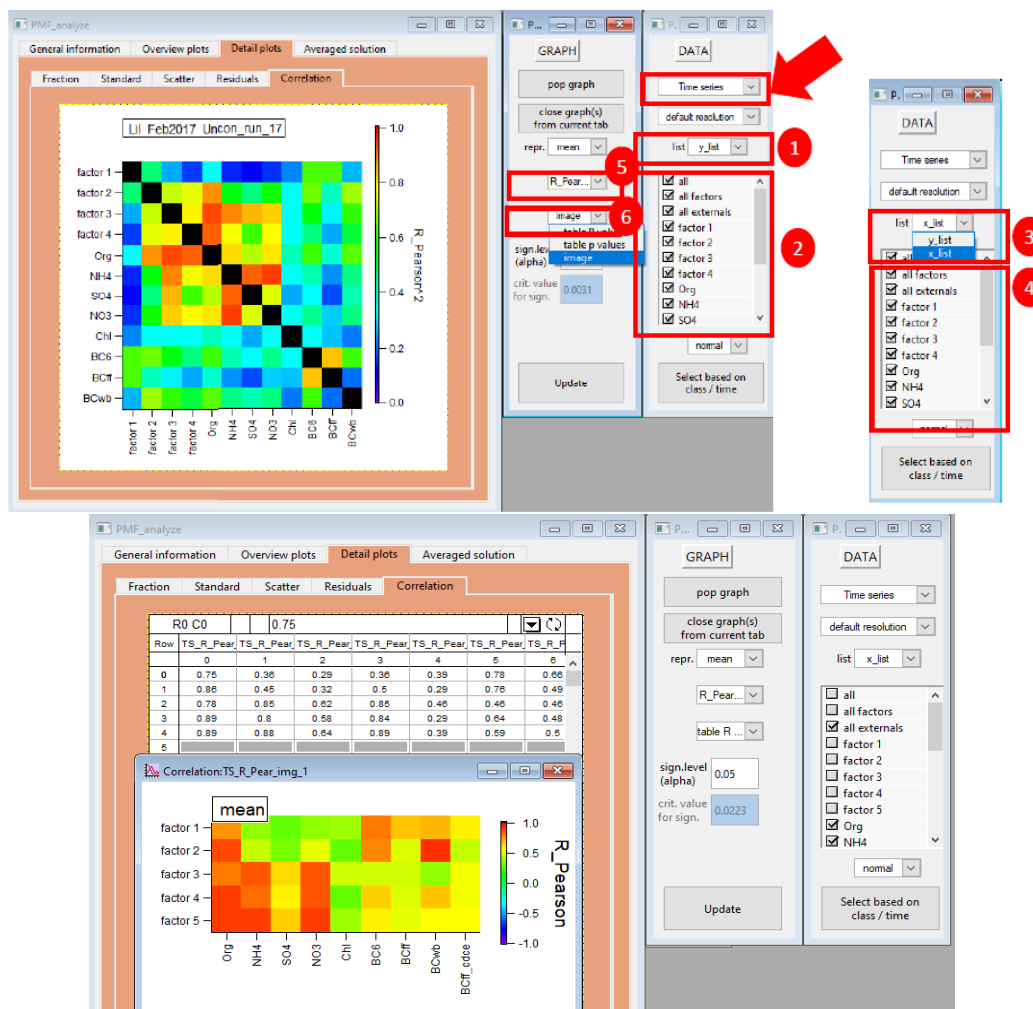


Figure 39 : Onget Corrélation ; Visualisation graphique et tableau de corrélations

Pour résumer la « PMF sans contrainte » :

Pour la PMF sans contrainte, l'utilisateur doit utiliser de 3 à 8 facteurs pour tester toutes les possibilités, vérifier les profils quotidiens des facteurs, la corrélation avec les données externes disponibles, etc. afin de déterminer le nombre de facteurs/sources possibles. Si les facteurs identifiés ne sont pas « propres », l'utilisateur peut passer au « PMF avec contrainte » présenté dans la section suivante.

3.5 Exécuter la « PMF avec contraintes »

Pour démarrer une autre PMF, l'utilisateur doit créer un nouveau fichier HDF comme indiqué dans la section précédente, puis enregistrer l'expérience. Pour définir les contraintes, il doit cliquer sur le panneau « Model Matrices ».

Dans cette partie, l'utilisateur peut exécuter la PMF contrainte, où il est possible de contraindre des facteurs ou des séries temporelles.

Trois options de contraintes peuvent être appliquées : **fpeak**, **a-value** et **pulling equations**. En parallèle, l'option « **seed** » permet de définir les profils de référence ou les séries temporelles. Il s'agit de la valeur de départ à partir de laquelle la PMF doit démarrer, et en fonction de ces valeurs, les facteurs sont contraints. L'option « factor specs » permet de définir les noms des facteurs.

Dans une analyse PMF, les options les plus utilisées sont « seed » et « a-value ». Cependant, dans ce guide, toutes les options sont détaillées.

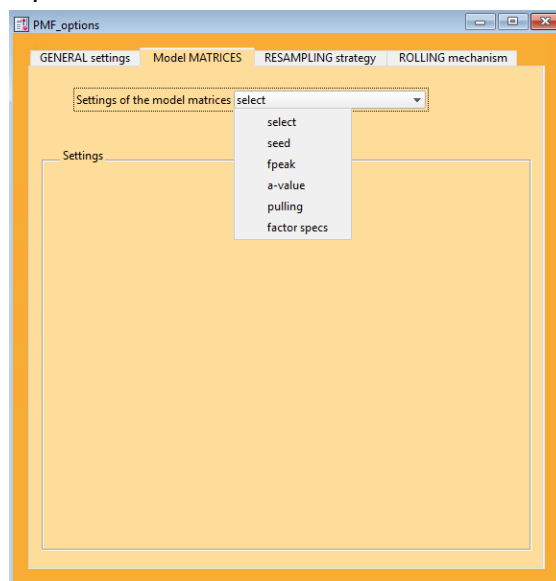


Figure 40 : Panneau Matrices du modèle

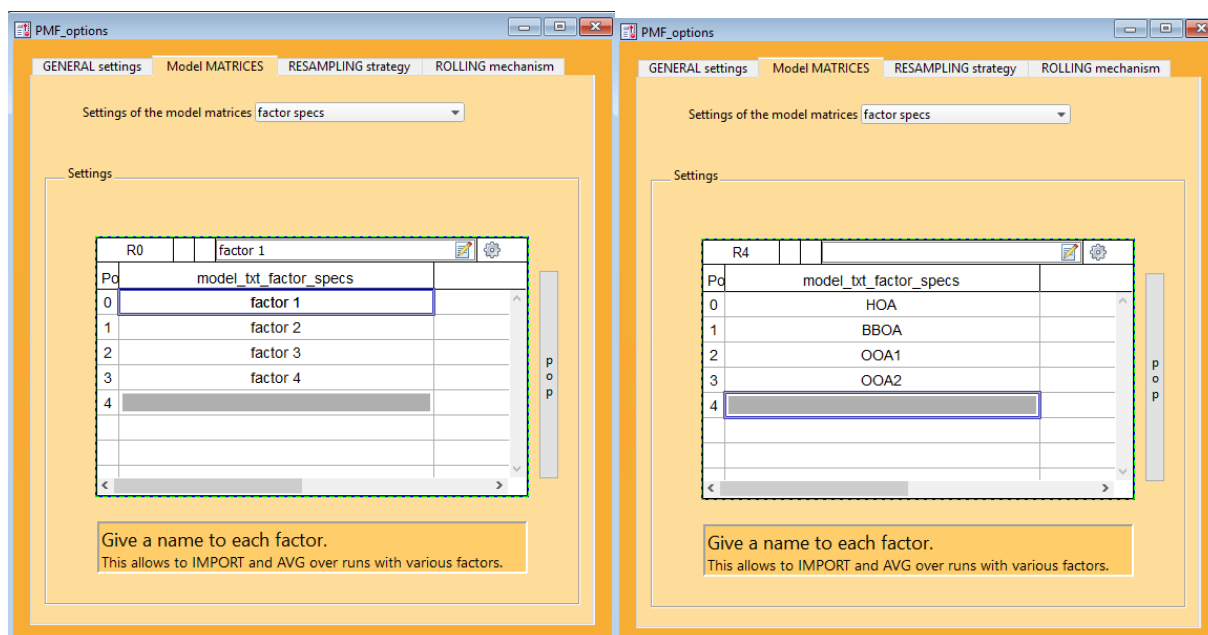


Figure 41 : Onglet « factor specs », noms par défaut (à gauche)

et noms définis par l'utilisateur (à droite)

3.5.1 Seed

Comme indiqué précédemment, 'seed' correspond au nombre de démarrages de PMF, c'est-à-dire aux valeurs initiales. Il permet d'initialiser les facteurs de référence (ou séries temporelles).

L'utilisateur sélectionne d'abord le « factor » et les « initial values » : « random », « pseudo-random » ou « from vector ». Si l'utilisateur souhaite conserver des caractéristiques spécifiques, « pseudo-random » signifie initialiser une seule fois et conserver la même valeur initiale aléatoire. L'orientation « random » comporte toujours des entrées entièrement aléatoires. À chaque exécution de PMF, SoFi initialise aléatoirement de manière complète ; soit pour les profils ou les séries temporelles. L'option « from vector » permet de contraindre le facteur à partir d'un vecteur. SoFi place alors automatiquement le facteur choisi dans le dossier « reference_pr ». Ensuite, pour l'option « variables », il est possible de contraindre des « single ions » du vecteur ou « all variables ». Ci-dessous, le graphique du profil de référence, où les variables sélectionnées sont présentées en rouge, permet de normaliser ou de dénormaliser le facteur. Si le profil complet est contraint, le facteur doit être normalisé à la somme de 1. L'utilisateur clique ensuite sur « write into the model matrix » pour importer le profil. Dans le tableau de gauche, l'utilisateur peut vérifier le rapport m/z et ses valeurs correspondantes. Le champ jaune « all model var. of factor defined » devient vert, ce qui signifie que le facteur est écrit. Les mêmes paramètres s'appliquent aux autres facteurs.

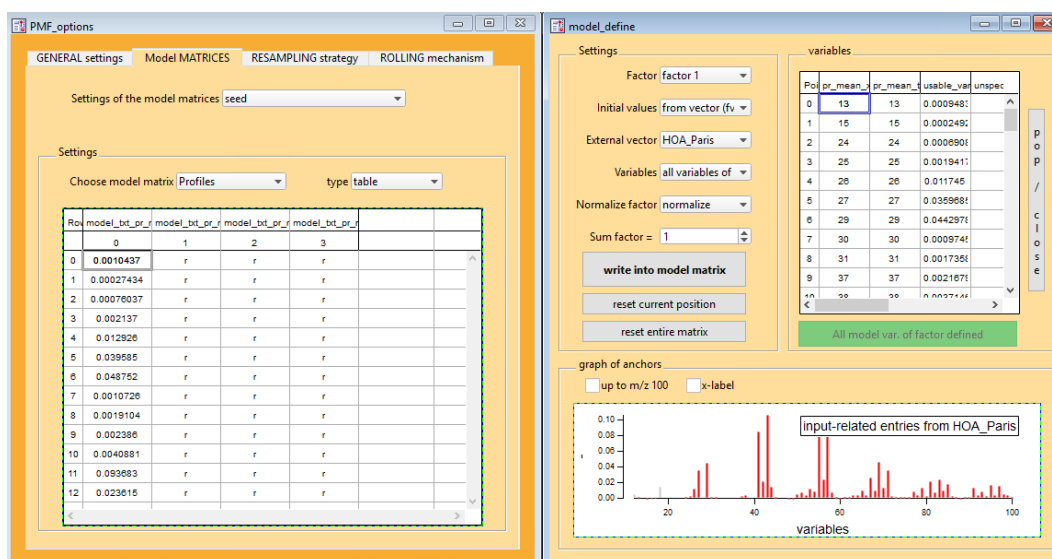


Figure 42 : Exemple 'seed'

3.5.2 Approche par a-value

L'approche par 'a-value' consiste à paramétrer le degré de différence autorisé entre le profil final et le profil de référence. Cela signifie que nous définissons une plage de variation pour le facteur de référence initialisé dans la partie « seed ». Cette étape s'inscrit dans la continuité de l'approche a-value. Il existe plusieurs types de contraintes pour l'approche a-value : « exact a-value », « random a-value » et « test of sensitivity » avec différentes dimensions. Cette approche permet également de contraindre les profils et les séries temporelles.

Cette partie se concentre sur les profils.

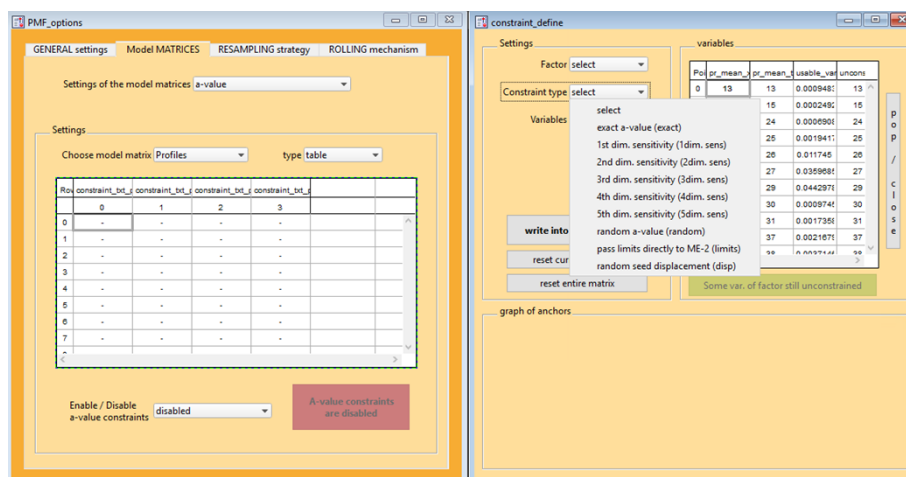


Figure 43 : Onglet « a-value »

L'utilisateur doit d'abord définir les valeurs de départ 'seed' comme expliqué précédemment, puis définir les paramètres de l'approche a-value. Dans l'onglet de gauche, il sélectionne la matrice de modèle : « Profiles » ou « Time series ». Dans la plupart des cas, les profils de contraintes sont plus utilisés. Ensuite, dans l'onglet de droite, il sélectionne le 'factor' et le 'Constraint type'. Il existe différents types de contraintes : « **Exact a-value** » avec une a-value spécifique (par exemple, une a-value = 0,1 signifie que le profil du facteur peut varier de $\pm 10\%$), « **sensitivity analysis** » avec différents tests de sensibilité de la a-value, par exemple, en exécutant la PMF une fois avec la contrainte de facteur 1 en utilisant la 1^{ère} dimension de sensibilité et en commençant par une a-value donnée, par exemple de 0 à 1 avec un pas de 0,1. Dans ce cas, l'analyse PMF est exécuté 11 fois, une première fois avec une a-value de 0, une seconde avec une a-value de 0,1, une troisième avec 0,2, et ainsi de suite. L'analyse de sensibilité prend en compte différentes dimensions de sensibilité. Ainsi, si l'utilisateur contraint plusieurs profils, les valeurs a-value de tous les profils peuvent être modifiées simultanément ou individuellement. Par exemple, pour contraindre un seul facteur, l'utilisateur peut utiliser la 1^{ère} dimension de sensibilité. Pour contraindre deux facteurs, il peut contraindre le premier facteur avec la 1^{ère} dim. de sensibilité et le second avec la 2nd dim. de sensibilité. Ainsi, tandis que la valeur a-value du premier facteur reste à 0, celle du second facteur passe de 0 à 1, puis a-value du 1^{er} facteur passe à 0,1, et celle du 2nd facteur est augmentée dans la même plage. L'analyse de sensibilité permet donc de tester toutes les possibilités de combinaisons des valeurs a-value. La troisième option consiste à utiliser une a-value aléatoire « **random a-value** », où l'utilisateur spécifie une valeur a-value supérieure et inférieure, afin d'obtenir une valeur aléatoire comprise entre les deux, en spécifiant une valeur a-delta qui représente l'incrément. Il existe une autre option, à savoir « pass limits directly to ME2 », afin d'avoir des limites différentes pour le profil. Mais elle n'est pas très utilisée.

Ainsi, pour les premiers tests, l'utilisateur peut contraindre les facteurs à l'aide d'une valeur a-value exacte, puis choisir une valeur a-value aléatoire pour étudier une plage de valeurs :

- Idéalement pour HOA : $a = 0,05$ ou $0,1$, mais envisager une plage plus large (0-0,3) ;
- Pour BBOA : $a = 0,3$ ou $0,5$, mais envisager une plage plus large (0-0,5).

Après avoir choisi le type de contraintes, l'utilisateur doit sélectionner les variables soumises à cette valeur a. Il existe une option permettant de définir des valeurs a pour chaque ion. Une autre option consiste à appliquer une valeur a à toutes les variables : « all model variables » ou « all variables of external vector ». Si le vecteur externe contient toutes les variables, les deux options sont correctes et l'utilisateur peut choisir « all model variables ». Il existe certains cas où elles sont différentes, comme dans le cas de l'utilisation de données HR (à l'aide d'autres instruments que ceux actuellement mis en œuvre dans le cadre du programme CARA).

L'utilisateur peut ensuite cliquer sur « write into the model matrix » pour insérer les paramètres définis dans le tableau de gauche. La même opération peut être effectuée pour les autres facteurs. Les facteurs primaires sont ceux qu'il est recommandé de contraindre. Enfin, l'utilisateur doit activer l'approche par a-value.

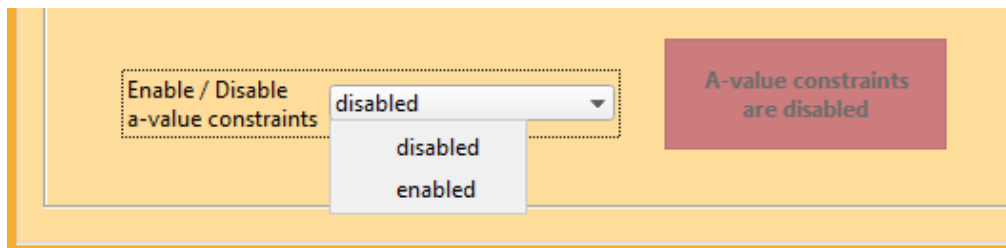


Figure 44 : « Enabled » a-value

Exemple d'application de l'analyse de sensibilité

L'analyse de sensibilité permet de ne pas modifier automatiquement ces valeurs a. Par exemple, si le profil est issu d'une étude et que l'utilisateur souhaite l'ajuster aux données, il peut le tester avec une valeur a à l'aide de l'analyse de sensibilité. Il existe 5 dimensions de sensibilité, soit 5 analyses de sensibilité indépendantes (5 dimensions of a-values matrix). Il est également possible d'utiliser la même dimension pour tous les facteurs si l'utilisateur souhaite une seule analyse de sensibilité. Cela dépend de l'utilisation souhaitée.

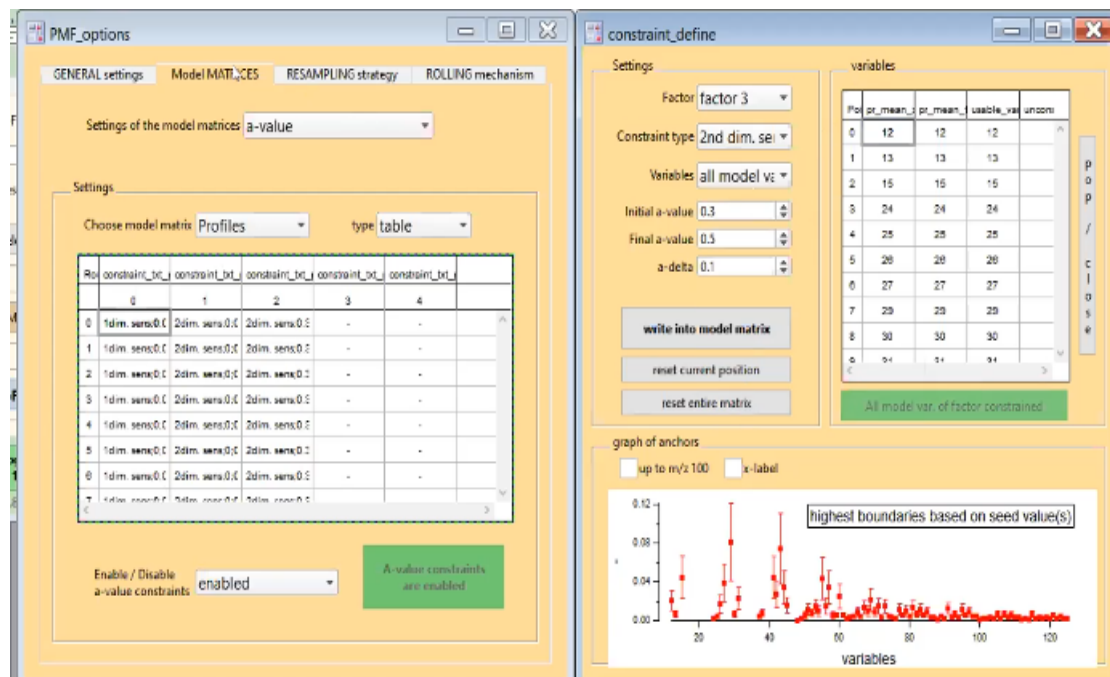


Figure 45 : Paramètres de valeur a

Dans l'exemple de la figure 46, une analyse de sensibilité est appliquée à BBOA (valeur a de 0,3 à 0,5 avec un pas de 0,1), COA (valeur a de 0 à 0,3 avec un pas de 0,1) et HOA (valeur a de 0 à 0,5 avec un pas de 0,1).

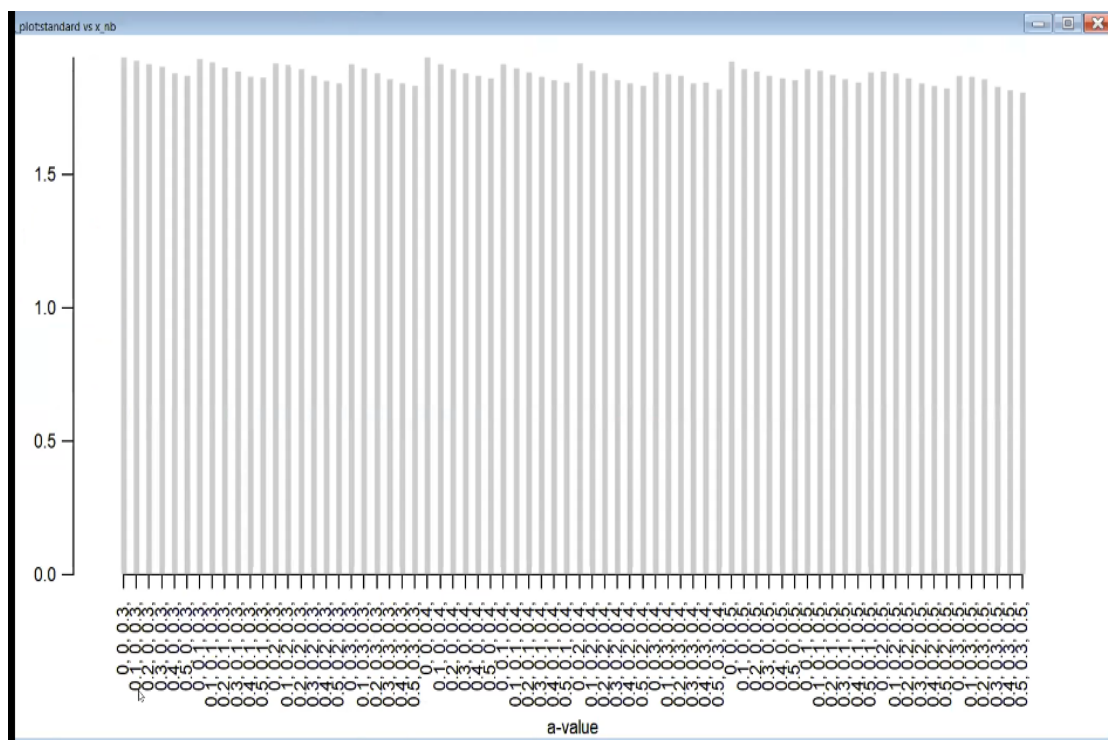


Figure 46 : Analyse de sensibilité

3.5.3 fpeak

Deux fonctionnalités de fpeak : « global fpeak » et « individual fpeak ». L'approche fpeak permet des localisations.

Le fpeak global permet une rotation de tous les facteurs, c'est-à-dire des rotations complètes entre toutes les combinaisons. Dans ce cas, trois options sont possibles : « exact fpeak », « sensitivity analysis » et « random fpeak ». Le fpeak exact n'a qu'une seule valeur de fpeak. L'analyse de sensibilité propose une plage de fpeaks : « initial fpeak » et « final fpeak », ainsi qu'une « user-defined fpeak resolution ». Le fpeak aléatoire génère des fpeaks compris entre -1 et 1 avec un pas de 0,1.

Le fpeak individuel permet d'analyser la façon dont d'autres solutions apparaissent, ce qui est important, notamment pour des rotations spécifiques. Si le fpeak individuel est activé, l'utilisateur dispose de la matrice de rotation. Autrement dit, si l'utilisateur contraint 5 facteurs, puis l'utilisateur transmet les positions qui doivent être pivotées.

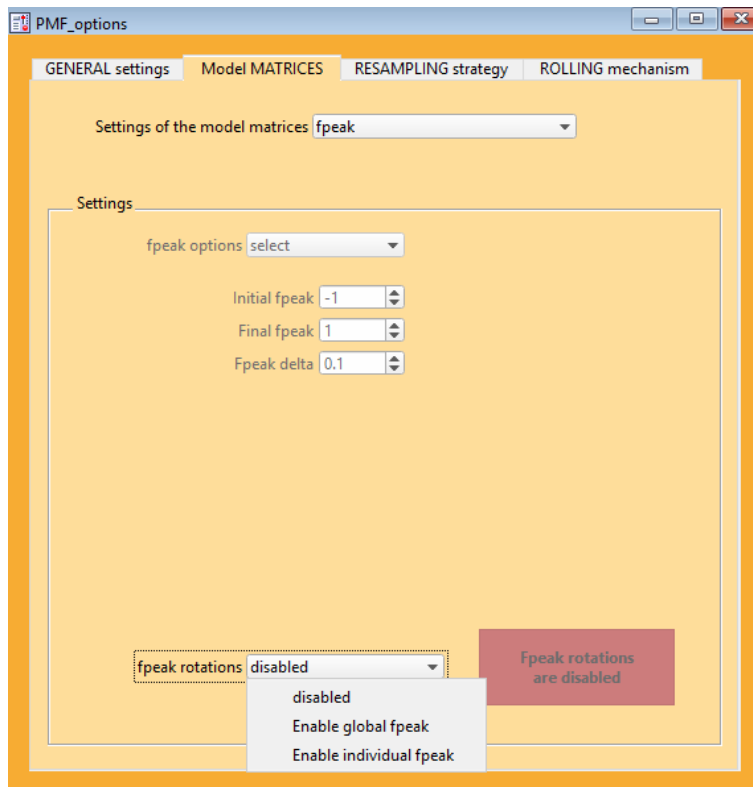


Figure 47 : Onglet fpeak

3.5.4 « Pulling equations »

Les équations d'extraction sont des équations supplémentaires qui peuvent être ajoutées au modèle PMF.

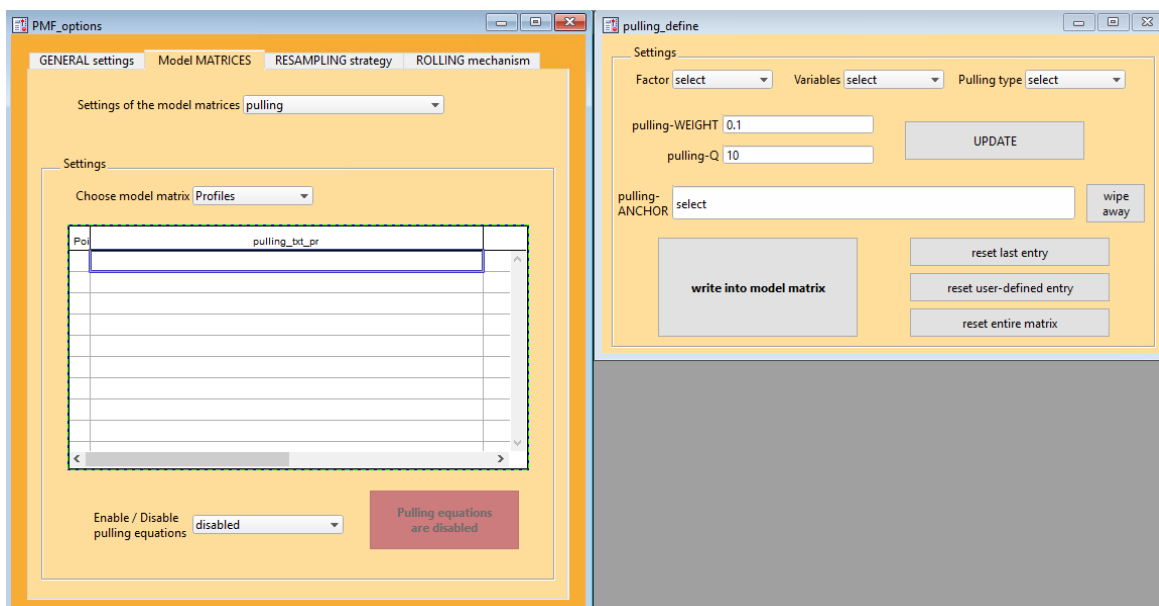


Figure 48 : Onglet 'pulling equations'

Voici un exemple de contrainte du BBOA à l'aide de l'équation d'extraction :

Par exemple, l'utilisateur peut définir une équation pour la variable m/z 60 du facteur de combustion de la biomasse (BBOA). Il doit d'abord définir ANCHOR (la valeur à extraire) et le WEIGHT (le poids d'extraction est de 0,1, qui représente l'incertitude transmise au modèle, est utilisé pour reposer l'équation d'extraction). Il peut ensuite modifier la valeur de l'équation. Il peut ensuite cliquer sur « write into model matrix » ; l'équation apparaît alors dans le tableau de gauche. L'utilisateur doit activer l'équation d'extraction pour appliquer les équations définies.

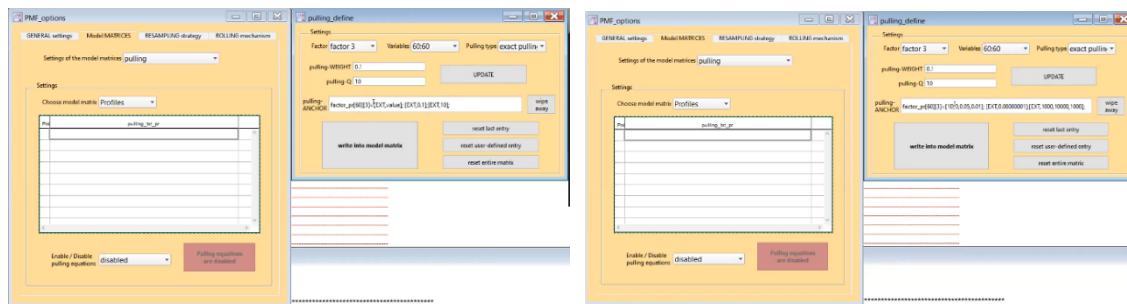


Figure 49 : Exemple de contrainte BBOA par équation d'extraction

L'utilisateur peut également tester l'analyse de sensibilité à l'aide de l'équation d'extraction :

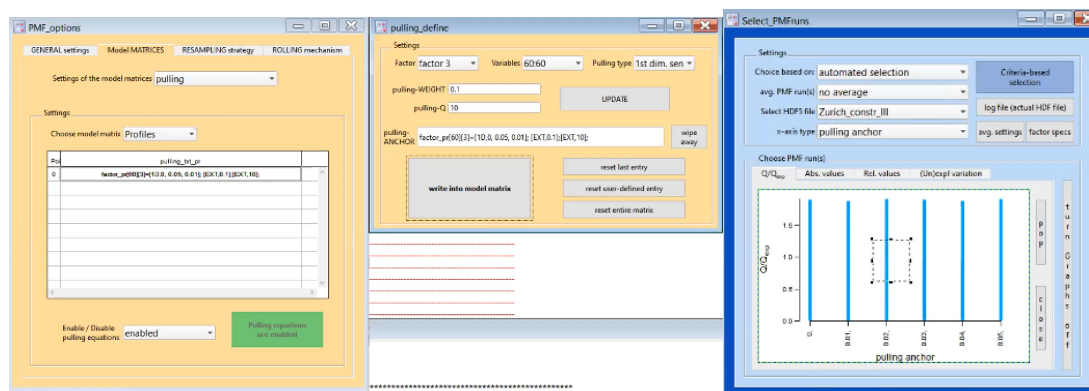


Figure 50 : Exemple de contrainte BBOA par équation d'extraction avec analyse de sensibilité

3.6 Utiliser la stratégie de rééchantillonnage « Bootstrap »

Cette stratégie consiste à activer le rééchantillonnage Bootstrap. Le bootstrap permet de tester la stabilité des résultats. Il génère une nouvelle matrice d'entrée à partir des données d'origine afin de vérifier la robustesse du résultat. Il est important d'activer la fonction Bootstrap, notamment pour les résultats finaux du PMF, afin d'estimer les incertitudes statistiques.



Figure 51 : Panneau de stratégie de rééchantillonnage

3.7 Exécuter la « PMF glissante »

Le mécanisme de 'rolling PMF' consiste à définir une petite fenêtre décalée par incréments de 1-jour sur l'ensemble des données, capturant ainsi les variations saisonnières des profils factoriels. Ceci est particulièrement utile pour les ensembles de données à long terme, car l'exécution et la validation des analyse PMF peuvent s'avérer très chronophages sur de grands jeux de données. Il est alors préférable d'utiliser le PMF glissant, qui consiste à appliquer de petites fenêtres aux données. Pour chaque fenêtre du PMF, les incertitudes sont propagées pour les résultats du PMF, avec par exemple une fenêtre temporelle de 14 jours et un décalage d'un jour, soit environ 14 répétitions.

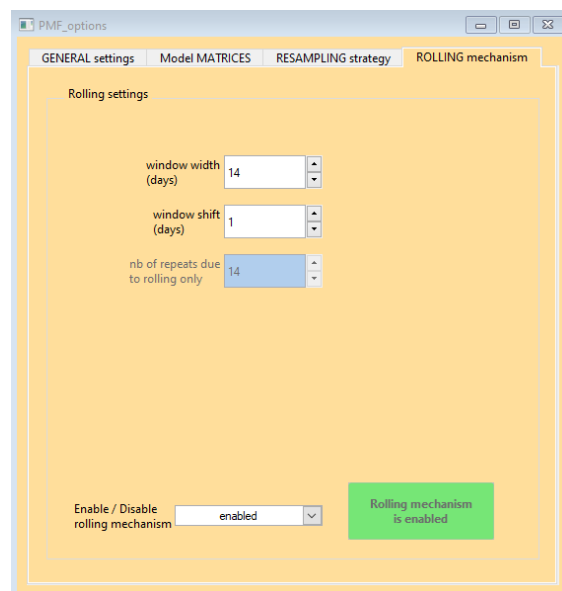


Figure 52 : Onglet PMF glissante

Dans cette partie, l'utilisateur doit définir la durée et l'intervalle de décalage de la fenêtre de temps élémentaire considérée. Les valeurs par défaut de ces paramètres sont de 14 jours pour chaque fenêtre et d'un jour de décalage, comme indiqué dans Canonaco et al. 2021. L'utilisateur peut néanmoins tester différentes fenêtres temporelles afin de choisir celle qui présente potentiellement le mieux les fréquences d'oscillations au sein du jeu de données considéré. Pour ce faire, l'utilisateur peut vérifier la variation de Q/Q_{exp} et des points non modélisés (ou des points manquants correspondant aux points non conformes aux critères définis par l'utilisateur pour toutes les exécutions) avec différentes fenêtres temporelles. Finalement, il choisit la fenêtre présentant les plus faibles valeurs de Q/Q_{exp} et le moins de points non modélisés. La fenêtre temporelle souhaitée est celle avec les valeurs supérieures de s_{win} (taille de la fenêtre), a_{max} (valeur a maximale) et x_{PMF} (répétitions PMF).

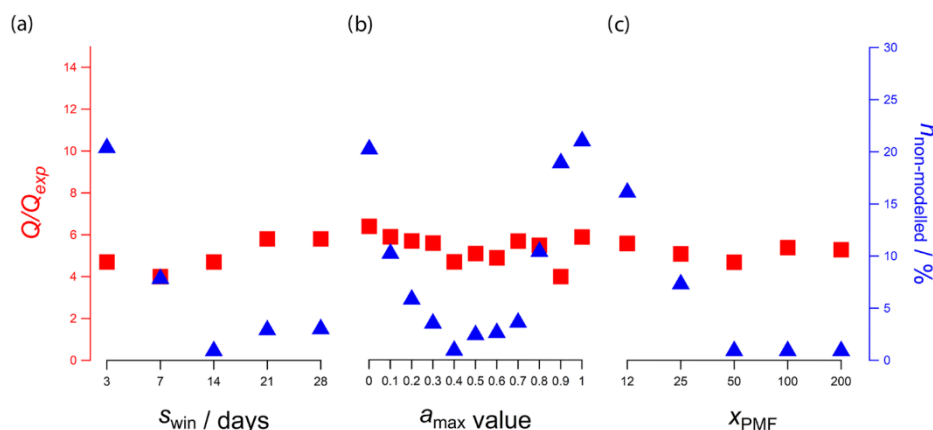


Figure 53 : Q/Q_{exp} et pourcentage de points temporels non modélisés sur l'ensemble des données en fonction de la taille de la fenêtre (s_{win}), de la valeur a maximale (a_{max}) et du nombre de répétitions PMF par fenêtre (x_{PMF}) (Canonaco et al. 2021)

3.8 Sélection par critères

La sélection par critères '**Criteria-based selection**' est particulièrement utilisée pour les résultats de PMF glissantes ou les résultats obtenus via bootstrap pour lesquels il existe de nombreuses exécutions, afin de sélectionner les meilleurs.

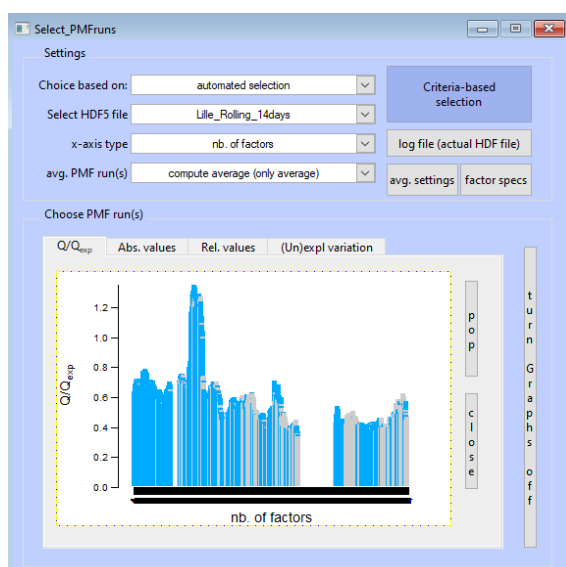


Figure 54 : Sélection des exécutions PMF selon le panneau de critères

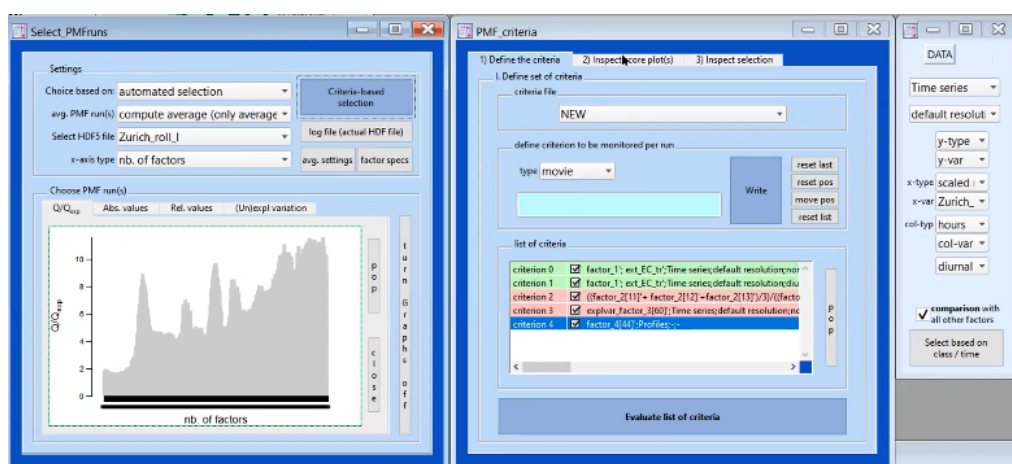


Figure 55 : Panneau de critères PMF

Trois panneaux sont disponibles pour les critères PMF.

Le **premier** est « **Define the criteria** », où l'utilisateur spécifie les critères, par exemple la corrélation entre les facteurs et les données externes à l'aide du panneau DATA. Il est possible de modifier le cycle (diurne, hebdomadaire, mensuel, etc.) et le type (abs ou fraction). L'utilisateur peut ensuite cliquer sur le bouton « Write » pour ajouter le critère à la liste. Le manuel SoFi (« sur le site Datalystica ») présente de nombreux exemples de critères de sélection ou de suivi. L'utilisateur vérifie ensuite si les critères sont cochés en cliquant sur le bouton « Evaluate list of criteria ».

Le **deuxième** panneau, « **Inspect score plots** », propose différentes options pour sélectionner les exécutions principales. Dans le panneau DATA, dans la section CRITERIA, l'utilisateur choisit le critère. Par exemple, dans la figure 56, le premier critère présente l'ensemble des corrélations obtenues entre le facteur n°1 (ici HOA) et BCtr (fraction du carbone suie provenant des émissions du transport automobile, utilisé comme contrainte externe). Cliquer sur « Pop batch table » permet ainsi de consulter le tableau résumant les principaux paramètres des critères et de visualiser la représentation complète de la corrélation au fil du temps, chaque point représentant une exécution PMF.

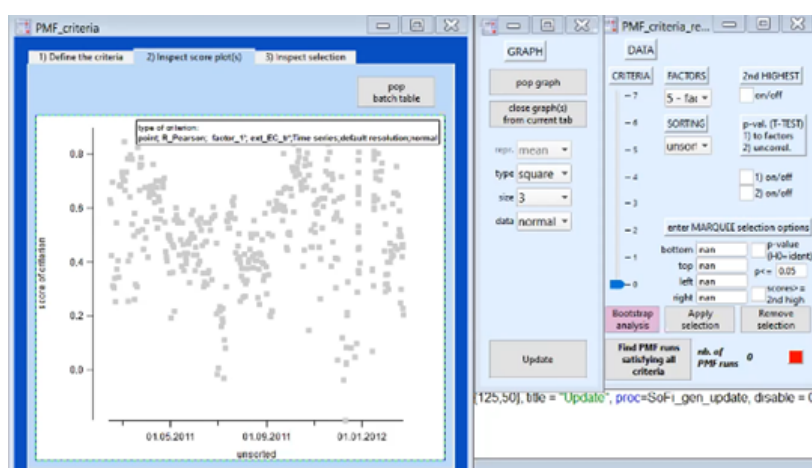


Figure 56 : premier critère « Corrélation entre le facteur 1 et BCtr »

Sur la droite (figure 56), des options permettent de sélectionner les meilleures séries avec certaines statistiques comme le T-TEST. Si la case « comparison with all other factors » du premier panneau est décochée, SoFi calcule uniquement les scores pour ces critères. Si elle est activée, SoFi compare les autres facteurs et effectue toutes les combinaisons. Ainsi, le bouton « 2nd Highest » du deuxième panneau signifie que SoFi calcule également les scores avec d'autres facteurs. Ensuite, après avoir cliqué sur « apply selection », les points bleus indiquent la corrélation entre le facteur 1 et BCtr, et les points noirs indiquent la corrélation entre le facteur 2 ou le facteur 3 et BCtr. Pour le T-TEST, deux options sont disponibles : « to factors » (pour sélectionner les points présentant une corrélation significative par rapport aux autres facteurs) et « uncorrel. » (pour une corrélation significative entre le facteur 1 et BCtr). Des options sont également disponibles pour la « p-value » et l'axe. Il est également possible de trier les facteurs et d'en modifier les solutions. Dans les versions de SoFi 8 et 9, d'autres types de critères (i.e., pas seulement des séries temporelles) peuvent être utilisés comme contraintes externes, tels que les images ou les vidéos.



Figure 57 : Définition des critères ; type image ou vidéo dans SoFi 8

Ensuite, l'utilisateur clique sur « Find PMF runs satisfying all criteria » et le nombre d'exécutions PMF sélectionnées s'affiche.

Il passe ensuite au **troisième panneau « Inspect selection »**. Les points bleus correspondent aux points sélectionnés et les points verts à toutes les exécutions PMF.

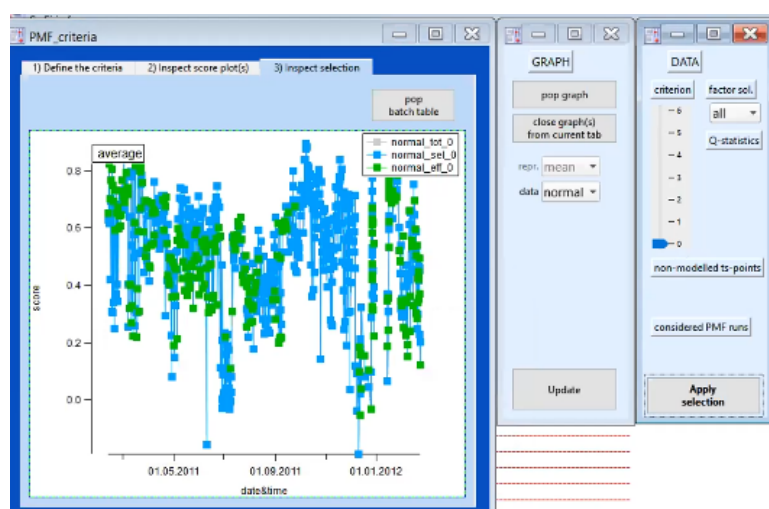


Figure 58 : panneau 'Inspect the selection'

L'utilisateur doit ensuite cliquer sur « apply selection » pour enregistrer les modifications, puis fermer la fenêtre.

La sélection est effectuée dans le panneau principal. L'utilisateur choisit ensuite l'option « Average ».

3.9 Solution moyenne 'Averaged solution'

Cette option correspond à l'option « averaged » du dernier panneau. Elle est uniquement disponible pour la solution moyenne.

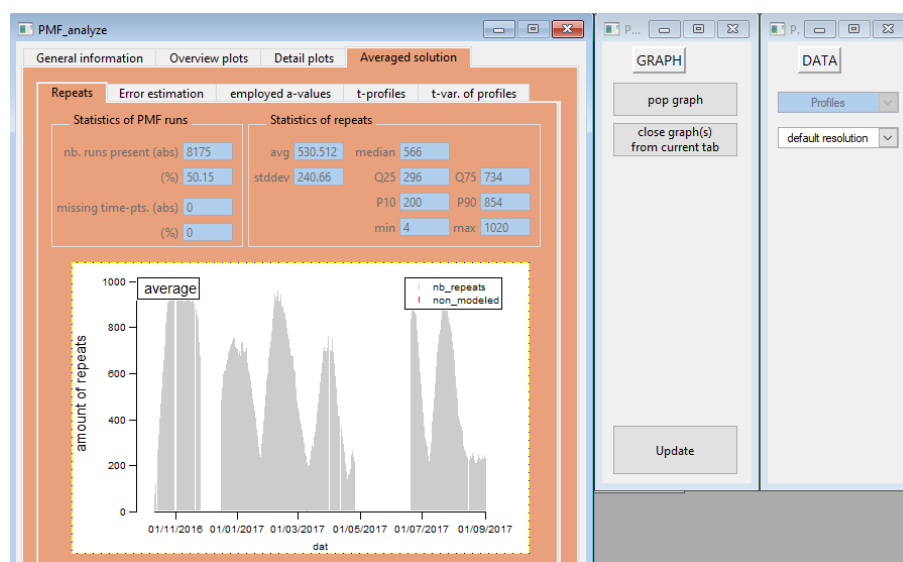


Figure 59 : Panneau « Averaged solution »

Cet onglet présente plusieurs panneaux relatifs à la solution moyenne (c.-à-d. paramètres et statistiques de la PMF) : « repeats » (statistiques sur les répétitions), « error estimation » (incertitudes et incertitudes de rotation, pour les explorer dans des scatter plots ou des histogrammes en traçant l'écart-type/la moyenne, les erreurs de la PMF), « employed a-values » (statistiques des valeurs a), « t-profiles » (profil temporel, variables au fil du temps), « t-var of profiles » (variables temporelles, variables de chaque facteur et leur évolution au fil du temps).

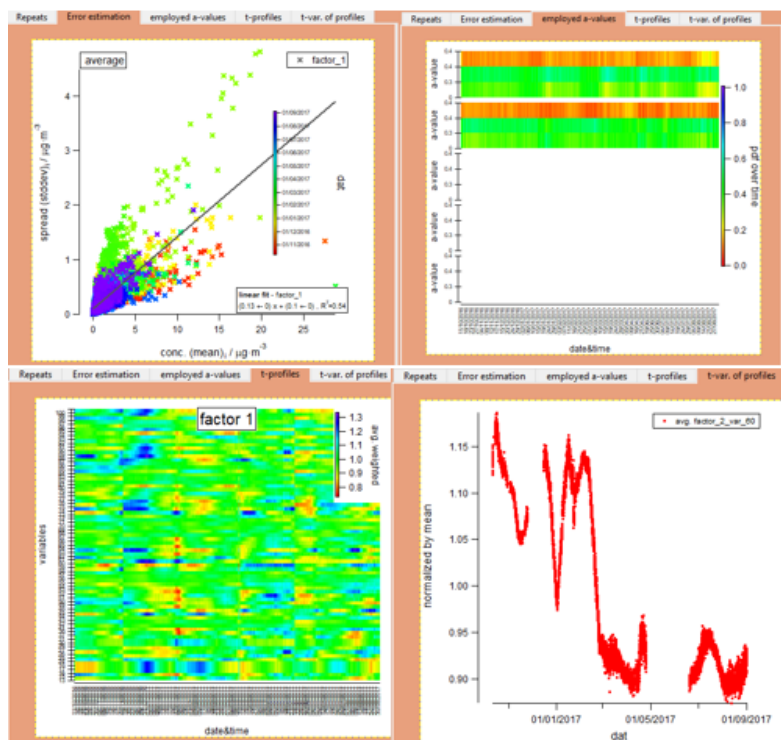


Figure 60 : Quelques graphiques de l'onglet « Averaged solution »

Il est possible de définir des couleurs pour les facteurs PMF dans le paramètre principal de SoFi et elles seront appliquées à tous les tracés générés dans l'expérience Igor.

4. Cas concrets d'application

4.1 PMF classique

Pour réaliser une PMF classique, l'utilisateur doit suivre les mêmes étapes que celles décrites dans ce guide. Il doit d'abord importer les données dans le panneau « Import data », puis les traiter dans le panneau « Pretreatment ». L'étape suivante consiste à spécifier les principaux paramètres d'exécution de la PMF.

1) Commencer par une PMF sans contrainte (de 2 à 8 facteurs) : l'utilisateur peut alors vérifier les spectres de masse des profils de facteurs, leurs cycles journaliers et leur corrélation avec les traceurs disponibles (par exemple, NO_x, BC, etc.) afin d'identifier le nombre de facteurs possibles. Si les facteurs sont corrects et propres, voir le point 4.

2) Sinon, l'utilisateur doit réaliser une PMF avec contrainte, notamment pour les facteurs primaires, en effectuant de nombreux tests des valeurs α et en testant de nombreux profils de référence. En commençant par HOA (par exemple, Crippa et al. (2013)), puis HOA et BBOA (par exemple, Ng et al., (2011)).

3) Tester des profils d'OA primaires spécifiques en cas de doute sur leur présence ou d'hésitation entre N et N+1 facteurs : OA de cuisson (COA), OA de combustion du charbon (CCOA), etc. L'utilisateur contraint alors les spectres de masse du facteur (par exemple, COA avec Crippa et al., (2013) pour vérifier si le profil journalier est raisonnable).

4) Une fois le meilleur cas de base (c'est-à-dire la solution PMF raisonnable) choisi, l'utilisateur doit tester la stabilité de cette solution en activant l'approche Bootstrap avec, par exemple, 50 répétitions.

Si l'utilisateur dispose de données à long terme, par exemple sur un an, il est préférable d'effectuer une analyse PMF saisonnière. L'utilisateur doit donc suivre ces étapes pour chaque saison (hiver, printemps, été, automne) et s'efforcer d'obtenir un nombre constant de facteurs oxygénés pour les différentes saisons, si cela est raisonnable environnementalement. Les profils extraits du bootstrap peuvent être utilisés pour contraindre les profils pour les analyses PMF glissantes ou en temps réel.

4.2 PMF glissante

La PMF glissante est une continuité de la PMF classique, où l'ensemble de données est analysé en une seule fois, en contraignant les profils avec des profils de référence ou des profils de site spécifiques obtenus à partir des résultats PMF saisonniers. L'utilisateur doit activer l'approche glissante en définissant la largeur et le décalage de la fenêtre. Après l'exécution du modèle, l'utilisateur obtient le fichier HdF et définit des critères pour choisir les meilleures solutions PMF.

4.2.1 Étapes de définition des paramètres d'exécution PMF

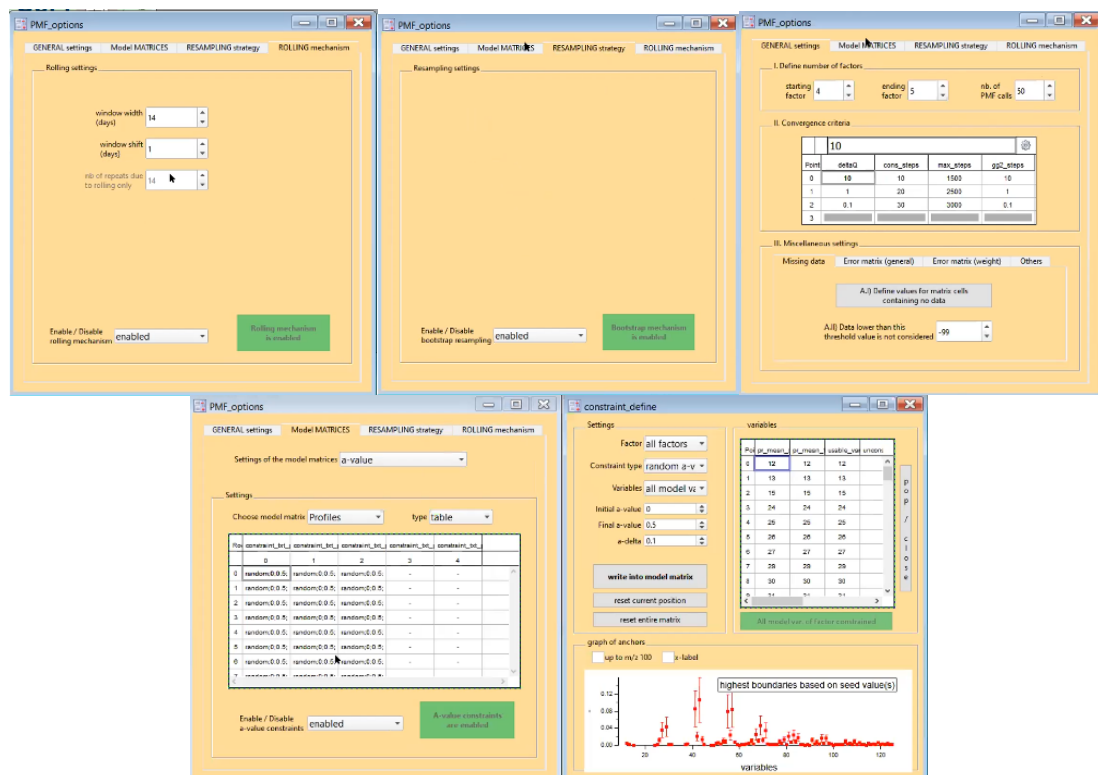


Figure 61 : Étapes de définition des paramètres d'exécution PMF

4.2.2 Étapes de définition des critères

L'utilisateur peut définir des critères pour chaque facteur. Il existe différents types de critères : certains critères de sélection, d'autres de monitoring. Si l'utilisateur dispose de plusieurs facteurs sans contrainte, il peut les trier. Par exemple, il est possible de trier les critères dont le rapport m/z est le plus élevé (44). L'utilisateur doit alors cliquer sur ce critère. La couleur noire indique qu'il s'agit d'un critère de tri.

Ci-dessous, un exemple de critère pour la variabilité de m/z 60 d'un facteur 3. Chaque point du graphique présente la variation expliquée de m/z 60 d'un facteur 3.

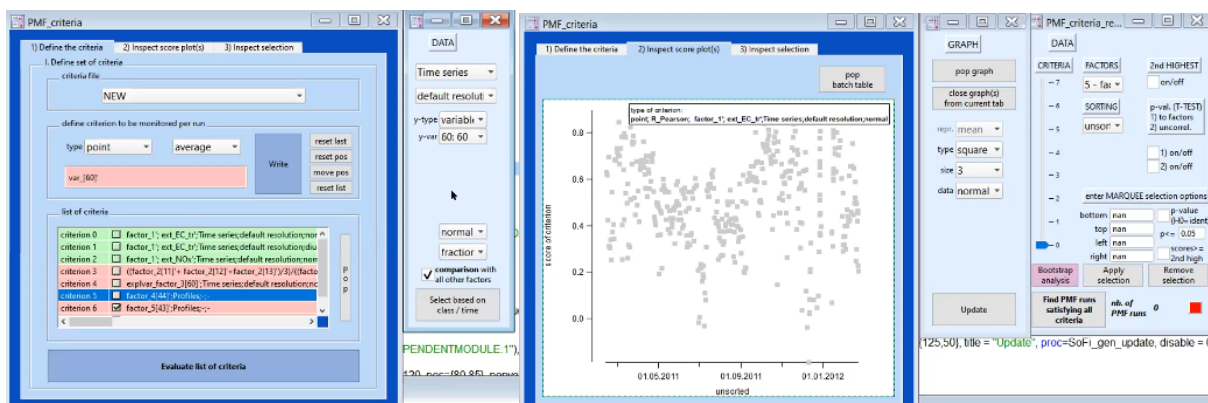


Figure 62 : Étapes de définition des critères

criteria_listsol_criteria_active		
R3 C1		((factor_2[11]+factor_2[12]+factor_2[13])/3)/((factor_2[9]+factor_2[10]+factor_2[14]+factor_2[15])/4);Time series;default resolution;diurnal cycle
	criteria_list[0]	criteria_list[1]
	0	1
criteria 0		factor_1'; ext_EC_tr';Time series;default resolution;normal
criteria 1		factor_1'; ext_EC_tr';Time series;default resolution;diurnal cycle
criteria 2		factor_1'; ext_NOx';Time series;default resolution;normal
criteria 3		((factor_2[11]+factor_2[12]+factor_2[13])/3)/((factor_2[9]+factor_2[10]+factor_2[14]+factor_2[15])/4);Time se
criteria 4		explvar_factor_3[60]';Time series;default resolution;normal
criteria 5		factor_4[44]';Profiles;-
criteria 6		factor_5[43]';Profiles;-

Figure 63 : Liste de quelques critères de sélection

4.3 « Multi-time PMF »

La fonctionnalité multi-time est utile pour combiner des données de différentes résolutions temporelles.

Tout d'abord, l'utilisateur importe les deux jeux de données dans le panneau « Import data » des spécifications de la matrice (les données ACSM et l'autre jeu de données), ainsi que la matrice d'erreur. Il définit ensuite les deux dimensions temporelles.

Dans le deuxième panneau « Pretreatment », l'utilisateur définit les classes en termes de séries temporelles. Il peut importer les deux classes ou les définir à l'aide d'un index. Il en va de même pour le profil afin de définir deux classes.

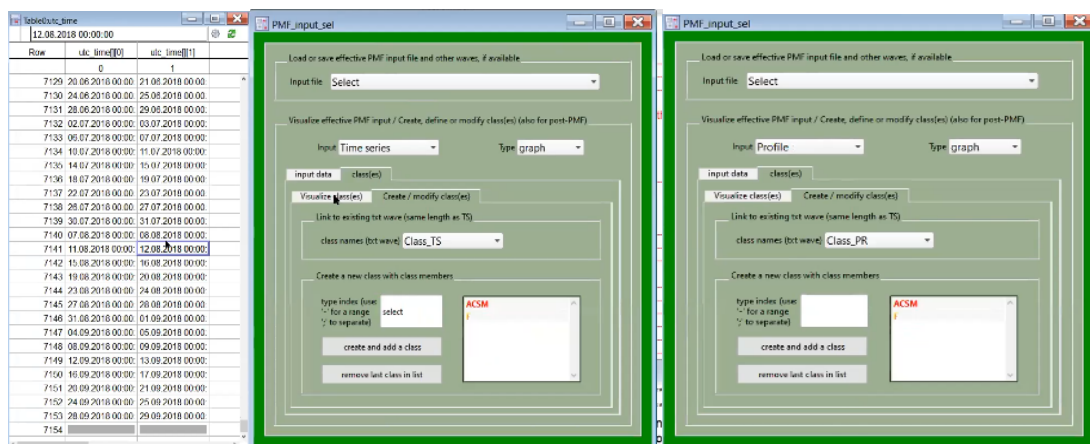


Figure 64 : Panneau de prétraitement ; Préparer l'entrée PMF

L'utilisateur peut ensuite activer l'option multi-time.

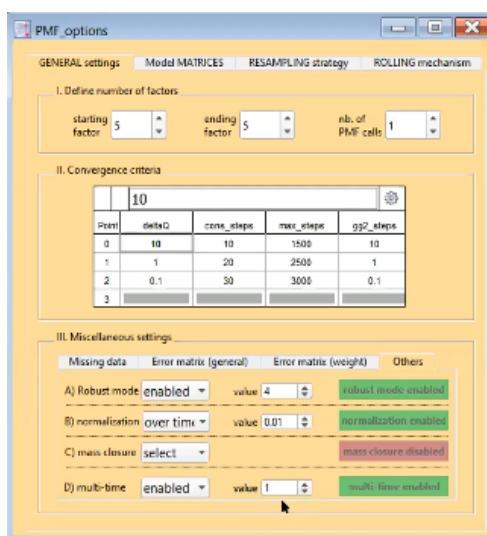


Figure 65 : Panneau d'exécution PMF ; Définir les options du modèle

L'utilisateur peut ensuite exécuter la PMF et vérifier le résultat.

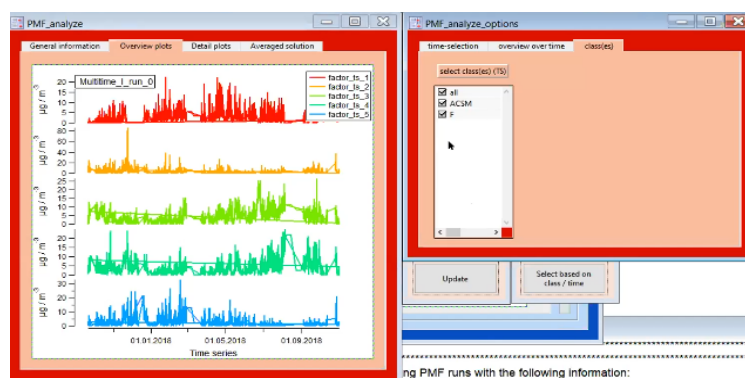


Figure 66 : Plot général

L'utilisateur peut vérifier les résidus pour les deux jeux de données/classes :

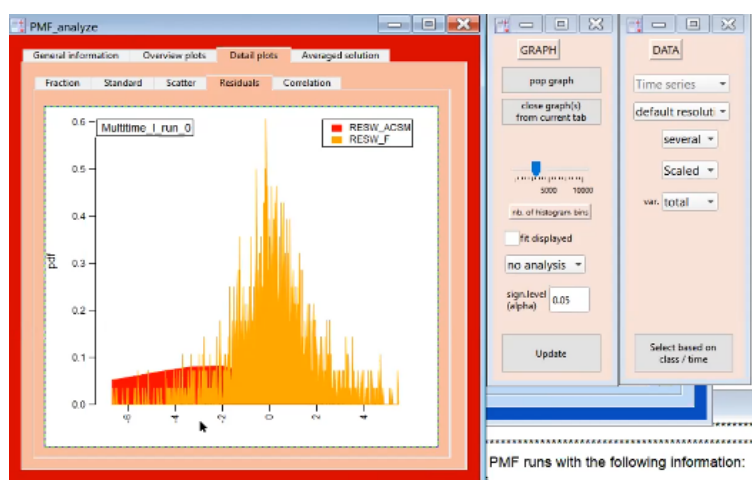


Figure 67 : Panneau des tracés détaillés ; onglet Résidus

Pour corriger ce problème et obtenir la même distribution sous-jacente pour les résidus (scaled), l'utilisateur peut utiliser l'approche C-value (voir le manuel SoFi, pages 117-118).

5. Références

Canonaco, Francesco, Anna Tobler, Gang Chen, Yulia Sosedova, Jay Gates Slowik, Carlo Bozzetti, Kaspar Rudolf Daellenbach, et al. 2021. « 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 ». *Atmospheric Measurement Techniques* 14 (2): 923-43.

Development, Office of Research &. s. d. « EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide ». Consulté le 5 juillet 2021. https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NERL&direntryid=308292
« Manual - Datalystica Knowledge Base ». s. d. Consulté le 5 juillet 2021. <https://datalystica.helpscoutdocs.com/article/19-manual>.

Paatero, Pentti, et Philip K. Hopke. 2003. « Discarding or Downweighting High-Noise Variables in Factor Analytic Models ». *Analytica Chimica Acta*, Papers presented at the 8th International Conference on Chemometrics and Analytical Chemistry, 490 (1): 277-89. [https://doi.org/10.1016/S0003-2670\(02\)01643-4](https://doi.org/10.1016/S0003-2670(02)01643-4).

Polissar, Alexandr V., Philip K. Hopke, Pentti Paatero, William C. Malm, et James F. Sisler. 1998. « Atmospheric Aerosol over Alaska: 2. Elemental Composition and Sources ». *Journal of Geophysical Research: Atmospheres* 103 (D15): 19045-57. <https://doi.org/10.1029/98JD01212>.

Ulbrich, I M, M R Canagaratna, Q Zhang, D R Worsnop, et J L Jimenez. 2009. « Interpretation of Organic Components from Positive Matrix Factorization of Aerosol Mass Spectrometric Data ». *Atmos. Chem. Phys.*, 28.

6. Annexe

Macro of Blacklist Button

```
#pragma rtGlobals=3          // Use modern global access method and strict wave access.
Macro BL_Button()
    PauseUpdate; Silent 1    // building window...
    ShowInfo
    //ShowTools
        controlinfo kwcontrolbar
    Variable width = 55, height = 22
    Variable line = v_height, col_1 = 0 * width+5 , col_2 = width + 5, col_3 = 2 * width + 5 , col_4
= 3 * width + 5,col_5 = 4 * width + 15

    ControlBar line + height * 1 // this 1 means there is only one line of buttons right now
        Button BL_but, pos={463,2}, size={80,20},title="Blacklist",proc=BLWindowProc
EndMacro
```

Macro of Clean Spikes Graphs

```
Window CleanUpSpikes() : Graph
    PauseUpdate; Silent 1    // building window...
    String fldrSav0= GetDataFolder(1)
    SetDataFolder root:SelectedIon_Diagnostic:
    Display /W=(766.5,336.5,1410.75,591.5)/L=L_18/B=B_28_18 mz18 vs mz28
    AppendToGraph/L=L_32/B=B_28_32 mz32 vs mz28
    AppendToGraph/L=L_40/B=B_28_40 mz40 vs mz28
    SetDataFolder fldrSav0
    ModifyGraph margin(bottom)=85
    ModifyGraph mode=3
    ModifyGraph marker=8
    ModifyGraph
    zColor(mz18)={:ACSM_Incoming:acsm_utc_time,*,*,Rainbow},zColor(mz32)={:ACSM_Incom
ing:acsm_utc_time,*,*,Rainbow}
    ModifyGraph zColor(mz40)={:ACSM_Incoming:acsm_utc_time,*,*,Rainbow}
    ModifyGraph standoff=0
    ModifyGraph notation=1
    ModifyGraph lblPos(L_18)=45,lblPos(B_28_18)=80,lblPos(L_32)=47,lblPos(B_28_32)=80
    ModifyGraph lblPos(L_40)=40,lblPos(B_28_40)=80
    ModifyGraph zapTZ=1
    ModifyGraph prescaleExp=12
    ModifyGraph freePos(L_18)=0
    ModifyGraph freePos(B_28_18)={0,L_18}
    ModifyGraph freePos(L_32)={0.35,kwFraction}
    ModifyGraph freePos(B_28_32)={0,L_32}
    ModifyGraph freePos(L_40)={0.7,kwFraction}
    ModifyGraph freePos(B_28_40)={0,L_40}
    ModifyGraph axisEnab(B_28_18)={0,0.25}
    ModifyGraph axisEnab(B_28_32)={0.35,0.6}
    ModifyGraph axisEnab(B_28_40)={0.7,1}
    Label L_18 "\um/z 18"
    Label B_28_18 "\um/z 28"
    Label L_32 "\um/z 32"
    Label B_28_32 "\um/z 28"
    Label L_40 "\um/z 40"
```

```

Label B_28_40 "\\um/z 28"
SetAxis/A=2/E=2 L_18
SetAxis/A/E=2 B_28_18
SetAxis/A=2/E=2 L_32
SetAxis/A/E=2 B_28_32
SetAxis/A=2/E=2 L_40
SetAxis/A/E=2 B_28_40
ColorScale/C/N=text0/F=0/B=1/A=MC/X=0.13/Y=-76.78 trace=mz18, vert=0, height=10
ColorScale/C/N=text0 widthPct=70, lblMargin=0
AppendText "UTC time"
EndMacro

```

Procedure of Clean Spikes

```

//Function adapted from the Aerodyne ACSM_SpikeRemover.ipf functions by V. Riffault (IMT
Nord Europe)
// to mimic the manual cleaning of data using the Selected Ion Plot in ACSM Local and stay
consistent with previous manual procedure
/////////////////////////////////////////////////////////////////
// Better be cautious and save the experiment before blacklisting too many points! //
/////////////////////////////////////////////////////////////////
Function ACSM_CleanSpikes(mz,deltapos,deltaneg)
variable deltapos,deltaneg, mz
// mz's should typically be air beam ions (28, 32,40) or water signal at mz18 (which can show
more variation so exert carefully!)
// deltapos and deltaneg are positive values in the function.
// The positive and negative values for delta
// The 2 deltas are mostly to account for differences in the variation of mz18, while mz28, mz32
and mz40 should exhibit a symmetrical level of noise. They must be set cautiously, according
to the level of noise for each mz.
wave mz18 = root:SelectedIon_Diagnostic:mz18
wave mz28 = root:SelectedIon_Diagnostic:mz28
wave mz32 = root:SelectedIon_Diagnostic:mz32
wave mz40 = root:SelectedIon_Diagnostic:mz40
wave time_w = root:Acsm_incoming:acsm_utc_time
variable i
switch(mz)
  case 18:
    for (i=0; i<numpts(time_w); i += 1)
      if ((mz18[i] > deltapos) || (mz18[i] < -deltaneg))
        BlackList(i,i)
        i=i-1
      endif
    endfor
    break
  case 28:
    for (i=0; i<numpts(time_w); i += 1)
      if ((mz28[i] > deltapos) || (mz28[i] < -deltaneg))
        BlackList(i,i)
        i=i-1
      endif
    endfor
    break
  case 32:
    for (i=0; i<numpts(time_w); i += 1)

```

```

        if ((mz32[i] > deltapos) || (mz32[i] < -deltaneg))
            BlackList(i,i)
            i=i-1
        endif
    endfor
    break
case 40:
    for (i=0; i<numpts(time_w); i += 1)
        if ((mz40[i] > deltapos) || (mz40[i] < -deltaneg))
            BlackList(i,i)
            i=i-1
        endif
    endfor
    break
default:
    print "Wrong mz. You must choose 18, 28, 32 or 40"
    break
endswitch
SIMZP_Make()
end

```

Procedure of RIT Time series

Window RIT_TS() : Graph

```

PauseUpdate; Silent 1      // building window...
String fldrSav0= GetDataFolder(1)
SetDataFolder root:RIT:
Display /W=(450,56,1413,439.25) RIT_Mat[*][1] vs ::ACSM_Incoming:acsm_utc_time
AppendToGraph RIT_Mat[*][3] vs ::ACSM_Incoming:acsm_utc_time
AppendToGraph RIT_Mat[*][4] vs ::ACSM_Incoming:acsm_utc_time
AppendToGraph RIT_Mat[*][2] vs ::ACSM_Incoming:acsm_utc_time
SetDataFolder fldrSav0
ModifyGraph mode=2
ModifyGraph ISize=1.5
ModifyGraph
rgb(RIT_Mat)=(43520,43520,43520),rgb(RIT_Mat#1)=(0,0,65280),rgb(RIT_Mat#3)=(0,65280,0)
ModifyGraph lblMargin(left)=15
ModifyGraph dateInfo(bottom)={0,0,0}
Label left "RIT"
Label bottom " "
ShowInfo
Legend/C/N=text0/J/H={0,5,10}/X=1.68/Y=-1.22 "\\s(RIT_Mat) mz 62\r\\s(RIT_Mat#3) mz 76\r\\s(RIT_Mat#1) mz 102\r\\s(RIT_Mat#2) mz 128"
ControlBar 22
Button BL_but,pos={463,2},size={80,20},proc=BLWindowProc,title="Blacklist"
EndMacro

```

Procedure of CDCE application

```
#pragma rtGlobals=3      // Use modern global access method and strict wave access.
```

```
Function CalcCE_fPhase(NH4_DL,CE_lowNH4)
Variable NH4_DL,CE_lowNH4
```

```
// Prior to running this procedure, all species must be calculated using the same CE.
// The sampling line relative humidity (if measured) should be named "RH_SL"
```

```
SetDataFolder root:
```

```
wave RH_SL
```

```
wave CE=root:CE
```

```
// Create waves of each species to smooth for the calculations.
```

```
duplicate/o root:time_series:SO4 SO4_raw
```

```
duplicate/o root:time_series:NH4 NH4_raw
```

```
duplicate/o root:time_series:NO3 NO3_raw
```

```
duplicate/o root:time_series:Chl Chl_raw
```

```
duplicate/o root:time_series:Org Org_raw
```

```
Variable i
```

```
Duplicate/o root:time_series:SO4 PredNH4_CE1, NH4_MeasToPredict, ANMF
```

```
Duplicate/o root:time_series:SO4 CE_dry,CE_fPhase
```

```
CE_dry=0.5
```

```
CE_fPhase=0.5
```

```
// Equation 3
```

```
PredNH4_CE1=18*(SO4_raw/96*2+NO3_raw/62+Chl_raw/35.45)
```

```
NH4_MeasToPredict=NH4_raw/PredNH4_CE1
```

```
// Equation 5
```

```
ANMF=(80/62)*NO3_raw/(NO3_raw+SO4_raw+NH4_raw+Org_raw+Chl_raw)
```

```
// Calculate the dry collection efficiency, CE_dry
```

```
For (i=0;i<(numpts(SO4_raw));i+=1)
```

```
// Nan negative NH4_MeasToPredict points
```

```
    If (NH4_MeasToPredict[i]<0)
```

```
        NH4_MeasToPredict[i]=nan
```

```
    EndIf
```

```
// Nan ANMF points if negative or more than 1
```

```
    If (ANMF[i]<0)
```

```
        ANMF[i]=nan
```

```
    Elseif (ANMF[i]>1)
```

```
        ANMF[i]=nan
```

```
    EndIf
```

```
    If (PredNH4_CE1[i]<NH4_DL)
```

```
// In general, do not calculate CE for these points.
```

```
        CE_dry[i]=0.5
```

```
// In the CE paper, applied CE for low ammonium mass
```

```
// CE_dry[i]=CE_lowNH4
```

```
        Elseif (NH4_MeasToPredict[i]>=0.75)
```

```
// Apply Equation 4
```

```
            CE_dry[i]= 0.0833+0.9167*ANMF[i]
```

```
        Elseif (NH4_MeasToPredict[i]<0.75)
```

```
// Apply Equation 6
```

```
            CE_dry[i]= 1-0.73*NH4_MeasToPredict[i]
```

```
        EndIf
```

```
EndFor
```

```
// Make CE_dry between 0.5 and 1
```

```
CE_dry=min(1,(max(0.5,CE_dry)))
```

```

If (WaveExists(RH_SL)==1)
// Apply Equation 7
    CE_fPhase=(5*CE_dry-4)+(1-CE_dry)/20*RH_SL
    For (i=0;i<numpnts(CE_fPhase)+1;i+=1)
        If (RH_SL[i]<80 || numtype(RH_SL[i])==2 )
            CE_fPhase[i]=CE_dry[i]
        EndIf
    EndFor
Else
    CE_fPhase=CE_dry

EndIf

Edit root:ACSM_Incoming:acsm_utc_time, CE_dry, CE_fPhase

KillWaves PredNH4_CE1

// Create CE-corrected waves for each species and their sum
SetDataFolder root:Time_Series:
wave NRPM1_CEcorr
Make/O/N=(numpnts(Org_raw)) Org=Org_raw*CE[0]/CE_fPhase
Make/O/N=(numpnts(Org_raw)) SO4=SO4_raw*CE[0]/CE_fPhase
Make/O/N=(numpnts(Org_raw)) NH4=NH4_raw*CE[0]/CE_fPhase
Make/O/N=(numpnts(Org_raw)) NO3=NO3_raw*CE[0]/CE_fPhase
Make/O/N=(numpnts(Org_raw)) Chl=Chl_raw*CE[0]/CE_fPhase

Make/O/N=(numpnts(Org)) NRPM1_CEcorr=Org+SO4+NH4+NO3+Chl

End

```

Le Laboratoire Central de Surveillance de la Qualité de l'Air est un groupement d'intérêt scientifique créé en 2005 et constitué des laboratoires de l'IMT Nord Europe, de l'Ineris et du LNE. Depuis 1991, ses équipes d'experts mènent des études et des recherches sur la qualité de l'air, en appui au ministère chargé de l'environnement, et en étroite collaboration avec les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA). Le LCSQA est désigné par le ministère pour assurer la coordination technique nationale du dispositif de surveillance de la qualité de l'air en France. Il est également l'organisme de référence requis par les directives européennes et ses missions sont définies par l'arrêté du 16 avril 2021 relatif au dispositif national de surveillance de la qualité de l'air ambiant.

L'objectif principal du LCSQA est ainsi de contribuer à l'amélioration continue du dispositif de surveillance en apportant un appui scientifique et technique au ministère et aux AASQA.

Ses travaux portent, d'une part, sur l'ensemble de la chaîne de surveillance : la qualité des mesures en air ambiant (du prélèvement à l'analyse), les inventaires d'émissions, la modélisation des concentrations et leur cartographie, la centralisation ainsi que le traitement et la mise à disposition des données, et d'autre part, sur les méthodes d'évaluation des plans d'actions. Cette action s'inscrit à la fois dans le cadre des réglementations nationales et européennes, et dans une démarche prospective visant à anticiper les besoins de la surveillance de demain.

Direction du LCSQA

Ineris - Parc technologique Alata 2 -
BP2 - F60550 Verneuil-en-Halatte
Tél : 03 44 55 69 16

www.lcsqa.org