



Sources de polluants dans les lichens

Estimation des contributions par méthode PMF

2025

REMERCIEMENTS

L'Institut Écocitoyen adresse ses remerciements amicaux à Amandine Durand (Laboratoire Chimie Environnement, Aix-Marseille Université), ainsi qu'à Emmanuel Wafo et Florence Chaspoul (Faculté de Pharmacie, Aix-Marseille Université) pour leurs accompagnements techniques dans les analyses chimiques.

L'Institut Écocitoyen remercie également l'ensemble des villes qui ont accueilli les prélèvements (voir Figure 1), ainsi que la Réserve Naturelle des Marais du Vigueirat, le Parc Naturel Régional de Camargue, et le Grand Port Maritime Marseille-Fos.

Les travaux présentés dans ce rapport ont été réalisés avec le soutien financier de la Métropole Aix-Marseille-Provence, de l'ADEME (dans le cadre du projet EVALVIE), de la ville de Port-Saint-Louis-du-Rhône (dans le cadre du projet CARE) et de la ville de Fos-sur-Mer.



Le présent rapport est publié sous licence CC-BY-ND



Les données, les interprétations et les images présentes dans ce rapport sont la propriété de l'Institut Écociroyen pour la Connaissance des Pollutions, libres d'accès et d'utilisation à condition de citer les références du rapport et d'indiquer la source des données, photos et graphiques ("Institut Écociroyen pour la Connaissance des Pollutions - Fos-sur-Mer").

Il ne sera pas apporté, à la version originale du rapport, de modifications à des fins de diffusion. L'Institut Écociroyen pour la Connaissance des Pollutions ne pourra être tenu responsable de toute interprétation réalisée par un tiers et qui n'est pas explicitement détaillée dans ce rapport.

Pour citer ce rapport :

Julien Dron. Contributions des sources de polluants atmosphériques : répartition géographique par l'application de la méthode PMF aux mesures de bioaccumulation dans les lichens. Institut Écociroyen pour la Connaissance des Pollutions, 2025.

Contact :

julien.dron@institut-ecociroyen.fr

*Institut Écociroyen pour la Connaissance des Pollutions
RD268 – Centre de vie la Fossette – 13270 Fos-sur-Mer*

04.90.55.49.94 – www.institut-ecociroyen.fr

SYNTHÈSE

Les travaux présentés dans ce rapport présentent une **méthodologie pour appliquer le modèle statistique multivarié PMF** (Positive Matrix Factorization) aux concentrations de polluants mesurées dans les lichens sur le territoire de la Métropole Aix Marseille Provence, afin d'y **estimer les contributions des sources de pollution**. La méthodologie mise au point a été **appliquée à la campagne « Lichens 2021 »**, pour laquelle ont été retenus 30 paramètres (contaminants, marqueurs) et 41 observations (sites de prélèvements), soit un total de 1 230 paires observations-paramètres.

- La **faisabilité de l'utilisation du modèle PMF** aux mesures de contaminants dans les lichens a été démontrée, et les précautions à adopter ont été indiquées.
- L'utilisation du modèle PMF sur les données de la campagne « Lichens 2021 » a fourni des résultats cohérents et en **bonne adéquation avec les études précédentes sur les sources de polluants atmosphériques dans la région**.
- L'application d'une contrainte sur le lévoglucosan, traceur de la combustion de biomasse, a permis de dégager une **signature marquée « combustion de biomasse »**, ce que ne permettait pas d'autres méthodes statistiques multivariées (ACP).
- Le modèle PMF a permis de distinguer **deux signatures industrielles**, dont une spécifique à la Zone Industriale-Portuaire (ZIP) de Fos.

- La **méthodologie statistique mise au point et utilisant le modèle PMF appliqué aux mesures de bioaccumulation dans les lichens permet de visualiser géographiquement les contributions des sources des contaminants**.

- Sur les données « Lichens 2021 », la **contribution des sources industrielles identifiées apparaît globalement la plus importante (37,2 % en faisant la somme des 2 facteurs industriels), suivi de la combustion de biomasse (26,9 %)**.

- La **répartition géographique des plus fortes contributions industrielles** est cohérente avec les activités industrielles connues dans la région : **ZIP de Fos (Fos-sur-Mer, Port-Saint-Louis-du-Rhône, Martigues, Port-de-Bouc) et ses alentours (Châteauneuf-les-Martigues, Vitrolles, Gardanne)**.

- La **répartition géographique des contributions du facteur « combustion de biomasse »** est aussi cohérente, homogène et, en relatif, plus importante **dans les zones rurales et urbaines**.

Ces résultats ouvrent de nombreuses perspectives :

- Appliquer la méthode aux campagnes 2017 et 2024 lors desquelles le lévoglucosan a été mesuré. Un plus grand nombre d'observations augmentera la fiabilité des résultats.
- Évaluer les points d'incertitudes ou d'optimisation, et tester des solutions pour améliorer la fiabilité et les capacités de la modélisation (contraintes supplémentaires, meilleure représentation de zones spécifiques...).
- Tenter d'identifier d'autres types de sources, comme les transports routiers ou maritimes.
- Évaluer l'évolution temporelle depuis 2017 (début des mesures du lévoglucosan) des contributions des sources identifiées.

SOMMAIRE

REMERCIEMENTS.....	1
SYNTHÈSE.....	3
1. INTRODUCTION.....	7
2. MÉTHODES.....	9
2.1. Principes de la biosurveillance lichénique par bioaccumulation des contaminants.....	9
2.2. Le programme LICHENS de l'Institut Écocitoyen.....	9
2.3. Données d'entrée.....	9
2.4. Analyses par composantes principales (ACP).....	10
2.5. Factorisation matricielle positive (PMF).....	10
2.6. PMF : paramètre « total » & incertitudes sur les données d'entrée.....	11
2.7. PMF : Choix du nombre de facteurs et application de contraintes.....	12
2.8. PMF : Calculs de l'incertitude statistique.....	13
2.9. Essais exploratoires et conduite des modélisations.....	13
3. CHOIX DU MODÈLE PMF : NOMBRE DE FACTEURS ET CONTRAINTES.....	15
4. JUSTESSE ET ERREURS RÉSIDUELLES DU MODÈLE PMF.....	17
5. RÉSULTATS DE LA MODÉLISATION PMF.....	23
5.1. Signatures des sources (facteurs PMF).....	23
5.2. Contributions des sources : répartition spatiale des facteurs PMF.....	25
6. COMPARAISON ENTRE MODÈLES PMF ET ACP.....	29
6.1. Analyse par composantes principales (ACP).....	29
6.2. Comparaison entre les résultats des modèles PMF et ACP.....	31
RÉFÉRENCES.....	33
ANNEXES.....	35

1. INTRODUCTION

Le territoire de l'ouest des Bouches-du-Rhône est marqué par la présence d'industries lourdes telles que la sidérurgie, la pétrochimie, le raffinage, l'industrie chimique ou encore l'incinération de déchets (zones industrialo-portuaires de Fos-sur-Mer, de Martigues-Lavera, La Mède et Berre-l'Étang, notamment). En outre, ces activités industrielles et portuaires induisent des flux logistiques, exposant une zone déjà très urbanisée aux émissions atmosphériques issues des transports maritimes et routiers. Dans l'étude de l'exposition aux polluants atmosphériques du territoire de la Métropole Aix-Marseille-Provence, où l'activité économique entraîne ainsi des émissions importantes et très diversifiées dans l'atmosphère, la question de la contribution des différentes sources d'émissions est prépondérante pour répondre aux besoins d'amélioration environnementale.

La caractérisation des sources contribuant aux niveaux de contaminants dans un site dit « récepteur » est essentiellement réalisée par des techniques instrumentales, telles que :

- Des prélèvements d'air sur filtre. L'Institut Écocitoyen, en partenariat avec Atmosud et le Laboratoire Chimie Environnement (LCE, Aix-Marseille Université), avait réalisé une étude en 2013-2014 utilisant cette méthodologie [Sylvestre 2016].
- Des mesures en ligne de la matière organique, comme expérimenté à Fos-Carabins en 2018-2019 [Benadda 2021] et depuis 2017 également à Marseille [Chazeau 2021], de métaux comme récemment employés à Marseille [Camman et al. 2024], ou encore de composés organiques volatils (COV) notamment expérimentés autour de Dunkerque [Zhang 2016].

Les résultats obtenus par ces techniques sont ensuite traités par méthodes d'analyses statistiques multivariées, basées sur des régressions linéaires multiples et des analyses en composantes principales (ACP), telles que PMF (« positive matrix factorization ») et son évolution ME-2 (« multilinear engine ») ou CMB (« chemical mass balance »).

Les prélèvements sur filtres ont en général une résolution temporelle de 24 h, mais les analyseurs en ligne permettent d'accéder à une très grande résolution temporelle (inférieure à 1 h).

Cependant, ces méthodes dites « instrumentales » font appel à du matériel coûteux et requièrent une logistique importante liée à la maintenance et l'entretien des appareils, l'accès électrique et des locaux climatisés, notamment. Par conséquent, elles ne peuvent être appliquées que sur un nombre de sites très restreint et souvent sur des périodes limitées. Elles ne permettent donc pas de cartographier les contributions des différentes sources sur un territoire.

Afin de compléter les mesures instrumentales, la biosurveillance est de plus en plus utilisée. Depuis 2011, l'Institut Écocitoyen réalise des campagnes de prélèvements de lichens, reconnus comme bioindicateurs de la qualité de l'air et bioaccumulateurs des polluants présents dans l'air ambiant. La dernière campagne s'est déroulée en 2021, sur 41 sites étudiés répartis sur la Métropole

Aix-Marseille-Provence, de Marseille à Port-Saint-Louis-du-Rhône, de Gardanne à Saint-Martin-de-Crau, en passant par Fos-sur-Mer, Istres, Martigues ou encore Berre-l'Étang. Les résultats obtenus ont été publiés récemment [Reuillard 2023].

La caractérisation des sources pourrait prendre une dimension spatiale en exploitant les résultats obtenus dans les lichens, qui portent sur de nombreuses familles de contaminants (métaux, HAP, PCB, dioxines/furanes) et des traceurs de la combustion de bois (lévoglucosan, produit par combustion de la cellulose). Comme cela n'a jamais été réalisé, des travaux de recherches scientifiques ont été nécessaires pour accéder à ce type d'interprétation. Ils sont présentés dans ce rapport, avec les premiers résultats cartographiés pour l'année 2021.

2. MÉTHODES

2.1. Principes de la biosurveillance lichénique par bioaccumulation des contaminants

La biosurveillance repose sur l'idée que les organismes vivants réagissent de manières différentes aux variations de leur environnement et donc en reflètent les changements et les altérations. Parmi les différentes méthodes de biosurveillance, la bioaccumulation consiste à déterminer les teneurs en éléments chimiques accumulés dans un ou plusieurs organismes vivants en réponse à une exposition plus ou moins diffuse. Dans le domaine de la qualité de l'air, de nombreux végétaux sont utilisés. Cependant, les lichens occupent une place particulière par leurs caractéristiques physiologiques qui en font de véritables « échantillonneurs passifs ».

Les lichens sont notamment dépourvus de racines et de système d'excrétion. Ils ont aussi une activité tout au long de l'année et ne sont pas capables de réguler leurs échanges gazeux avec l'atmosphère, ce qui les expose continuellement aux polluants atmosphériques sans influence de leur support. L'ensemble des éléments chimiques présents dans l'air est susceptible d'être adsorbé par les lichens, parmi lesquels des contaminants ou d'autres composés, comme le lévoglucosan qui est un traceur spécifique de la combustion du bois, issu de la combustion de la cellulose. Si les concentrations dans l'atmosphère changent, ces variations seront répercutées dans les lichens progressivement avec un délai de plusieurs mois (temps d'intégration). Le processus de bioaccumulation résulte d'un équilibre entre les concentrations atmosphériques et lichéniques, les niveaux accumulés dans les lichens sont ainsi représentatifs de la pollution de l'air.

2.2. Le programme LICHENS de l'Institut Écocitoyen

L'Institut Écocitoyen a mis en place depuis 2011, un programme de suivi de la bioaccumulation des contaminants par les lichens, qui ont été mesurés en 2011, 2012, 2013, 2014, 2017 et 2021 [Dron et al. 2016; Dauphin et al. 2020; Reuillard 2023]. Par ailleurs, des travaux de recherches pour mieux estimer le temps d'intégration des contaminants par les lichens ont montré une durée de 6 à 12 mois au moins [Dron et al. 2021], ainsi que pour mettre au point une méthode d'analyse du lévoglucosan [Dauphin et al. 2020]. Si une estimation des sources avait déjà pu être faite sur la campagne 2014 [Ratier et al. 2018], la mesure du lévoglucosan et l'utilisation de méthodes statistiques plus avancées ouvrent de nouvelles voies pour déterminer de façon plus précise et plus cohérente la contribution des divers types d'émissions à la présence des contaminants chimiques dans les lichens, et donc dans l'air ambiant.

Les travaux présentés ici portent sur la mise au point d'une méthode de détermination de la contribution des sources de façon plus précise et cartographiée. Ils exploitent les résultats des analyses de bioaccumulation dans les prélèvements réalisés lors de la campagne 2021. La localisation, les conditions météorologiques, et autres éléments de contexte, sont détaillés dans le rapport dédié à cette campagne [Dauphin et al. 2020; Reuillard 2023].

2.3. Données d'entrée

La technique d'analyse statistique par PMF du logiciel USEPA PMF 5.0, comme l'analyse ACP, permettent d'inclure autant de paramètres ou variables que souhaités (contaminants, traceurs...). Les analyses menées sur les prélèvements de 2021 ont porté sur un grand nombre de contaminants : 20 éléments métalliques ou métalloïdes, 17 dioxines/furanes (PCDD-F), 40 PCB, 16 HAP, et 1 traceur spécifique de la combustion de bois (lévoglucosan).

Les pesticides ont été exclus de l'analyse car ils n'ont pas été déterminés sur la totalité des prélèvements, ce qui peut entraîner des biais importants dans l'analyse. De plus, nous n'avons aucun recul sur ces mesures, à la fois au niveau local comme au niveau national ou international vu

qu'il s'agissait d'une totale innovation scientifique, jamais réalisée auparavant [Durand et al. 2024].

Nous disposons donc de 94 paramètres. Or même si l'analyse statistique admet un nombre quasi illimité de paramètres, il convient de se limiter à un nombre raisonnablement interprétable. Comme pour les études précédentes par ACP [Ratier et al. 2018], nous avons donc rassemblé :

- Les dioxines et furanes en 2 groupes (furanes PCDF, et dioxines PCDD). Ces deux groupes ont montré des caractéristiques distinctes en fonction des influences industrielles (favorisant les PCDF) et des autres (favorisant les PCDD) [Ratier et al. 2018].
- Les PCB par nombre d'atomes de chlore, soit 6 groupes contenant 3, 4, 5, 6, 7 ou 8 chlores. Le nombre de chlore peut aussi dépendre de l'origine des PCB ou de leur vieillissement dans l'atmosphère [Ratier et al. 2018; Dron et al. 2022].
- Les HAP par nombre de cycles benzéniques, soit 5 groupes contenant 2, 3, 4, 5 ou 6 cycles. Comme pour les PCB, l'origine des HAP (notamment la température de combustion, le type de carburant...) a une influence sur le nombre de cycles, qui est donc une caractéristique potentiellement intéressante à interpréter [Ratier et al. 2018; Reuillard 2023].
- Les 16 métaux qui ont été détectés dans au moins 80 % des sites (sur 20 recherchés) ont été conservés, pour maintenir de bonnes conditions statistiques. Les métaux exclus sont le mercure (Hg), l'étain (Sn), l'argent (Ag) et le thallium (Tl).

Nous obtenons ainsi un nombre réduit de 30 paramètres, bien équilibrés entre métaux (16) et contaminants organiques (13), ainsi que le traceur de combustion de biomasse (lévoglucosan). Chacun de ces paramètres a été détecté dans au moins 80 % des 41 sites de prélèvements de la campagne 2021.

Certains sites n'ont pas bénéficié des mesures de PCDD-F (MS3, MN1, SME, SMO, PSL-5, PSL-9, PSL-13, PSL-12) ou de lévoglucosan (MS3, MN1). Nous leur avons attribué les valeurs des médianes calculées sur les autres sites.

2.4. Analyses par composantes principales (ACP)

Dans un premier temps, des analyses par composantes principales (ACP) ont été réalisées, de la même manière que sur les résultats de la campagne 2014. L'objectif de ces ACP préliminaires est à la fois de vérifier les résultats obtenus dans les mêmes conditions qu'antérieurement [Ratier et al. 2018], mais aussi d'avoir une première idée des signatures et du nombre de composantes (ou facteurs) pertinent(e)s. Les analyses ACP ont été réalisées avec le logiciel R (version 4.1.2, [R Core Team 2021]) et la fonction « prcomp » (package « stats »).

2.5. Factorisation matricielle positive (PMF)

Les analyses statistiques avec le modèle PMF (Factorisation matricielle positive) ont été réalisées avec le logiciel PMF 5.0 mis à disposition par USEPA (<https://www.epa.gov/air-research/positive-matrix-factorization-model-environmental-data-analyses>) et dont les principes statistiques sont décrits dans le guide d'utilisation ([Norris et al. 2014]). Le modèle PMF repose sur des bases semblables à celles de l'ACP, et sont tous les deux des méthodes d'analyses statistiques multivariées (nombre de variables > 2). Comme pour l'ACP, il en résulte des facteurs (similaires aux composantes de l'ACP). Cependant, le modèle impose des résultats positifs, plus faciles à matérialiser d'un point de vue environnemental.

Il fournit aussi la possibilité de contraindre le modèle notamment en déterminant un nombre de facteurs prédéfini, de forcer certains traceurs sur des facteurs (...). L'application de ces contraintes est réalisée au prix d'une erreur statistique non-négligeable qui peut être calculée, au contraire de l'ACP qui fournira autant de composantes que nécessaire pour approcher une erreur statistique

nulle. Il convient de noter qu'une erreur statistique nulle ne garantit pas la justesse des résultats ! Dans les sections suivantes de la méthodologie, nous décrivons les principaux paramètres et aspects qui doivent être considérés dans l'analyse PMF.

2.6. PMF : paramètre « total » & incertitudes sur les données d'entrée

L'analyse statistique PMF du logiciel USEPA PMF 5.0 requiert d'indiquer les incertitudes sur chacun des paramètres d'entrée et les limites de détection. Les incertitudes incluant le prélèvement et la préparation de l'échantillon ainsi que l'analyse chimique ont été déterminées par un travail de recherche de l'Institut Écocitoyen, et sont de 30 % pour les contaminants organiques et les métaux [Dron et al. 2021]. Les limites de détection sont indiquées dans le rapport de la campagne 2021 [Reuillard 2023].

L'analyse statistique PMF permet aussi d'inclure un paramètre « total ». En chimie atmosphérique, c'est généralement la masse des particules PM1, PM2.5 ou PM10, dans lesquelles sont mesurés les contaminants et traceurs, qui est renseignée comme paramètre « total ». Dans le cas des lichens, la masse totale de lichens n'a aucun sens physique pour la détermination des sources. Nous avons donc choisi de considérer la somme des concentrations normalisées (ou « réduites », c'est à dire pondérées par l'écart-type) de tous les contaminants (métaux, HAP, PCDD-F, PCB) et traceurs (lévoglucosan) mesurés, pour chacune des observations (site). Chaque valeur de concentration est ainsi divisée (réduite ou normalisée) par l'écart-type de la variable, permettant d'avoir des valeurs de même ordre de grandeur pour tous les éléments, quelle que soit sa gamme de concentrations. Ainsi, pour chacun des sites, la somme pondérée (S_j) est calculée selon l'équation suivante,

$$S_j = \sum_i \frac{\text{concentration de } i \text{ au site } j}{\text{écart-type de } i} \quad (\text{équation 1})$$

i étant un élément (variable) et j un site (observation).

par exemple pour le site CAB, la somme des concentrations réduites s'écrit,

$$S_{CAB} = \frac{\text{conc. Al au site CAB}}{\text{écart-type Al}} + \frac{\text{conc. V au site CAB}}{\text{écart-type V}} + (\dots) + \frac{\text{conc. furanes au site CAB}}{\text{écart-type furanes}} \quad (\text{équation 2})$$

Les sommes des concentrations ainsi « réduites » constituent donc un indicateur de la contamination globale des échantillons. Elles sont représentées de façon cartographique sur la Figure 1. Comme cela a été décrit en détail dans le rapport dédié aux résultats de la campagne 2021 [Reuillard 2023], des niveaux nettement plus élevés sont retrouvés aux abords des grandes zones industrielles de la Métropole (Fos, Gardanne, Martigues). Les zones résidentielles alentours (Fos-sur-Mer,

Port-de-Bouc, Châteauneuf-les-Martigues...) ont des valeurs S_j moins élevées mais relativement proches des grands centres urbains comme Marseille, Vitrolles ou Aix-en-Provence, malgré une urbanisation beaucoup moins dense. Enfin, les sites de référence comme les marais du Vigueirat ou Entressen présentent les niveaux les plus faibles. Il faut toutefois relativiser cette interprétation globale, qui masque d'éventuels « pics » pour certains composés et qui ne prend pas en compte les pesticides analysés lors de cette campagne, avec des niveaux élevés notamment sur Entressen comme cela est bien détaillé par Reuillard [2023]. Quoi qu'il en soit, le terme S_j représente une bonne approximation de l'exposition globale et constitue donc un paramètre « total » pertinent comme le sont les PM10 pour les études de sources basées sur les signatures mesurées dans les particules.



Figure 1. Somme des concentrations normalisées ou « réduites » (paramètre « total » ou S_j dans la modélisation PMF) des contaminants pour chacun des sites étudiés (campagne Lichen 2021). Cartographie Coraline Soul (QGIS)

2.7. PMF : Choix du nombre de facteurs et application de contraintes

Comme indiqué dans la section 2.5., une analyse ACP fait appel à une infinité de facteurs pour approcher une erreur statistique nulle. On ne considère en général que les 4 ou 5 premiers facteurs au maximum, les suivants n'ayant qu'une incidence négligeable (< 5 %) et auxquels on ne peut généralement pas attribuer de signification concrète. Au contraire, dans une analyse PMF, le nombre de facteurs est prédéfini par l'opérateur, et il en résulte une erreur statistique non négligeable. Cependant, cela permet d'attribuer une signification aux facteurs choisis (origine, physico-chimie...). Quoiqu'il en soit, avant d'interpréter les résultats d'une analyse PMF, il convient de faire varier le nombre de facteurs jusqu'à obtenir une erreur statistique acceptable et minimisée, tout en gardant une signification physico-chimique pertinente.

Quel que soit le nombre de facteurs choisis, il est possible en PMF d'y appliquer des contraintes sur les paramètres qui le composent. Elles peuvent apparaître selon différentes formes. Par exemple, un paramètre (contaminant, traceur...) peut être volontairement exclu d'un ou plusieurs facteurs, sur lequel les connaissances préalables indiquent qu'il ne peut pas apparaître. Il peut aussi être maximisé, par exemple si ce paramètre est un traceur d'un facteur caractéristique d'une source particulière. Les contraintes sur les paramètres peuvent revêtir des formes plus complexes, on peut imposer un ratio constant entre 2 paramètres au sein d'un facteur, une égalité, une supériorité, ou une relation mathématique plus complexe.

Dans notre cas, nous considérons un traceur spécifique de la combustion de bois, le lévoglucosan. Par conséquent, si un facteur semble correspondre à cette source, nous pourrions maximiser le traceur lévoglucosan dans ce facteur, et le minimiser dans les autres. Les connaissances acquises au cours des campagnes « LICHENS » depuis 2011, ainsi que l'étude de sources en 2013-2014 et

d'autres travaux, ont aussi permis d'identifier d'autres paramètres qui semblent caractéristiques de certaines sources. Par exemple, un ratio furanes/dioxines élevé semble caractéristique des émissions industrielles [Ratier et al. 2018], et le molybdène (Mo) est très spécifique aux émissions de Lyondell-Fos jusqu'en 2017 au moins [Dron & Chamaret 2015].

2.8. PMF : Calculs de l'incertitude statistique

Le logiciel PMF 5.0 développé par l'USEPA fournit plusieurs méthodes pour estimer les erreurs de la modélisation statistique par PMF. La première, dite des « déplacements d'intervalles » (DISP) prend en compte les ambiguïtés rotationnelles, mais pas les erreurs aléatoires. Cette méthode reste une bonne approche si les incertitudes de mesure sont fiables, ce qui est le cas ici. Pour cette première étude, qui vise à évaluer la faisabilité de la méthode PMF, on se contentera donc de cette méthode DISP.

La méthode « bootstrap » (BS) demande un temps de calcul à l'ordinateur plus important, et il est conseillé de procéder de manière progressive jusqu'à arriver à la précision requise, ce qui nécessite de répéter plusieurs fois l'opération en affinant les paramètres de calcul. Elle prend en compte les erreurs aléatoires et en partie les ambiguïtés rotationnelles, et n'est pas impactée par les incertitudes de mesures. Elle pourra être testée lors d'une future étude.

Enfin, la méthode mixte « BS-DISP », qui prend en compte les deux approches, est la plus robuste, elle profite des avantages des deux méthodes. Cependant, il est nécessaire de réaliser avant les 2 méthodes indépendamment.

2.9. Essais exploratoires et conduite des modélisations

Il est important de noter que sur un ordinateur de bureau aux performances classiques, la durée des calculs entraînés par la modélisation PMF peut aller jusqu'à plusieurs heures selon de premiers essais effectués sur un processeur classique type « i5 » 4 cœurs et 8 Go de RAM, par exemple. Les calculs d'erreur sont généralement plus longs encore et finissent souvent par ne pas aboutir avec ce type de matériel. Par conséquent, nous avons utilisé un ordinateur aux performances supérieures (12 cœurs, 32 Go de RAM) qui a permis d'abaisser le temps de calcul à une quinzaine de minutes au maximum, pour les modélisations réalisées dans le cadre de la présente étude de faisabilité. Quoiqu'il en soit, il convient malgré tout d'avoir réalisé une bonne réflexion en amont afin de minimiser le nombre de calculs !

Suite à une période de prise en main du logiciel (non prise en charge des valeurs absentes – « NA » – qui impose d'attribuer une valeur, ajout des incertitudes, formats de fichiers...), divers essais exploratoires ont été effectués. C'est ainsi qu'il a été constaté que l'absence de facteur total était peu attractive car tous les facteurs sont alors normalisés par le logiciel dans le résultat, avec des contributions moyennes qui sont ainsi égales. Or, il est souhaitable que le sens physique donné aux résultats puisse montrer que certaines origines de contaminations ont un impact global plus faible que d'autres, si c'est le cas. L'attribution de données réduites pour l'ensemble des valeurs a conduit à des résultats incohérents. L'utilisation d'une somme totale également, en raison de la grande disparité des niveaux de concentrations en fonction des contaminants. C'est ainsi que nous avons finalement proposé une somme totale basée sur la somme des données réduites, mais en conservant les valeurs réelles pour les concentrations individuelles des contaminants considérés, comme cela est décrit dans la section 2.6.

Suite à la mise au point exploratoire, il a pu être possible de réaliser des modélisations systématiques avec les données d'entrée de la campagne 2021 ainsi mises en forme. Comme le logiciel ne conserve pas en mémoire ses résultats si le nombre de facteurs peut être modifié à chaque modélisation, nous avons ainsi procédé en augmentant progressivement le nombre de facteurs, en commençant par 3 facteurs.

- a) Réalisation de la première modélisation.
- b) Estimation de l'erreur DISP.
- c) Ensuite, consultation des résultats et tentative d'identifier des signatures pour certains facteurs qui pourraient correspondre à certaines origines de contaminants (industries, combustion de biomasse, trafic maritime, trafic routier...). Le facteur « combustion de biomasse » a été particulièrement scruté afin de poursuivre l'analyse. En effet, la modélisation PMF permettant d'imposer des contraintes, l'identification du facteur « combustion de biomasse » permet de lui attribuer une contrainte maximisant la contribution du lévoglucosan (traceur de la combustion de biomasse) sur ce facteur. Et au contraire de minimiser sa contribution sur les autres facteurs, étant donné qu'il n'est spécifique qu'à cette origine de pollution.
- d) Une fois le facteur « combustion de biomasse » identifié, une contrainte pour lui attribuer une concentration maximisée de lévoglucosan lui est appliquée. Inversement sur les autres facteurs, une contrainte pour minimiser sa contribution est appliquée. Et une seconde modélisation est donc lancée.
- e) Estimation de l'erreur DISP sur cette seconde modélisation.
- f) Consultation des résultats.
- g) Recommencer au a) avec un nombre de facteurs différent.

L'ensemble de ces étapes a ainsi été réalisé pour 3, 4, 5, 6, 7, et 8 facteurs.

3. CHOIX DU MODÈLE PMF : NOMBRE DE FACTEURS ET CONTRAINTES

Afin de décrire au mieux les contributions des différentes origines de pollutions atmosphériques (combustion de biomasse, émissions industrielles, trafic routier...) aux teneurs en contaminants visés ici (métaux, HAP, PCB, PCDD-F, lévoglucosan), il est nécessaire d'étudier et de déterminer le nombre de facteurs adéquat, que l'on fera ensuite correspondre autant que possible à ces différentes origines. Comme indiqué précédemment, dans le modèle PMF, c'est à l'opérateur de déterminer le nombre de facteurs souhaités. Pour chaque modélisation et nombre de facteurs choisis, l'erreur est calculée (on se limitera ici à l'erreur dite « DISP »). Le choix vers le nombre de facteurs qui entraîne une erreur statistique minimale est ainsi optimisé. Ce choix doit également prendre en compte la pertinence du résultat obtenu vis-à-vis des connaissances existantes (sources potentielles, cohérence des signatures...).

Étant donné l'environnement d'étude, au moins 3 origines distinctes peuvent être distinguées, a priori, pour les contaminants recherchés, comme par exemple : les émissions industrielles (potentiellement plusieurs différenciées), les transports (routier, maritime, potentiellement différenciées), la combustion de biomasse (chauffage au bois, brûlages agricoles, incendies), les poussières de sol...

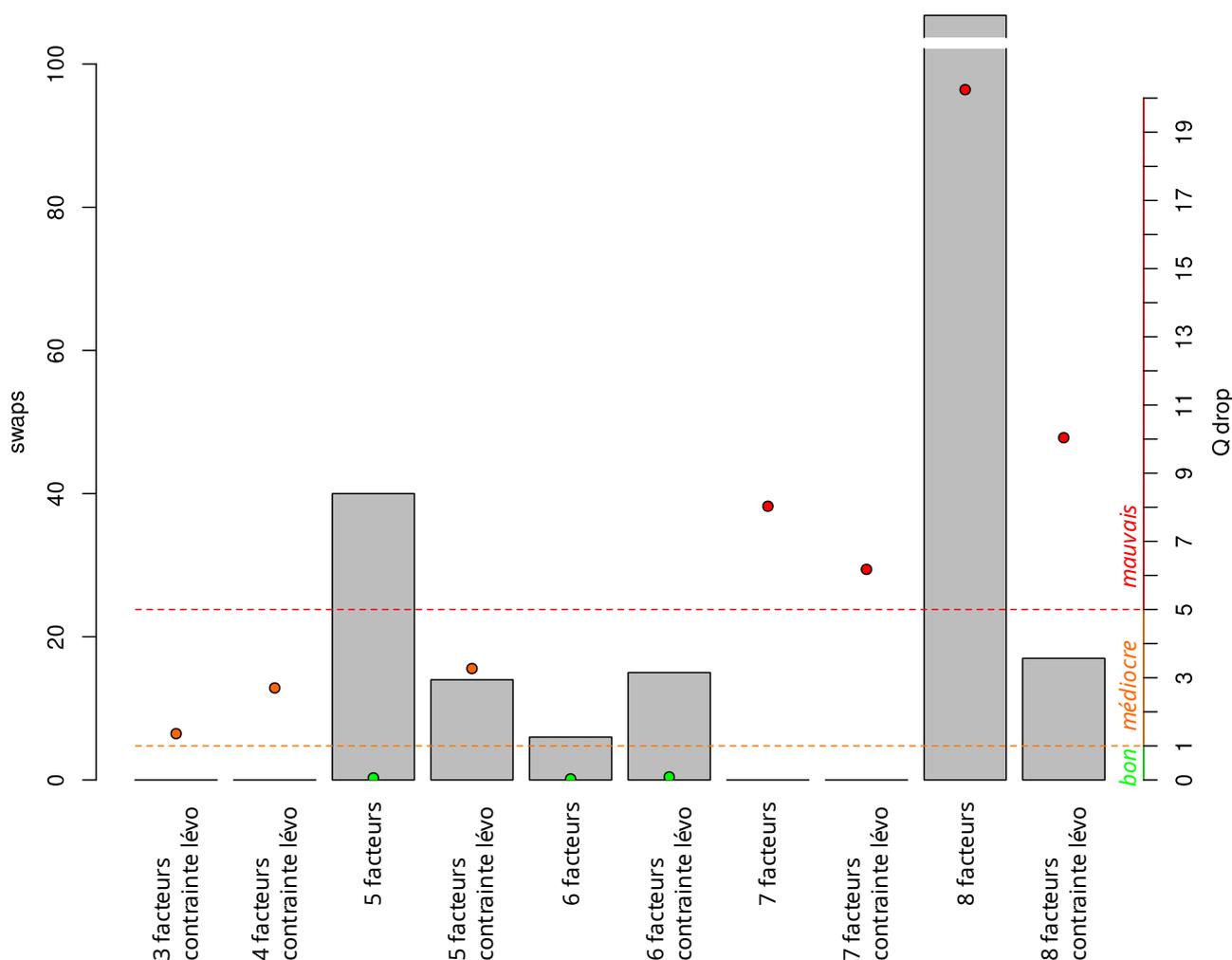


Figure 2. Baisse de la valeur Q (Q drop) et nombre de mélanges (swaps) dans les solutions convergentes obtenues avec 3 à 8 facteurs, avec et sans contrainte sur le lévoglucosan.

Par ailleurs, les résultats obtenus par ACP lors des travaux sur la campagne 2014 ont montré jusqu'à

7 composantes potentiellement pertinentes [Ratier et al. 2018]. Par conséquent, et comme évoqué précédemment, les essais ont été démarrés avec 3 facteurs, puis 4, 5 (...) au moins jusqu'à 7 et tant que les résultats restaient cohérents avec une erreur statistique acceptable.

L'estimation de l'erreur DISP fournit notamment deux paramètres pour estimer le niveau d'erreur :

- La baisse de la valeur Q, qui caractérise la performance du modèle à représenter les mesures. La baisse doit rester < 1 % pour rester acceptable. Au-delà de 5 %, le modèle est inexploitable.
- Le nombre de « swaps », qui indique l'instabilité de la modélisation et le niveau de mélange entre facteurs. Idéalement, le nombre de swaps doit être de 0, ou à défaut rester le plus faible possible.

Enfin, la modélisation peut ne pas converger vers une solution, ce qui entraîne une erreur et donc l'absence de résultats. L'ensemble des résultats est indiqué en Annexe A, et représenté graphiquement dans la Figure 2.

Ainsi, aucune convergence n'a été atteinte pour les modèles PMF sans contrainte avec seulement 3 et 4 facteurs, qui ne fournissent donc pas de solution. Par ailleurs, les modèles PMF à 7 et 8 facteurs entraînent des baisses de la valeur Q trop élevées (> 5 %) pour que l'on puisse envisager leurs solutions.

Les modélisations à 3, 4 et 5 facteurs sur lesquelles nous avons appliqué une contrainte au lévoglucosan (maximale sur le facteur identifié « biomasse », minimale sur les autres) entraînent également des baisses de la valeur Q importantes (> 1 %) et ont donc aussi été écartées, même si elles présentaient peu de mélanges entre facteurs (« swaps »).

Sur les données utilisées (41 sites : campagne 2021, 30 paramètres : métaux, HAP, PCB, PCDD-F, lévoglucosan, voir section 2.3), seuls 3 modèles PMF permettent une convergence des solutions avec des erreurs acceptables : 5 facteurs sans contrainte, 6 facteurs sans contrainte et 6 facteurs avec les contraintes sur le lévoglucosan (Figure 2, Annexe A). La solution à 5 facteurs sans contrainte a été exclue, car elle présente un niveau de mélange des facteurs élevé (swaps = 40).

La modélisation PMF avec 6 facteurs retenue est ainsi proche des résultats obtenus sur les données 2014 par ACP, au niveau du nombre de facteurs [Ratier et al. 2018]. La solution sans contrainte présente moins de mélanges entre facteurs (swaps = 6) que la solution avec les contraintes sur le lévoglucosan (swaps = 15). Dans les deux cas, ils sont répartis équitablement entre le facteur identifié « biomasse », un facteur « organique », et un facteur reflétant a priori les émissions industrielles.

La solution à 6 facteurs avec les contraintes sur le lévoglucosan (maximal dans le facteur « biomasse » et minimal dans les autres facteurs), malgré un mélange entre facteurs plus élevé que dans la solution sans contrainte, présente une meilleure représentation des sources auxquelles on peut s'attendre. Notamment, la maximisation du lévoglucosan dans le facteur désigné comme « combustion de biomasse » entraîne que la variation du lévoglucosan est principalement expliquée par ce facteur (53,6 % alors qu'il n'y était présent qu'à seulement 21,2 % dans la solution sans contrainte), conformément au fait que la combustion de biomasse est *a priori* la seule émettrice de lévoglucosan. Le maintien du lévoglucosan dans d'autres facteurs (18,5 % dans le facteur « PCB », 7,8 % dans le facteur « calcium » et 20,1 % dans le facteur « PCDD-F ») résulte de diverses incertitudes, notamment statistiques, mais aussi des mélanges dans les facteurs, possiblement liés à des émissions de type « combustion de biomasse » en milieu industriel par exemple (incendies, incinération de déchets contenant de la biomasse...).

Par conséquent, la solution à 6 facteurs avec contraintes sur le lévoglucosan a été retenue.

4. JUSTESSE ET ERREURS RÉSIDUELLES DU MODÈLE PMF

Par définition, toute modélisation comporte des approximations et des erreurs, même si elles sont minimisées par le choix de la solution offrant les meilleurs résultats. Les erreurs qui subsistent peuvent être étudiées et visualisées au travers des différences entre les valeurs de contaminants calculées par le modèle PMF et celles qui ont été mesurées. Ces différences sont nommées « résidus ».

4.1. Exemples caractéristiques

La Figure 3 permet de visualiser la performance du modèle PMF choisi (6 facteurs, contraintes sur le lévoglucosan) à représenter la réalité des valeurs mesurées pour Ni, Al, PCDF et lévoglucosan.

Dans un cas idéal, les valeurs calculées par le modèle sont identiques à celles mesurées, ce qui est représenté graphiquement par la droite bleue. **Dans le cas du nickel (Ni, Figure 3a),** on peut

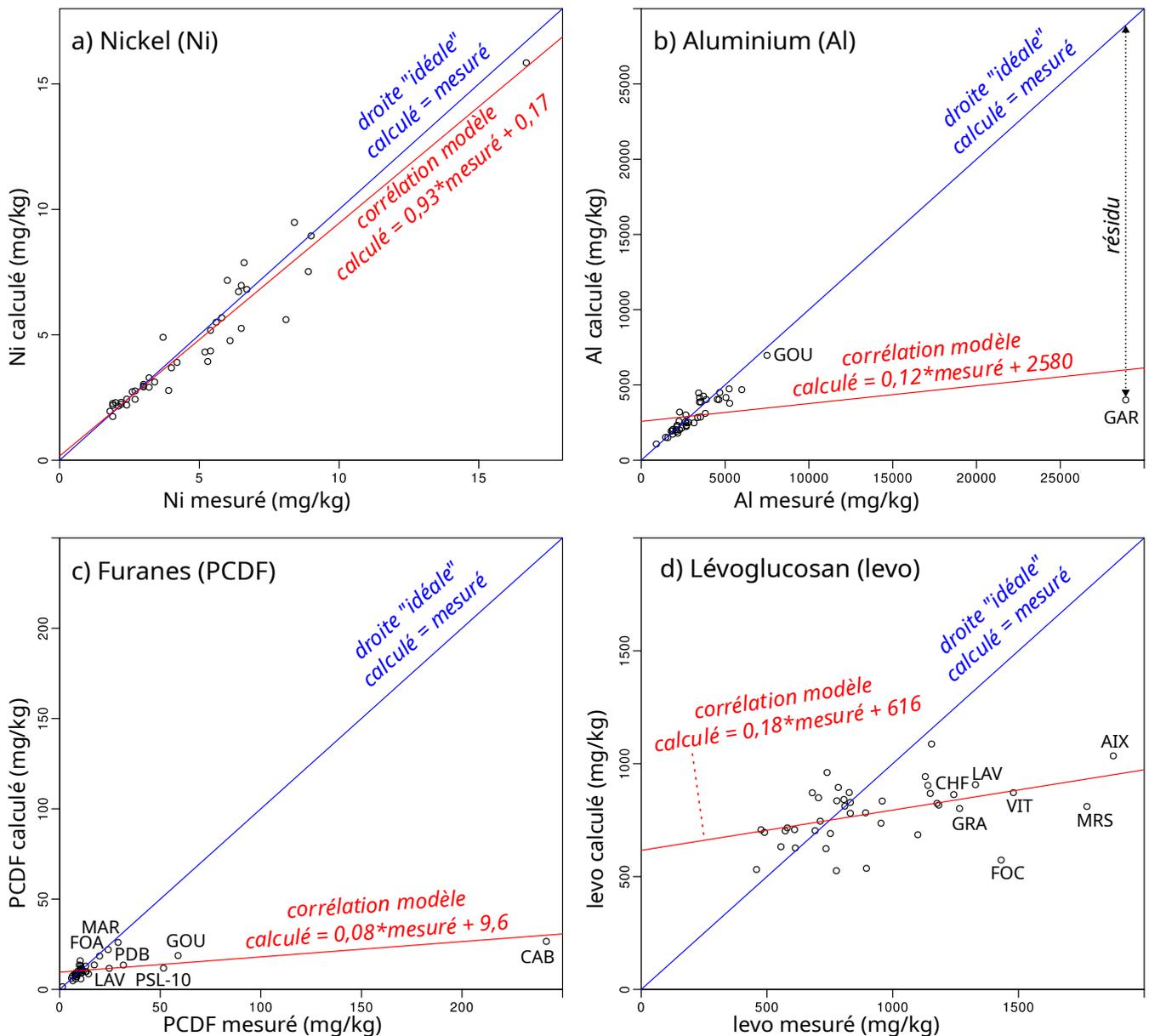


Figure 3. Représentation des valeurs calculées par le modèle PMF (6 facteurs, contraintes sur le lévoglucosan) pour a) le nickel, b) l'aluminium, c) les furanes, et d) le lévoglucosan.

En bleu, la droite idéale où toutes les valeurs calculées sont égales aux valeurs mesurées (« calculé = mesuré »).

En rouge, les droites de corrélation réelles (« calculé = a*mesuré + b »).

En noir dans l'encadré b), visualisation graphique du résidu entre ce qui est mesuré et calculé.

constater que le modèle représente presque parfaitement la réalité. Ainsi, la corrélation réelle montre une pente proche de 1 (qui est la pente idéale), et croise l'axe des ordonnées (intersection) proche de 0 (mesure à 0 bien représentée par une valeur calculée à 0). **Dans le 2^{ème} exemple (Al, Figure 3b)**, tous les points sont bien alignés avec la droite idéale, à l'exception de GAR. Comme celui-ci présente un niveau de Al largement plus élevé, il entraîne une très forte dérive de la droite de corrélation (en rouge), et les résultats qui en découlent sont très faibles (pente faible, intersection très élevée, et mauvais coefficient de corrélation). Le contaminant Al est donc très bien modélisé sur l'ensemble des sites à l'exception de GAR. Cela peut indiquer une source mal prise en compte par le modèle, car très spécifique à cet unique site. Dans ce cas, il est très probable que cela soit lié à l'incidence marquée de l'usine d'alumine de Gardanne qui n'impacte que ce site. **Le cas des PCDF (Figure 3c)** est similaire à celui de Al, mais ici plusieurs sites sont concernés (CAB, GOU, PSL-10, PDB, LAV). Cela produit la même conséquence sur la corrélation qui est très éloignée de la pente idéale. Ces sites sont tous industriels, le modèle montre donc des difficultés à représenter la sur-exposition de ces sites aux PCDF. Cela est potentiellement lié aux écarts importants observés avec d'autres sites industriels où les concentrations en PCDF sont faibles (TON, LAF, FOA, ENG par exemple). Ces contrastes en PCDF selon les sites de typologies très industrielles restent difficiles à expliquer, car ils semblent a priori soumis aux mêmes sources. Cela se ressent ainsi sur le modèle, qui permet cependant de très bien représenter les PCDF dans la grande majorité des sites. **Dans l'exemple du lévoglucosan (Figure 3d)**, l'erreur du modèle à représenter les mesures s'avère plus généralisée, et induit à nouveau des résidus plus élevés pour les hautes valeurs, ici retrouvées plutôt dans des sites urbains comme AIX, MRS, VIT, FOC, GRA et CHF. Il est probable que ces écarts entre modèle et mesures soient, au moins en partie, liés aux contraintes appliquées au modèle sur le lévoglucosan qui « force » le résultat selon la demande de l'utilisateur plutôt que la justesse statistique.

4.2. Justesse du modèle

Le Tableau 1 résume l'ensemble des statistiques pour l'évaluation de la justesse du modèle. Tout d'abord, les résultats des tests de Kolmogorov-Smirnov (K-S) indiquent tous qu'il n'y a pas de différence significative entre les distributions des valeurs calculées et des valeurs mesurées, pour tous les contaminants étudiés. Cela confirme que le modèle choisi est globalement performant et ses résultats acceptables d'un point de vue statistique.

On retrouve par ailleurs, les excellents résultats obtenus pour le Ni illustrés par la Figure 3a, avec un coefficient de corrélation $R^2 = 0,93$. Dans la Figure 3, des chiffres médiocres en termes de résultats statistiques sont souvent imputables à seulement une très faible minorité d'observations, et masquent de très bons résultats. Finalement, pour une très large majorité de paramètres, le modèle représente bien la réalité malgré une mauvaise pente ou un mauvais coefficient de corrélation. Parmi les métaux, c'est le cas pour Al (seul GAR est mal représenté, Figure 3b), mais aussi pour V, Cr, Mn, Fe, Cu, As (Annexe B). On peut constater qu'une poignée d'observations concentre la plupart des écarts entre valeurs calculées et valeurs mesurées, GAR (Al, V, Cr, Fe), PSL-9 (Cr, Mn) et PSL-12 (Cu), TON (Fe). On peut ainsi soupçonner, notamment à GAR et PSL, la présence de sources très spécifiques et intenses à proximité de ces sites conduisant à de mauvaises estimations du modèle.

En ce qui concerne les HAP, c'est principalement le site GOU qui concentre les erreurs de modélisation, ainsi que le site PSL pour les HAP3 (Annexe C). Cela affecte principalement les pentes des HAP2, HAP3 et HAP6, mais les corrélations restent très bonnes malgré tout ($R^2 > 0,79$). Pour les PCDD (Annexe C), seul SMN (Saint-Martin Nord) se révèle mal évalué par le modèle.

Tableau 1. Résultats de régression entre les valeurs mesurées et les valeurs modélisées (K-S stat et K-S p-value sont le résultat et la valeur-p du test de Kolmogorov-Smirnov). Les valeurs en rouges indiquent des résultats médiocres (interception > min. ; ou pente < 0,5 ; ou R² < 0,5, ou K-S p-value < 0,05).

Paramètres	Interception	Min. mesuré	Pente	Corrélation (R ²)	K-S stat	K-S p-value
Al	2580	900	0,12	0,17	0,14	0,41
V	6,4	3,1	0,22	0,23	0,2	0,08
Cr	7,8	3,7	0,42	0,6	0,16	0,22
Mn	41	32	0,47	0,52	0,1	0,76
Fe	1770	1050	0,47	0,62	0,16	0,23
Co	0,08	0,40	0,88	0,87	0,1	0,8
Ni	0,17	1,8	0,93	0,93	0,09	0,92
Cu	12	6,4	0,06	0,08	0,12	0,61
Zn	6,3	21	0,81	0,92	0,09	0,92
As	0,62	0,51	0,4	0,43	0,13	0,49
Mo	0,16	0,35	0,74	0,73	0,16	0,27
Cd	0,01	0,05	0,81	0,7	0,08	0,96
Sb	0,10	0,05	0,71	0,93	0,12	0,61
Pb	1,6	2,30	0,72	0,84	0,07	0,98
Ca	455	270	0,74	0,87	0,17	0,17
Ti	28	37	0,62	0,75	0,1	0,77
HAP2	17	16	0,37	0,79	0,06	1
HAP3	66	63	0,39	0,82	0,12	0,6
HAP4	7,6	231	0,92	0,82	0,08	0,95
HAP5	22	90	0,81	0,9	0,09	0,88
HAP6	56	34	0,18	0,8	0,13	0,52
PCDD	18	5,9	0,47	0,47	0,14	0,39
PCDF	9,6	1,4	0,08	0,38	0,12	0,57
levo	616	458	0,18	0,21	0,09	0,86
PCB3	3,5	2,4	0	0	0,1	0,85
PCB4	0,88	1,1	0,58	0,69	0,08	0,93
PCB5	3,8	5,7	0,62	0,8	0,07	0,98
PCB6	4,5	2,1	0,4	0,33	0,15	0,28
PCB7	3,9	2,6	0,24	0,33	0,09	0,86
PCB8	1,1	1,0	0,46	0,61	0,1	0,84
S _j (somme normée)	1,65	12	0,91	0,94	0,2	0,08

Enfin, pour les PCB (Annexe D), le site MIR (Miramas) est mal représenté pour les PCB₆, PCB₇, et PCB₈. D'autre part, le modèle a totalement échoué à représenter les PCB₃. Ce paramètre ne pourra donc pas être interprété dans le cadre de cette modélisation.

Dans les autres cas (Co, Ni, Zn, Mo, Cd, Sb, Pb, Ca, Ti, HAP₄, HAP₅, PCB₄ et PCB₅), les correspondances entre le modèle et les valeurs mesurées sont excellentes. On constate aussi que la somme normée des contaminants est très bien représentée par le modèle PMF choisi.

4.3. Analyse des résidus

Comme indiqué en section 4.1., les résidus sont les écarts entre les valeurs calculées par le modèle et les valeurs réellement mesurées. Ils peuvent être calculés pour chaque contaminant (paramètre) et chaque site d'étude (observation). Ces écarts, appelés « résidus », sont donc calculés ainsi :

$$X_{résidu\ ij} = X_{mesure\ ij} - X_{modèle\ ij} \quad (\text{équation 3})$$

Afin que les résidus soient comparables sur une même échelle, quel que soit le contaminant considéré, on pourra également utiliser les valeurs dites « réduites » des résidus (c'est-à-dire pondérées par les écarts-types), comme cela a été fait pour le paramètre « somme des contaminants » S_j (équations 1 et 2, section 2.6). Alors, l'équation 3 devient :

$$X_{résidu\ ij} = \frac{X_{résidu\ ij}}{\text{écart-type de } j} \quad (\text{équation 4})$$

où i est le site considéré (observation) et j le contaminant considéré (paramètre).

Aussi, et comme indiqué en section 4.1., les résidus peuvent être facilement visualisés pour un paramètre (contaminant) en particulier. On constate par exemple des résidus très importants pour GAR sur la modélisation de Al et pour CAB sur la modélisation de PCDF (Figure 3).

Lorsque les résidus de chaque site sont cumulés, il peut ainsi être évalué si un site ou plusieurs d'entre eux rassemblent une grande quantité d'erreurs (et donc de résidus). La Figure 4 représente ainsi les valeurs de résidus (réduites selon l'équation 4, et sommées en valeurs absolues) cumulées pour chacun des sites étudiés.

Il ressort que le site GAR (Gardanne) concentre un niveau de résidus, et donc d'erreur de modélisation, très élevé comparé aux autres sites, tel qu'on a pu le voir en exemple pour Al (Figure 3), mais aussi V, Cr, Fe (Annexe B). Cette mauvaise modélisation de plusieurs contaminants au site GAR est très probablement liée à l'existence d'une (ou plusieurs) source(s) d'émissions très spécifiques à ce site relativement isolé des autres, en l'occurrence l'usine Altéo et la centrale thermique de Provence, qui ont des signatures différentes des émissions industrielles de la ZIP de Fos ou de Lavéra. Cela s'est d'ailleurs bien confirmé dans un travail dédié à la zone de Gardanne/Meyreuil [Reuillard et al. 2024].

Par ailleurs, dans les 25 % des sites qui cumulent le plus de résidus (barres au-delà du 75^e percentile dans la Figure 4), on retrouve 3 sites de Port-Saint-Louis-du-Rhône (PSL-9, PSL-12 et PSL-13), dont le modèle sous-estime notamment les niveaux de PCB et métaux ainsi que certains HAP. La spécificité de l'exposition de Port-Saint-Louis-du-Rhône aux HAP avait déjà été relevée par l'Institut Ecocitoyen lors des études HAPGIS et CARE [Austruy & Reuillard 2021; Reuillard & Austruy 2022]. Elle semble donc aussi montrer des teneurs particulières pour certains métaux et PCB. Il est toutefois possible que le grand nombre de mesures réalisées sur cette ville ait permis de compenser la spécificité locale, limitant les erreurs de modélisation et les résidus, comparé à Gardanne.

Enfin, des erreurs marquées sont aussi retrouvées sur les sites les plus exposés de la ZIP de Fos, GOU (Goulevielle), ENG (Engrenier) et CAB (Caban). Ces sites enregistrent des valeurs extrêmes de bioaccumulation de plusieurs contaminants, liées à leur situation géographique et donc à leur proximité des émissions industrielles de la ZIP de Fos, qui se révèlent mal évaluées par le modèle. C'est une difficulté classique de la modélisation en général, qui n'est pas destinée à modéliser des situations particulières (GAR, PSL) ou extrêmes (GOU, ENG, CAB).

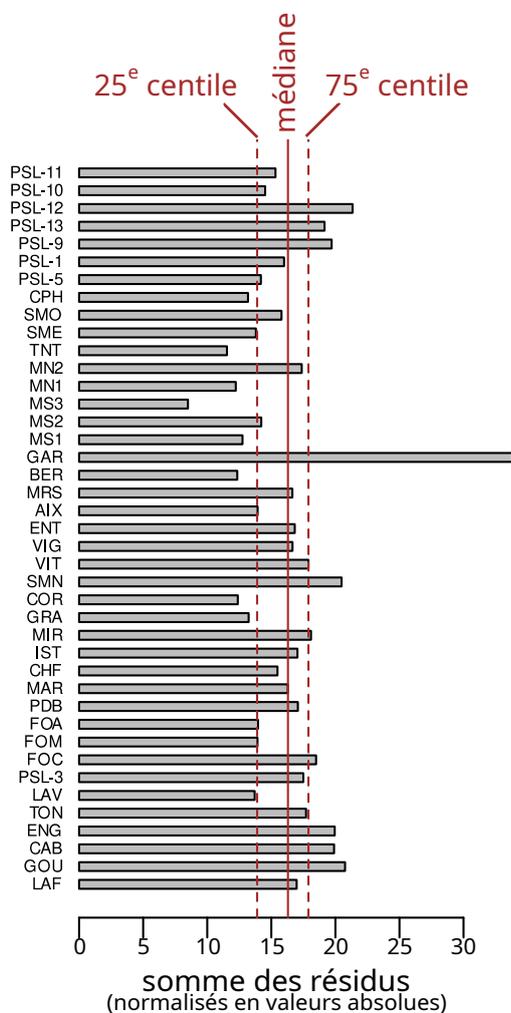


Figure 4. Cumul des résidus (en valeur absolue et normalisée) pour chaque site étudié.

Pour conclure, il semble malgré tout que le site GAR (Gardanne) se détache largement dans les résultats de la modélisation sur les données de la campagne 2021. Il est possible que le retrait de ce site puisse améliorer la performance globale de la modélisation PMF de manière générale, car la présence de ce site particulier peut perturber la modélisation PMF dans son ensemble. C'est une option à envisager dans de futurs travaux pour affiner la modélisation PMF des sources des contaminants dans les lichens de la Métropole-Aix-Marseille-Provence (MAMP).

5. RÉSULTATS DE LA MODÉLISATION PMF

Les 6 facteurs obtenus par la modélisation PMF avec l'application des contraintes sur le lévoglucosan sont présentés dans la Figure 5.

5.1. Signatures des sources (facteurs PMF)

Le facteur « **combustion de biomasse** » est caractérisé par la plus forte contribution de lévoglucosan (53,6 %), logiquement puisque celui-ci a été contraint à être maximisé sur ce facteur. On retrouve également des parts importantes des HAP, ainsi que des PCB. Globalement, à l'échelle de la MAMP, ce facteur est le plus présent, à hauteur de 26,9 %.

Deux facteurs industriels, « **industriel ZIP** » et « **industriel 2** » viennent ensuite. Avec des présences à l'échelle de la MAMP de 20,6 % et 16,6 %, respectivement, la somme de ces facteurs identifiés comme industriels représente donc au total 37,2 % de contribution aux contaminants analysés. Ils sont constitués de fractions importantes de la plupart des métaux. Le facteur « **industriel ZIP** » est identifié grâce au molybdène (où plus de 60 % du molybdène sont expliqués), un métal très spécifique au site Lyondell de la ZIP de Fos [Dron & Chamaret 2015]. Il comporte aussi de nombreux métaux typiques des émissions de la ZIP (Pb, Cr, Mn, Zn...) ainsi qu'une fraction importante de HAP. Le facteur « **industriel 2** » comporte également de grandes parts des métaux (Al, V, Cr, Mn, Fe, As, Ti...) ainsi qu'une fraction notable des PCDD-F avec une prédominance des PCDF, caractéristiques des émissions industrielles [Ratier et al. 2018].

Le facteur « **PCB** » est caractérisé par une signature réservée aux PCB, dont la part atteint ici 40 % à 60 % de leur répartition entre les facteurs. Ce facteur représente une contribution globale de 15,9 %, expliquant donc notamment la variabilité des PCB. Il n'est cependant pas possible de le relier à un type de source connue d'après sa seule signature, car les émissions atmosphériques de PCB demeurent mal évaluées.

Le facteur « **calcium** » est caractérisé par le fait qu'il explique 80 % des teneurs en calcium (Ca). Il présente aussi une forte part de l'antimoine (Sb). Compte tenu de la nature calcaire de la plupart des sols de la zone étudiée, cela suggère que ce facteur est peut-être représentatif de la remise en suspension de poussières de sols. La contribution globale de ce facteur est de 10%.

Le facteur « **PCDD-F** », caractérisé par les plus fortes fractions en PCDD-F est aussi difficile à relier à une catégorie spécifique d'émissions atmosphériques. Contrairement à ce qui est généralement observé en milieu industriel, ce sont les PCDD qui dominent dans ce facteur dont la contribution à l'échelle de la MAMP est de 10 %.

Quoi qu'il en soit, ces signatures, et leurs contributions globales aux concentrations en contaminants recherchés ici dans les sites étudiés de la MAMP, montrent que les émissions industrielles et celles liées à la combustion de biomasse (chauffage au bois, feux agricoles, incendies, écobuage...) sont prédominantes. Cela confirme les premières estimations réalisées dans les lichens pour la campagne 2014 [Ratier et al. 2018] et celles issues de modélisations PMF appliquées aux particules PM2.5 prélevées en 2013 à Fos-sur-Mer [Sylvestre 2016].

Aucune corrélation entre les facteurs n'a pu être mise en évidence. Toutefois, l'étude de la répartition géographique des facteurs pourra fournir des précisions éventuelles sur leurs origines.

5.2. Contributions des sources : répartition spatiale des facteurs PMF

Les répartitions géographiques des contributions brutes au paramètre « total » des 6 facteurs sont présentées individuellement en Annexe E.

Il devient aussi particulièrement intéressant d'observer les contributions des 6 facteurs de façon relative sur chacun des sites étudiés et les répartitions géographiques de ces contributions relatives, telles qu'elles sont représentées sur la carte de la Figure 6. Le détail des contributions relatives des 6 facteurs pour chacun des sites étudiés est aussi indiqué dans le Tableau 2.

Les facteurs identifiés comme industriels sont majoritaires sur l'ensemble de la ZIP de Fos, dans pratiquement tous les sites, de Port-Saint-Louis-du-Rhône à Martigues (dont Lavéra), Port-de-Bouc, Engrenier et Fos-Carabins, ainsi qu'à Châteauneuf-les-Martigues, Vitrolles et Gardanne (Figure 6). Cette répartition géographique correspond bien aux activités industrielles présentes. Le facteur « Industriel ZIP » est ainsi logiquement majoritaire dans la ZIP, jusqu'à 74,6 % au site GOU en plein cœur de la ZIP, et sur plusieurs sites de Port-Saint-Louis-du-Rhône (16,4 % à PSL-5 jusqu'à 56,6 % à PSL-11). Dans les sites périphériques, c'est le facteur « industriel 2 » qui prend le dessus (Fos-sur-Mer, Port-de-Bouc, Martigues) jusqu'à devenir largement majoritaire dans les sites à caractère industriel éloignés de la ZIP, comme Châteauneuf-les-Martigues, Vitrolles et Gardanne (26,3 % ; 35 % ; et 41,2 % respectivement). Cela coïncide bien avec leur exposition à des émissions industrielles différentes de celles rencontrées dans la ZIP de Fos.

Le facteur « industriel 2 » est aussi présent au niveau de la ZIP de Fos et alentour, indiquant que les émissions de la ZIP de Fos sont aussi en partie représentées par ce facteur (Figure 6).

Le facteur « combustion de biomasse » est quant à lui majoritaire dans les zones rurales comme Entressen (65,9 %), Saint-Martin-de-Crau (40,4 % à 55,0 %), Vigueirat (42,3 %), Grans (41,9 %) ou Cornillon-Confoux (43,7 %). Il est également majoritaire ou très important dans la plupart des sites urbains à Marseille, Istres ou Miramas, mais aussi à Fos-sur-Mer, Port-Saint-Louis-du-Rhône, Châteauneuf-les-Martigues, Vitrolles, Gardanne et Aix-en-Provence (Figure 6). Sa répartition est globalement homogène sur tout le territoire, même en zone industrielle, mais la forte incidence des facteurs industriels autour de la ZIP de Fos le rend malgré tout minoritaire (Annexe E). Cette répartition géographique reflète bien que la combustion de bois est une source souvent prépondérante dans les milieux ruraux et urbains. Elle est supérieure à 30 % dans 26 des 41 sites étudiés (Tableau 2).

Tout comme les concentrations en PCB, **le facteur « PCB »** a une répartition homogène sur la zone étudiée. Sa présence à des niveaux élevés dans les sites de Berre-l'Étang (39,1 %), Lavéra (30,1 %) et plusieurs sites de Port-Saint-Louis-du-Rhône (jusqu'à 29,9 %) ne permet pas d'exclure une contribution industrielle à ce facteur, notamment pétrochimique, ni d'écarter d'autres sources (transports notamment) ou même combustion de bois dans lequel les PCB sont bien présents (Figure 5) puisqu'il est élevé aussi dans la plupart des sites urbains, comme Istres (47,7 %) ou Miramas (28,8 %) (Figure 6).

Le facteur « calcium » semble au contraire très marqué sur certaines zones, Marseille (20,6 % à 40,2 %), Fos-sur-Mer (21,5 % à 48,3 %, sauf FOC à 5,3 %), et Saint-Martin-de-Crau (14,5 % à 25,0 %). Aussi, les deux sites les plus marqués par ce facteur « calcium » sont très exposés aux envols de poussières dans des lieux où les sols sont très calcaires (ENG, FOA et MN2), de par la présence notamment de sites importants d'exploitations de carrières. Sont relevés aussi les sites en zone agricole à Saint-Martin-de-Crau et Entressen notamment, pouvant être impactés par l'envol de poussières de sols laissés à nu entre les cultures. Sa présence dans la ZIP de Fos (CAB, GOU et TON) peut être liée au terminal minéralier dont les stockages sont en extérieur et au stockage de minerai et scories à proximité des sites sidérurgiques [Sylvestre et al. 2017]. Il en est de même pour Marseille sud (MS1, MS2 et MS3), où de nombreux dépôts de scories sont encore présents. Il semble donc que ce facteur soit lié en grande partie à l'envol de poussières de sols

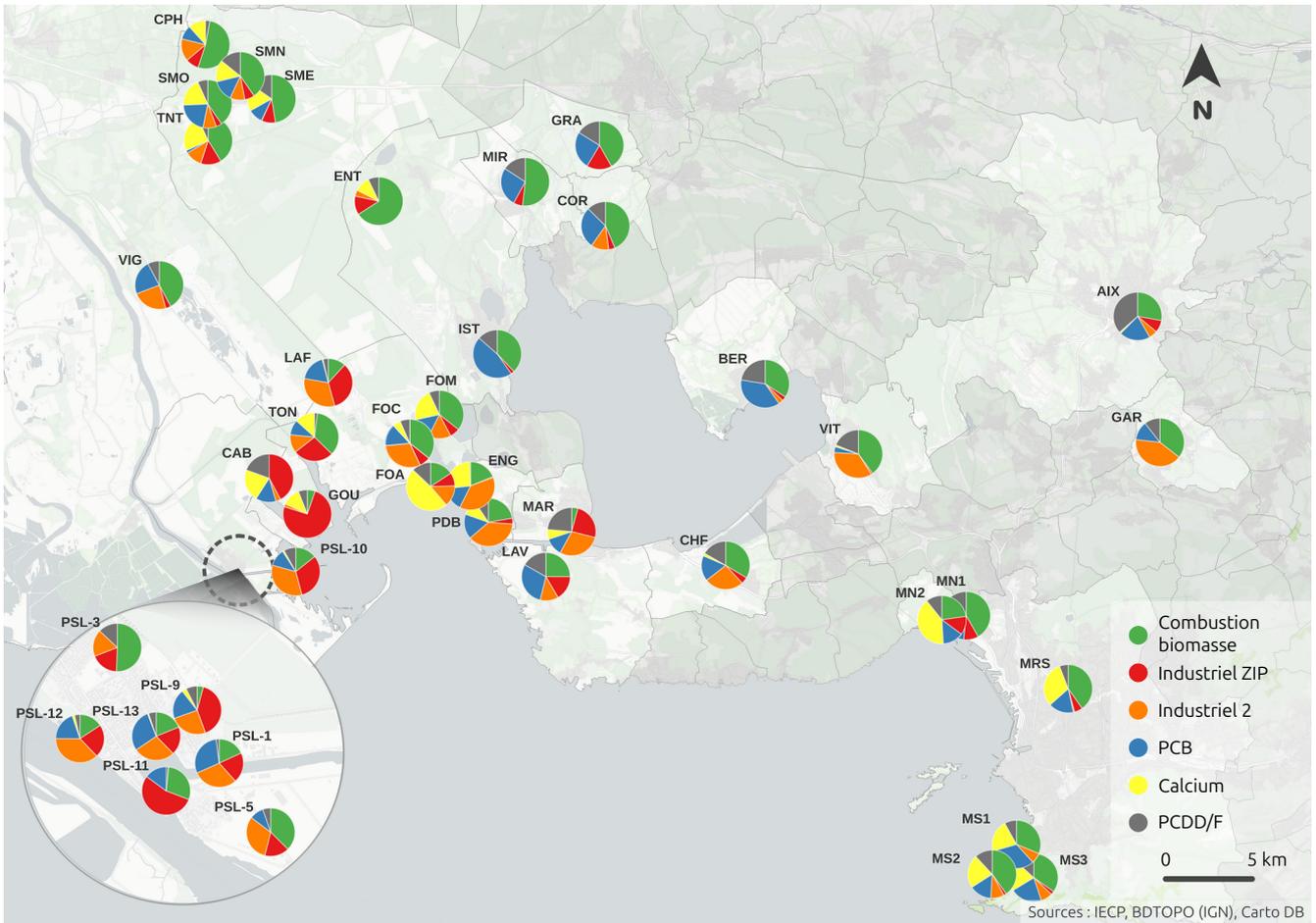


Figure 6. Contributions relatives des 6 facteurs dans chacun des sites étudiés.
Cartographie Coraline Soul (QGIS)

calcaires, mais il est probablement mélangé avec des parts industrielles ou urbaines. Le site FOA est ainsi marqué par de forts niveaux de métaux issus des émissions industrielles, et le site MN2 se situe à proximité immédiate d'une friche industrielle. La sur-représentation de l'antimoine (Sb) dans ce facteur (Figure 5) peut aussi témoigner de l'impact d'émissions industrielles spécifiques [Sylvestre 2016].

Enfin le facteur « PCDD-F » a des contributions relatives distribuées de façon très homogène sur le territoire d'étude. Il est compris entre 10 % et 25 % dans 19 des 41 sites (Tableau 2), et trouve sa contribution maximale à Aix-en-Provence (36,6 %). Si sa présence est relativement inférieure dans la ZIP de Fos, à Port-de-Bouc, Fos-sur-Mer et Port-Saint-Louis-du-Rhône (Figure 6), c'est surtout par la domination des facteurs industriels dans ce secteur. En effet, la contribution brute dans la ZIP et alentour reste élevée suggérant qu'il soit aussi en partie industriel (Annexe E). Ce facteur est marqué par une légère prépondérance des dioxines (PCDD) par rapport aux furanes (PCDF), ce qui n'est généralement pas caractéristique des émissions industrielles, où les PCDF sont majoritaires comme dans le facteur « industriel 2 » (Figure 5). Aussi, la présence de cuivre (Cu) et d'antimoine (Sb) peuvent suggérer une participation du transport routier (métaux émis par l'abrasion des freins) ou des émissions agricoles et traitements phytosanitaires. En effet, considérant que les PCDD-F ont fortement baissé dans les sites industriels, il est aussi possible que les émissions industrielles y contribuent moins et que d'autres sources de PCDD-F (notamment PCDD) prennent le dessus (pesticides notamment), conduisant à un mélange des origines de pollutions dans ce facteur.

Tableau 2. Contributions relatives des 6 facteurs issus de la modélisation PMF pour chacun des 41 sites étudiés. Les contributions supérieures à 30 % sont indiquées en rouge.

La modélisation PMF impose que la somme des contributions soit égale à 100 % pour chaque site, conduisant parfois à des valeurs légèrement négatives qui peuvent être considérées comme égales à 0.

Site	Combust. bois	Industriel ZIP	Industriel 2	PCB	Calcium	PCDD/F
LAF	11,8	34,0	32,0	18,1	0,5	3,5
GOU	5,3	74,6	3,0	-1,1	12,2	5,9
CAB	-2,9	45,2	3,3	13,5	21,5	19,5
ENG	18,8	0,5	38,0	15,9	26,3	0,4
TON	37,4	27,0	12,2	10,2	15,1	-1,9
LAV	25,0	16,6	12,2	30,1	-0,6	16,7
PSL-3	50,7	18,4	24,8	-4,2	-2,6	13,0
FOC	35,7	7,4	30,4	14,7	5,3	6,4
FOM	35,7	6,6	15,0	14,2	21,5	7,0
FOA	15,7	8,8	17,1	-2,8	48,3	12,9
PDB	22,3	3,9	37,5	16,8	10,3	9,3
MAR	4,2	24,8	28,9	12,0	6,6	23,6
CHF	33,7	4,5	26,2	17,1	1,9	16,6
IST	38,0	4,5	-2,2	47,7	-1,5	13,5
MIR	51,8	10,0	-4,0	28,8	-2,4	15,7
GRA	41,9	17,5	-0,8	26,2	-0,9	16,1
COR	43,7	4,2	11,9	28,6	-0,9	12,6
SMN	40,5	6,6	9,9	14,3	14,7	14,2
VIT	40,0	1,0	35,0	3,9	1,2	18,9
VIG	42,3	3,5	23,4	24,0	-0,9	7,8
ENT	65,9	12,6	7,1	-3,0	10,4	7,1
AIX	27,9	8,2	5,6	20,7	1,0	36,6
MRS	40,3	5,6	0,8	16,7	30,9	5,8
BER	34,3	2,9	3,2	39,1	-1,8	22,4
GAR	39,9	-4,3	41,2	14,8	-2,1	10,5
MS1	36,9	-5,7	7,1	32,3	21,5	7,9
MS2	40,2	1,8	8,9	14,9	22,3	11,8
MS3	34,8	2,3	8,0	21,3	20,6	13,1
MN1	41,8	12,4	-2,5	7,0	23,5	17,8
MN2	22,8	16,0	-3,5	13,9	40,2	10,7
TNT	41,3	13,7	11,3	1,9	24,9	6,9
SME	47,6	9,2	0,2	9,9	17,9	15,3
SMO	40,4	3,8	8,9	21,2	19,0	6,6
CPH	55,0	8,7	15,1	9,6	14,5	-2,9
PSL-5	37,6	16,4	31,4	10,5	-1,1	5,3
PSL-1	18,2	20,2	29,9	29,9	-0,3	2,0
PSL-9	4,2	40,3	25,1	20,0	3,0	7,5
PSL-13	18,9	18,9	27,7	28,8	0,7	5,0
PSL-12	15,8	22,1	37,2	19,8	1,7	3,5
PSL-10	14,3	31,7	33,4	11,6	1,1	8,0
PSL-11	30,9	56,6	-2,3	16,4	-0,2	-1,4
GLOBA L	26,9	20,6	16,6	15,9	10,0	10,0

6. COMPARAISON ENTRE MODÈLES PMF ET ACP

6.1. Analyse par composantes principales (ACP)

Comme indiqué en section 2.4, la méthode d'analyse statistique ACP détermine une infinité de composantes (facteurs) afin que l'erreur du modèle se rapproche de 0. Cependant, il est rare de considérer les résultats au-delà de la 10^{ème} composante, où les contributions sont très faibles (généralement < 3 %) et sujettes à des interprétations hasardeuses. Ici, nous n'avons considéré que les composantes avec une contribution > 5 %, soit dans le cas présent, les 5 premières. Les contributions de ces composantes représentent 6,3 % (5^{ème} composante) à 38,6 % (1^{ère} composante), et au total 73,2 % de la variance totale (composantes 1+2+3+4+5).

La première composante, qui fournit la plus forte contribution (38,6 % en 2021), a une signature marquée par l'ensemble des métaux, à l'exception du calcium (Ca) et une participation moyenne du cuivre (Cu). Cette signature est aussi marquée par la forte présence des HAP (toutes les classes) et d'une proportion de furanes (PCDF) nettement supérieure aux dioxines (PCDD). Cette prépondérance des furanes, ainsi que la forte présence des HAP et des métaux, indiquent qu'il s'agit d'une composante reflétant les émissions industrielles de manière générale, sans distinction entre activités industrielles métallurgiques, pétrochimiques, incinération (...) ou associées, telles que le transport maritime notamment (Figure 7). Les contributions de cette composante sont particulièrement fortes autour de la ZIP de Fos et du complexe pétrochimique de Lavéra (communes de Port-Saint-Louis-du-Rhône, Fos-sur-Mer, Port-de-Bouc, Martigues). Elle est aussi très importante à Gardanne. Sa répartition géographique confirme donc sa nature fortement influencée par les émissions industrielles de toutes sortes (Figure 7). Elle est très similaire aux premières composantes des campagnes 2014 et 2017, par sa signature et sa répartition géographique. Sa part relative est en légère augmentation, passant de 34,1 % en 2014 [Ratier et al. 2018] à 37,5 % en 2017 et donc 38,6 % en 2021.

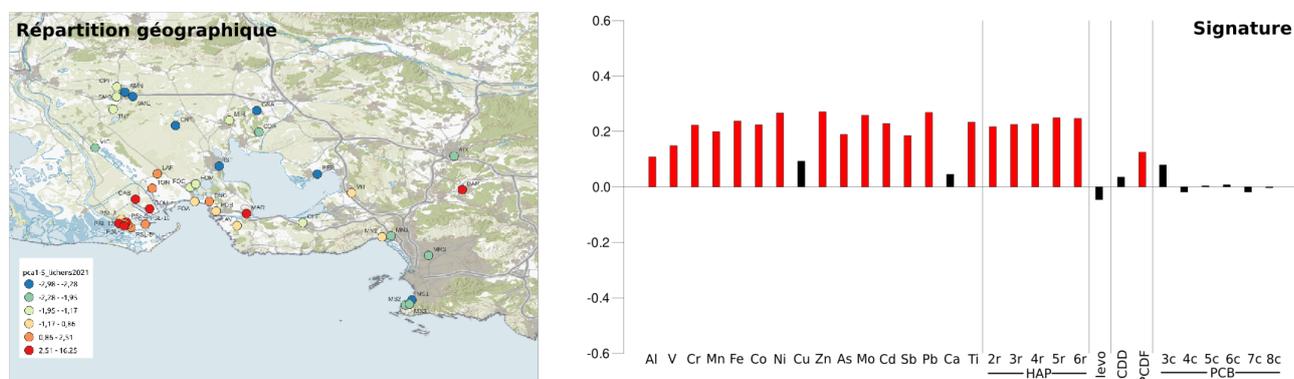


Figure 7. Composante 1 de l'ACP sur les données 2021, répartition géographique (unités arbitraires) et signature (unités arbitraires). La couleur rouge souligne les facteurs-clé de la signature.

La seconde composante est caractérisée par une opposition entre les PCB et quelques métaux, calcium (Ca) et antimoine (Sb) notamment. Il convient de rappeler que les résultats d'une analyse ACP peuvent fournir des résultats positifs comme négatifs, à la fois dans les signatures et les contributions, mais que le caractère positif ou négatif n'a aucune signification physique. Dans le cas présent, la signature de la composante est donc influencée à la fois par les PCB d'un côté, et par le Ca et le Sb de l'autre, de façon opposée mais avec des intensités comparables (Figure 8).

Cette composante 2 représente 12,6 % de la variance. Le calcium peut caractériser la remise en suspension de poussières de sol et Sb de poussières liées au trafic routier issues des systèmes de frein notamment. Comme ils sont de signe négatif, ils sont principalement représentés dans les

contributions négatives de la répartition géographique, à Marseille, Fos-ville et ZIP, et Saint-Martin-de-Crau. Au contraire, les PCB sont surtout présents dans les contributions positives, notamment à Martigues, Port-Saint-Louis, Berre-l'étang, Gardanne, Miramas et Istres. Les sources de PCB dans l'atmosphère sont multiples et difficiles à appréhender, mais selon plusieurs études, la plupart provient de lieux de stockage ou d'incinération d'anciens équipements électriques, de peintures (...), ou de friches contaminées [Jahnke et al. 2022; Oh et al. 2024], ce qui reste relativement cohérent avec la répartition géographique. La partie « PCB » ne peut pas être comparée aux résultats 2014 (pas d'analyses de PCB), mais elle était bien présente en seconde composante également (11,4 %) en 2017.

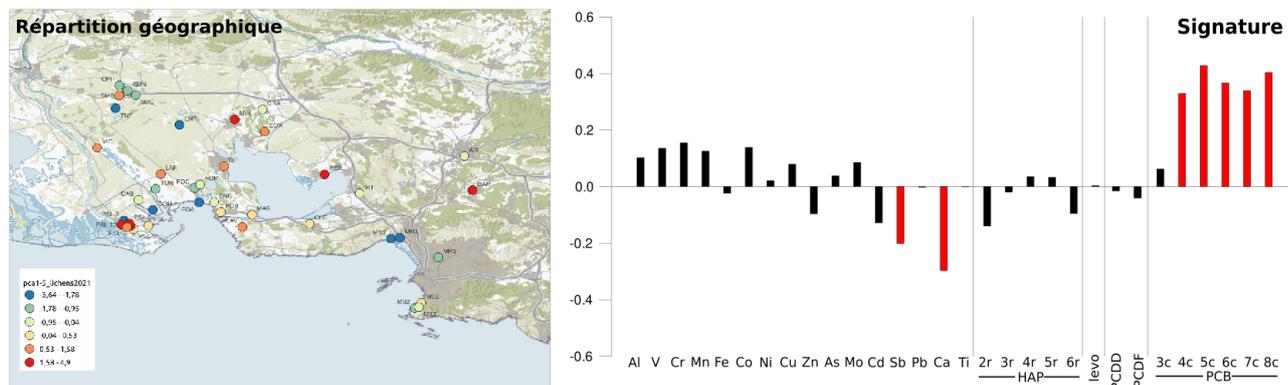


Figure 8. Composante 2 de l'ACP sur les données 2021, répartition géographique (unités arbitraires) et signature (unités arbitraires). La couleur rouge souligne les facteurs-clé de la signature.

Les composantes 3, 4 et 5 contribuent à la variance à hauteur de 8,6 %, 7,1 % et 6,3 %, respectivement. Les éléments Al et V sont présents de façon importante dans leurs signatures, et opposés aux composés organiques (HAP, PCB, PCDF), et à Ca et Sb dans la composante 3, accompagnés de Sb, Ca, des PCDD, PCDF et du lévoglucosan dans la composante 4 et accompagnés des HAP mais opposés aux PCDD, PCDF, lévoglucosan et Cu dans la composante 5. On observe donc des tendances sur les mêmes paramètres, mais de manières contradictoires sur ces 3 composantes. Cela rend difficile l'association de ces composantes à des origines de contamination.

Les métaux Al et V (et Cr dans la composante 3) peuvent correspondre à des émissions industrielles spécifiques au bassin de Gardanne, et la répartition géographique des contributions de ces 3 composantes montrent une influence de ces métaux systématiquement élevée sur Gardanne. En revanche, aucune tendance claire en milieu urbain ou rural ou même sur les autres sites industriels ne se dégage. La ZIP de Fos semble concernée, mais sur des sites différents selon la composante. En revanche, les sites de Marseille et du nord du territoire d'étude (Grans, Cornillon, Istres) sont peu influencés par ces composantes. On peut donc en conclure une incidence notable des émissions industrielles du bassin de Gardanne, répartie sur les 3 composantes, mais sans identifier les autres contributions.

Pour conclure, l'ACP permet de dégager une première composante très nette d'influence industrielle (ZIP de Fos et Gardanne), avec une signature composée de nombreux métaux (à l'exception de Cu et Ca), des HAP et des PCDF, qui correspond bien à ce qui avait été dégagé précédemment [Ratier et al. 2018]. La deuxième composante reste assez nette mais reflète déjà une opposition entre 2 influences, Ca et Sb d'un côté, et les PCB de l'autre.

Les facteurs Ca et Sb peuvent caractériser les envols de poussières de sol (avec Ca notamment), mais l'attribution de l'origine des PCB est plus complexe. Enfin, dans les composantes 3, 4 et 5, il est très difficile de faire ressortir des influences marquées, que ce soit par les signatures ou les répartitions géographiques. Ces composantes sont probablement issues de mélanges d'influences, d'où ces difficultés d'interprétation.

6.2. Comparaison entre les résultats des modèles PMF et ACP

La première composante de l'ACP, au **caractère industriel**, ainsi que les facteurs du modèle PMF « industriel ZIP » et « industriel 2 » montrent plusieurs similitudes. Cette première composante de l'ACP apparaît comme la combinaison des deux facteurs PMF industriels. En effet, elle comporte les éléments métalliques des deux facteurs PMF, les HAP du facteur « industriel ZIP », et les PCDD/F avec une prépondérance des PCDF comme le facteur « industriel 2 ». D'ailleurs sa contribution globale est de 38,6 %, ce qui est très proche de la somme des contributions des deux facteurs PMF industriels qui est de 37,2 %. Ainsi, le modèle PMF permet de séparer 2 origines industrielles pour les contaminants considérés et de fournir une réponse plus détaillée. La répartition géographique de la première composante ACP est d'ailleurs très proche de la combinaison des répartitions des deux facteurs PMF, incluant la ZIP, Port-Saint-Louis-du-Rhône, Fos-sur-Mer marqués par le facteur « industriel ZIP » mais aussi les sites de Châteauneuf-les-Martigues, Gardanne, Vitrolles, Martigues et Fos-sur-Mer marqués par le facteur « industriel 2 ». Dans la répartition de la composante ACP, on peut toutefois noter la présence supplémentaire des sites FOA (Fos-Arènes) et MN2 (Marseille Nord 2), pour lesquels le modèle PMF avait privilégié le facteur « calcium ». Cela pourrait donc appuyer l'hypothèse d'une nature, au moins en partie, industrielle du facteur PMF « calcium ».

La seconde composante de l'ACP, avec d'un côté un **caractère marqué par les PCB** et quelques métaux, et à l'opposé une **signature marquée par le calcium (Ca) et l'antimoine (Sb)**, évoque clairement les facteurs PMF, « PCB » et « calcium », respectivement. Comme pour la première composante, le modèle PMF permet de les décomposer en deux facteurs distincts, permettant une meilleure lecture et facilitant les interprétations. La contribution globale de la deuxième composante de l'ACP est de 12,6 %. Cela correspond assez logiquement à la moyenne des deux facteurs PMF qui la composent (12,9 %), mais ici de façon opposée. La signature « calcium » se retrouve bien à FOA et MN2, mais aussi Saint-Martin-de-Crau, et les autres sites de Fos-sur-Mer et Marseille, à la fois pour la composante ACP et le facteur « calcium » du modèle PMF. La signature « PCB » est aussi répartie géographiquement de façon comparable pour l'ACP et la PMF (Port-Saint-Louis-du-Rhône, Miramas, Istres, Grans, Berre-l'étang), à l'exception de Gardanne où elle est presque absente dans le modèle PMF, mais élevée dans le modèle ACP.

Les composantes suivantes de l'ACP (3, 4 et 5) sont toutes marquées par la présence des PCDD-F, tout comme le facteur « PCDD-F » du modèle PMF. Les composantes 4 et 5 sont aussi marquées par la présence du lévoglucosan, comme le facteur « combustion biomasse » du modèle PMF. Cependant, on a pu voir que ces composantes restaient difficilement interprétables, du fait de tendances contradictoires entre les différents paramètres impliqués (PCDD-F, lévoglucosan, HAP, et quelques métaux). Au contraire, le modèle PMF a permis de dégager un **facteur « combustion biomasse »** clairement identifié par le lévoglucosan, maximisé de façon contrainte mais pertinente, étant donné sa contribution globale (26,9 %), sa répartition géographique homogène mais plus importante en milieux ruraux et urbains, ainsi que par la présence de HAP dans sa signature (émis par tous types de combustion, dont celle de la biomasse).

RÉFÉRENCES

- Austruy, A. and Reuillard, M. (2021).** *HAPs atmosphériques à Port-Saint-Louis-du-Rhône : suivi intégré et sources*. Rapport final HAPSYS, Institut Ecocitoyen pour la Connaissance des Pollutions.
- Benadda, C. (2021).** *Caractéristiques et sources des particules atmosphériques en proximité industrielle par spectrométrie de masse aérosol(ACSM)*, Rapport Master2, Institut Ecocitoyen pour la Connaissance des Pollutions & Aix-Marseille Université.
- Camman, J.; Chazeau, B.; Marchand, N.; Durand, A.; Gille, G.; Lanzi, L.; Jaffrezo, J.-L.; Wortham, H. and Uzu, G. (2024).** *Oxidative potential apportionment of atmospheric PM₁: a new approach combining high-sensitive online analysers for chemical composition and offline OP measurement technique*, Atmospheric Chemistry and Physics 24 : 3257-3278.
- Chazeau, B. (2021).** *Phénoménologie, sources et dynamique de la pollution atmosphérique par les particules fines : observatoire de Marseille*, Thèse de doctorat, Aix-Marseille Université.
- Dauphin, C.-E.; Durand, A.; Lubonis, K.; Wortham, H. and Dron, J. (2020).** *Quantification of monosaccharide anhydrides by gas chromatography/mass spectrometry in lichen samples*, Journal of Chromatography A 1612 : 460675.
- Dron, J.; Austruy, A.; Agnan, Y.; Ratier, A. and Chamaret, P. (2016).** *Utilisation de la biosurveillance lichénique sur la zone industrialo-portuaire de Fos-sur-Mer : retour sur trois ans de suivi à l'échelle d'un territoire intercommunal*, Pollution Atmosphérique 228.
- Dron, J. and Chamaret, P. (2015).** *Étude des émissions en molybdène du site industriel Lyondell Chimie France à Fos-sur-Mer*, Rapport final MOLY, Institut Ecocitoyen pour la Connaissance des Pollutions.
- Dron, J.; Ratier, A.; Austruy, A.; Revenko, G.; Chaspoul, F. and Wafo, E. (2021).** *Effects of meteorological conditions and topography on the bioaccumulation of PAHs and metal elements by native lichen (Xanthoria parietina)*, Journal of Environmental Sciences 109 : 193-205.
- Dron, J.; Wafo, E.; Boissery, P.; Dhermain, F.; Bouchoucha, M.; Chamaret, P. and Lafitte, D. (2022).** *Trends of banned pesticides and PCBs in different tissues of striped dolphins (Stenella coeruleoalba) stranded in the Northwestern Mediterranean reflect changing contamination patterns*, Marine Pollution Bulletin 174 : 113198.
- Durand, A.; Dron, J.; Prudent, P.; Wortham, H.; Dalquier, C.; Reuillard, M. and Austruy, A. (2024).** *Evaluation of the atmospheric pollution by pesticides using lichens as biomonitors*, Science of The Total Environment 955 : 177286.
- Jahnke, J. C.; Martinez, A. and Hornbuckle, K. C. (2022).** *Distinguishing Aroclor and non-Aroclor sources to Chicago Air*, Science of The Total Environment 823 : 153263.
- Norris, G.; Duvall, R.; Brown, S. and Bai, S. (2014).** *EPA positive matrix factorization (PMF) 5.0 fundamentals and user guide*, <https://www.epa.gov/air-research/epa-positive-matrix-factorization-50-fundamentals-and-user-guide>, PMF 5.0 User Guide, US-EPA.
- Oh, J.; Shunthirasingham, C.; Zhan, F.; Li, Y.; Lei, Y. D.; Ben Chaaben, A.; Lu, Z.; Lee, K.; Gobas, F. A. P. C.; Hung, H. and Wania, F. (2024).** *Passive Air Sampling Networks Combined with Multivariate Statistics Reveal Widespread Non-Aroclor Polychlorinated Biphenyl Sources to the Canadian Atmosphere*, ACS ES&T Air 1 : 481-491.

- Ratier, A.; Dron, J.; Revenko, G.; Austruy, A.; Dauphin, C.-E.; Chaspoul, F. and Wafo, E. (2018).** *Characterization of atmospheric emission sources in lichen from metal and organic contaminant patterns*, Environmental Science and Pollution Research 25 : 8364-8376.
- Reuillard, M. (2023).** *Biosurveillance atmosphérique de la Métropole Aix-Marseille-Provence - campagne lichens 2021*, Rapport final LICHENS-2021, Institut Ecocitoyen pour la Connaissance des Pollutions.
- Reuillard, M. and Austruy, A. (2022).** *Cartographie et revégétalisation de sols contaminés aux HAPs –CARE Phase I*, Rapport CARE phase I, Institut Ecocitoyen pour la Connaissance des Pollutions.
- Reuillard, M.; Austruy, A. and Dron, J. (2024).** *Biosurveillance lichénique: exposition atmosphérique des territoires de Gardanne et Meyreuil*, Rapport final, Institut Ecocitoyen pour la Connaissance des Pollutions.
- Sylvestre, A. (2016).** *Caractérisation de l'aérosol industriel et quantification de sa contribution aux PM 2.5 atmosphériques*, Thèse de doctorat, Université Aix-Marseille.
- Sylvestre, A.; Mizzi, A.; Mathiot, S.; Masson, F.; Jaffrezo, J. L.; Dron, J.; Mesbah, B.; Wortham, H. and Marchand, N. (2017).** *Comprehensive chemical characterization of industrial PM2.5 from steel industry activities*, Atmospheric Environment 152 : 180-190.
- R Core Team (2021).** *R: A Language and Environment for Statistical Computing*, <https://www.R-project.org/>, R Core Team, Vienna.
- Zhang, S. (2016).** *Analyse dynamique, en champ proche et à résolution temporelle fine, de l'aérosol submicronique en situation urbaine sous influence industrielle*, Thèse de doctorat, Université du Littoral Côte d'Opale.

ANNEXES

Annexe A :
Détail des résultats des modélisations PMF

Annexe B :
Correspondances entre valeurs calculées et mesurées (métaux)

Annexe C :
Correspondances entre valeurs calculées et mesurées (HAP et PCDD)

Annexe D :
Correspondances entre valeurs calculées et mesurées (PCB)

Annexe E :
Répartition des facteurs en valeurs de contribution au paramètre « total »

Annexe A
Détail des résultats des erreurs « DISP » des modélisations PMF

Mélange entre facteurs (swaps), baisse de la valeur Q (%dQ drop), nombre de résidus dont les écarts entre valeurs mesurées et modélisées est > 3 écarts-type (> 3·SD), valeur Q (Qtrue), rapport entre Q modélisé et mesuré, et nombre d'étapes nécessaires pour converger.

* x F = nombre de facteurs et c = contrainte

** Nombre total de swaps (nombre de facteurs avec au moins 1 swap)

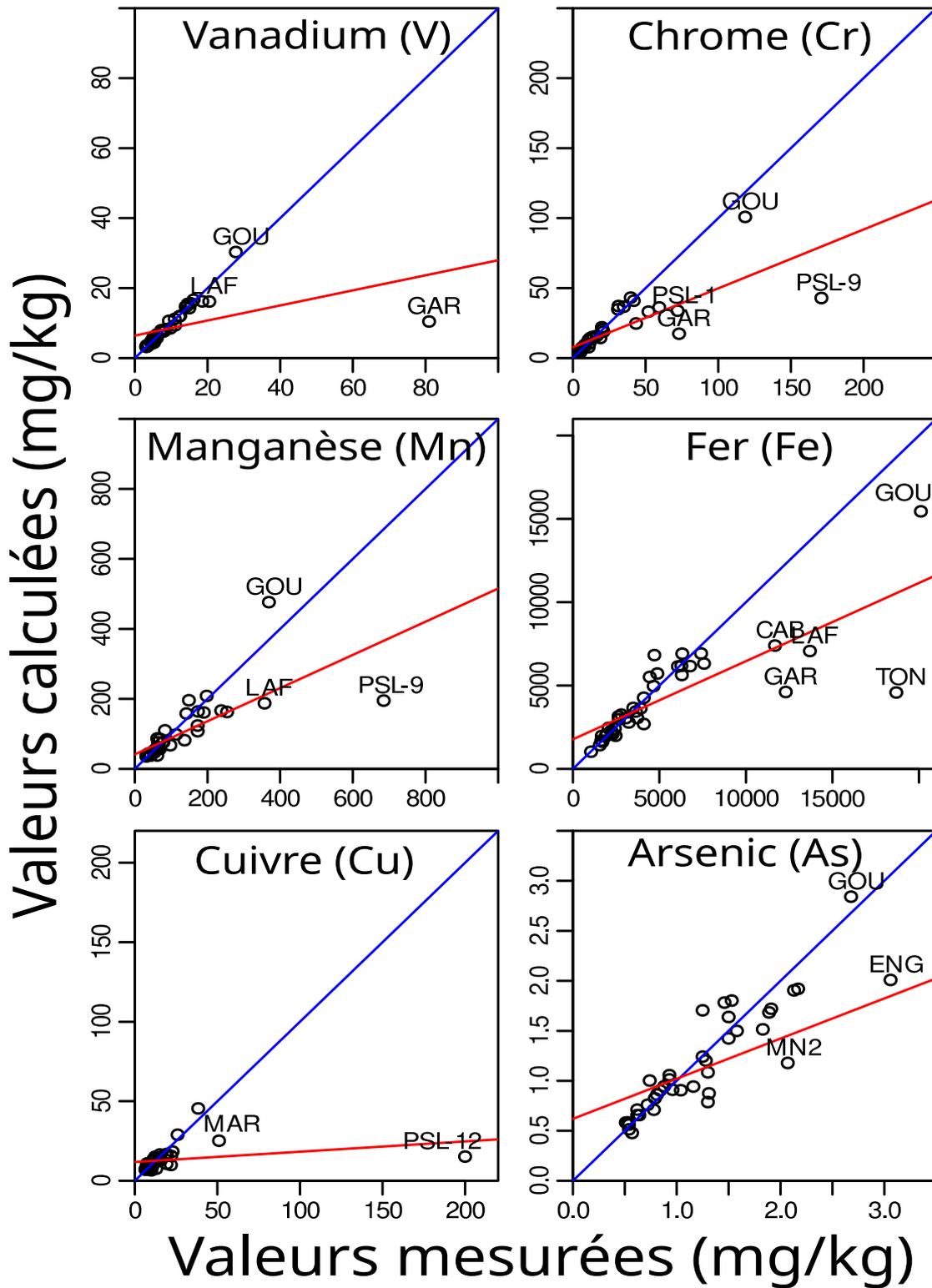
Modèle*	swaps (facteurs)**	% dQ drop	résidus > 3·SD	Qtrue	Qtrue/Qexp	steps
3 F	NA (NA)	na	Al-GAR, V-GAR, Cu-PSL12, DD-GOU, DD-CPH, DF-GOU, DF-CAB, PCB6-PSL3	1139	1.124	432
3 F + c	0 (0)	1.36	na	1143		262
4 F	NA (NA)	na	Al-GAR, V-GAR, Cu-PSL12, PCB6-PSL3	918	0.974	351
4 F + c	0 (0)	2.70	na	922		190
5 F	40 (3)	0.06	Cu-PSL12 & Mo-GAR	769	0.884	505
5 F + c	14 (4)	3.27	na	773		230
6 F	6 (3)	0.03	Cu-PSL12 & Mo-GAR	646	0.810	438
6 F + c	15 (3)	0.09	na	651		311
7 F	0 (0)	8.03	Cu-PSL12	561	0.772	1162
7 F + c	0 (0)	6.18	na	564		309
8 F	576 (8)	20.25	Cu-PSL12	482	0.737	930
8 F + c	17 (4)	10.04	na	485		930

Annexe B
Correspondances entre valeurs calculées et mesurées (métaux)

Valeurs calculées par le modèle PMF en fonction des valeurs mesurées pour le vanadium (V), le chrome (Cr), le manganèse (Mn), le fer (Fe), le cuivre (Cu) et l'arsenic (As).

En bleu : droites idéales « calculé = mesuré ».

En rouge : droites de corrélation réelles entre valeurs calculées par le modèle PMF et valeurs mesurées « calculé = a*mesuré + b » où « a » est la pente et « b » l'intersection à 0.



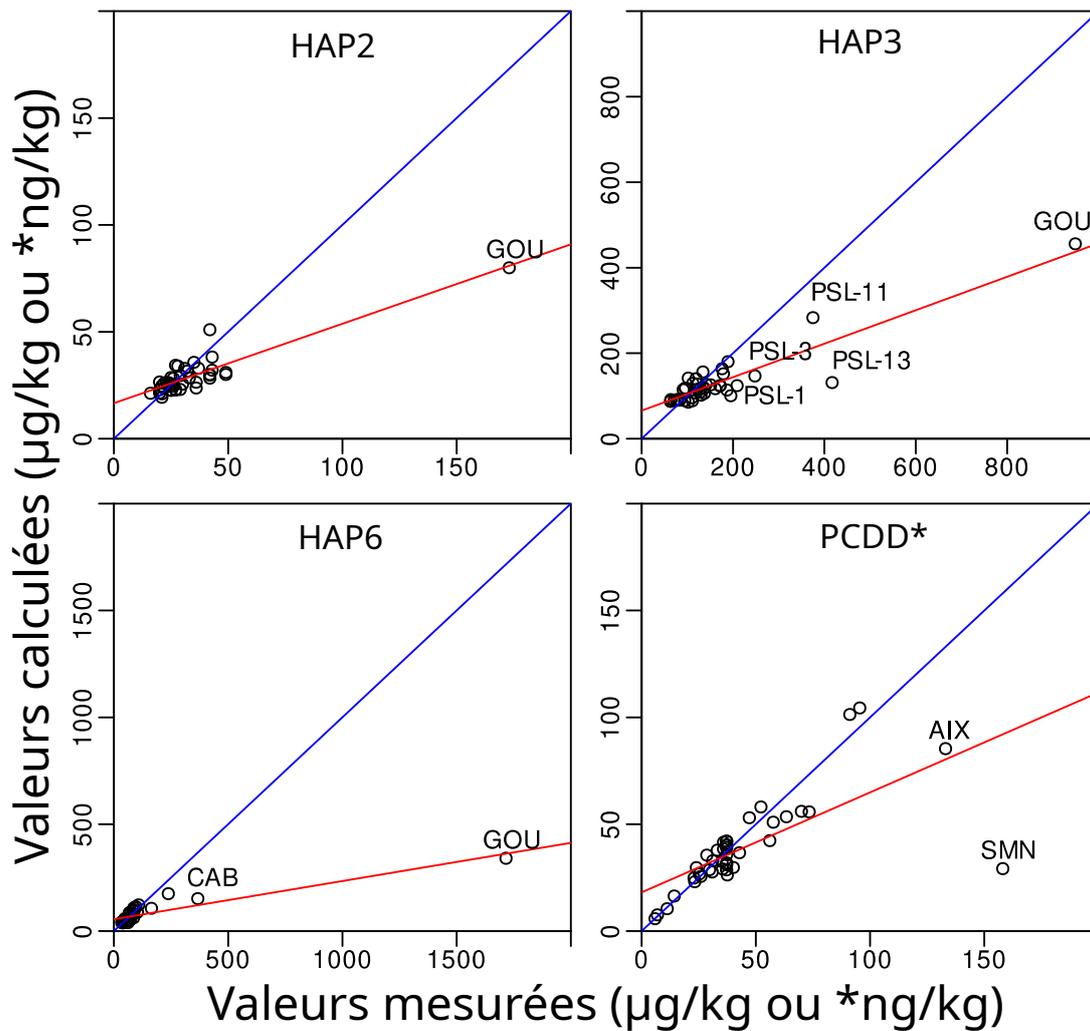
Annexe C

Correspondances entre valeurs calculées et mesurées (HAP et PCDD)

Valeurs calculées par le modèle PMF en fonction des valeurs mesurées pour les HAP à 2 cycles (HAP2), 3 cycles (HAP3) et 6 cycles (HAP6) et pour les dioxines (PCDD).

En bleu : droites idéales « calculé = mesuré »

En rouge : droites de corrélation réelles entre valeurs calculées par le modèle PMF et valeurs mesurées « calculé = $a \cdot \text{mesuré} + b$ » où « a » est la pente et « b » l'intersection à 0.

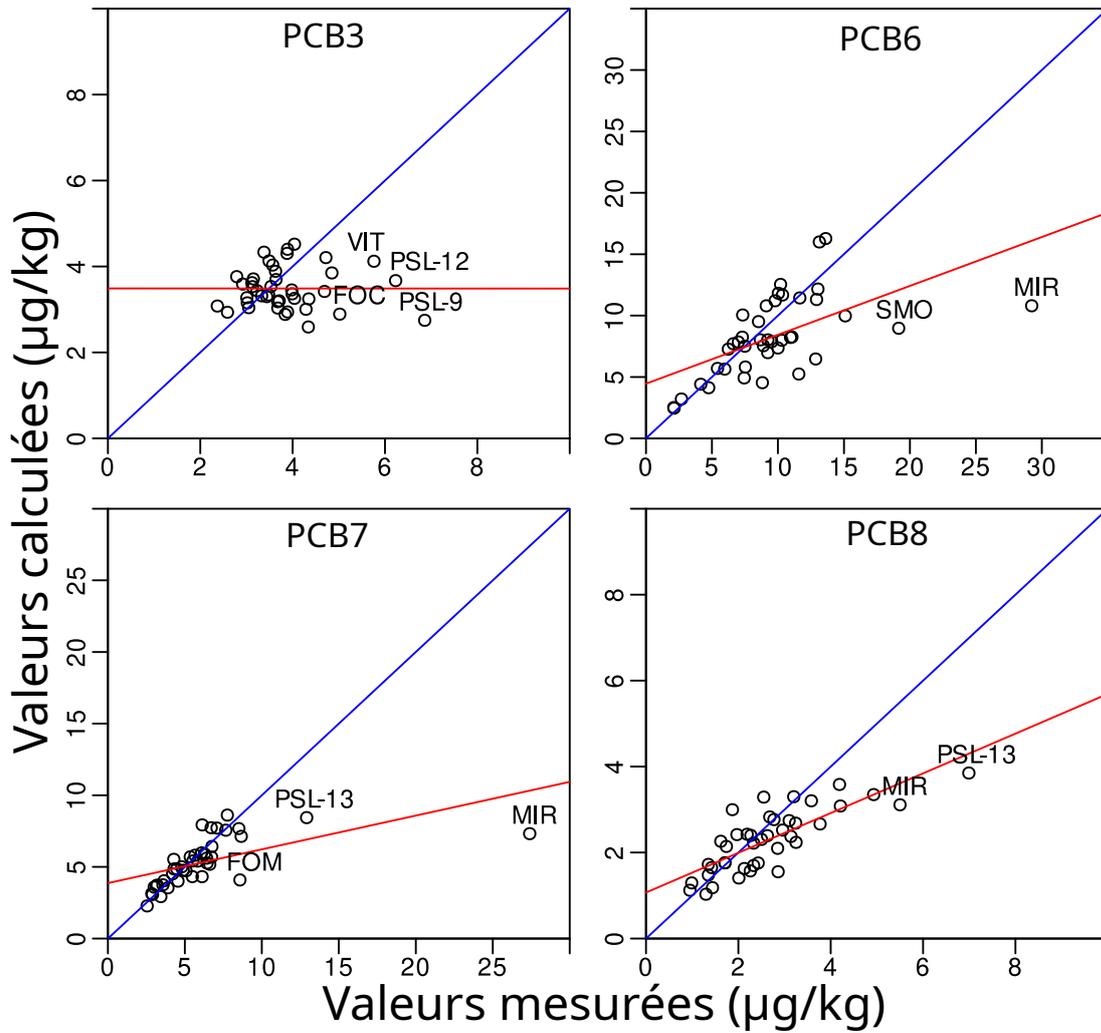


Annexe D
Correspondances entre valeurs calculées et mesurées (PCB)

Valeurs calculées par le modèle PMF en fonction des valeurs mesurées pour les PCB à 3 chlores (PCB3), 6 chlores (PCB6), 7 chlores (PCB7) et 8 chlores (PCB8).

En bleu : droites idéales « calculé = mesuré »

En rouge : droites de corrélation réelles entre valeurs calculées par le modèle PMF et valeurs mesurées « calculé = a*mesuré + b » où « a » est la pente et « b » l'intersection à 0.



Annexe E

Répartition des facteurs en valeurs de contribution au paramètre « total »

Les valeurs représentées sont les parts brutes de contribution des facteurs au paramètre « total » (somme des concentrations normalisées S_j , voir section 2.6). Pour un site donné, la somme des contributions brutes des 6 facteurs doit être égale à celle représentée pour ce site dans la Figure 1. Les unités sont arbitraires, mais sur la même base pour les 6 graphiques. Cartes Coraline SouL.

