

Note technique

Travaux financés par le ministère chargé de l'environnement

GUIDE EUROPEEN POUR LA MISE EN ŒUVRE D'OUTILS STATISTIQUES DE TYPE « MODELES RECEPTEURS » : PRINCIPAUX POINTS D'ATTENTION

O. FAVEZ (Ineris)

SYNTHESE

L'élaboration et l'évaluation des plans d'action visant à améliorer la qualité de l'air nécessitent l'identification et la quantification des principales sources d'émissions. Deux grands types de méthodologies sont alors principalement utilisées : (i) celles se basant sur l'utilisation de modèles numériques permettant de simuler le devenir des polluants dans l'atmosphère à partir de cadastres d'émission, de la paramétrisation des conditions météorologiques et des processus physico-chimiques de (trans-)formation des PM ; (ii) celles se basant sur la mesure des propriétés physico-chimiques des particules sur un site récepteur (« modèles récepteurs »). Ces derniers rendent compte des situations réelles et sont aujourd'hui couramment utilisés au sein de la communauté scientifique et par les acteurs de la surveillance de la qualité de l'air.

Un guide méthodologique pour leur mise en œuvre a été édité en 2014 et révisé en 2019 dans le cadre des travaux du forum Européen pour la modalisation de la qualité de l'air (FAIRMODE).¹ Le LCSQA/Ineris s'est fortement impliqué dans ces travaux, en collaboration avec le *Joint Research Center*. **La présente note relie les différents chapitres du guide européen (*indiqués en bleu*) à certains des principaux points d'attention à considérer lors de la mise en œuvre de ce type d'outils statistiques, et en particulier de la *Positive Matrix Factorization* (PMF). Elle ne constitue donc pas un substitut à une lecture attentive et une application rigoureuse du guide européen.**

¹ <https://ec.europa.eu/jrc/en/publication/european-guide-air-pollution-source-apportionment-receptor-models>

1. GENERALITES

Les « modèles récepteurs » sont aujourd'hui couramment utilisés pour l'identification et la quantification des sources de polluants atmosphériques, en particuliers les particules fines ([introduction](#)). Pour ces dernières, les méthodologies appliquées incluent généralement la détermination d'une grande diversité de traceurs spécifiques, suivie de l'utilisation d'outils statistiques basés soit sur une bonne connaissance *a priori* des principales sources et de leur profil d'émission (par ex. CMB : *Chemical Mass Balance*, [chapitres 8](#)), soit sur une détermination *a posteriori* de ces sources à l'aide de profils d'émission déterminés par ailleurs (par ex. PMF : *Positive Matrix Factorization*, [chapitres 9-13](#)).

Le jeu de données peut être obtenu par analyse d'échantillons prélevés sur le site d'étude ou par mesures en ligne ([chapitre 4](#)). Outre les méthodologies analytiques classiquement utilisées pour la caractérisation chimique des particules (e.g., chromatographie ionique, spectrométrie de masse, analyses thermiques et/ou optiques), d'autres approches, telles que comme l'analyse isotopique ([chapitre 15](#)), la résonance magnétique nucléaire ([chapitre 16](#)) ou encore la spectroscopie infrarouge à transformée de Fourier ([chapitre 17](#)), peuvent également être mises en œuvre.

Quelle que soit l'approche utilisée, l'obtention de résultats pertinents nécessite **une bonne connaissance de l'environnement immédiat du site d'étude** et des principales sources d'émission pouvant y influencer les concentrations de particules ([chapitre 1](#)), ainsi que **l'élaboration d'une stratégie expérimentale adaptée** ([chapitre 2](#)).

Les **informations fournies par les « modèles récepteurs » sont directement exploitables** (et communicables) **pour les sources primaires** de particules (e.g., combustion de biomasse, émissions directes à l'échappement automobile, sels de mer, ...). En revanche, **pour les aérosols secondaires, les contributions obtenues ne peuvent permettre de statuer sur l'impact réel d'une modification des concentrations de leurs précurseurs gazeux** (dont les sources restent, en outre, fréquemment à déterminer) **sur leurs niveaux de concentrations particulaires**. En effet, comme les processus de transformations secondaires ne sont généralement pas linéaires, les polluants émis par les différentes sources (anthropiques ou naturelles) réagissent entre eux, selon des processus que les modèles récepteurs ne peuvent ni discriminer ni quantifier. Pour exemple, les sorties PMF permettent d'estimer la contribution des émissions directes de particules par le transport routier, mais n'apportent pas d'information chiffrée sur l'influence réelle de cette même source dans la formation d'aérosols secondaires (en particulier le nitrate d'ammonium et les aérosols organiques secondaires).

Par ailleurs, **ce type d'approche n'apporte pas d'information sur l'origine géographique des facteurs de sources** mis en évidence. Une analyse complémentaire des variables météorologiques (dont les directions et vitesse de vent) et/ou des rétro-trajectoires des masses d'air peut néanmoins permettre de donner des informations qualitatives sur ces origines géographiques ([chapitre 18](#)).

L'application de « modèles récepteurs » implique généralement l'utilisation de nombreuses hypothèses conceptuelles et empiriques. Elle requière donc **une bonne maîtrise de leur principe de fonctionnement et de la littérature scientifique afférente**. De ce fait, il est également recommandé d'accompagner les rapports d'étude ou publications scientifiques

présentant les résultats obtenus d'un **maximum d'information sur les choix réalisés au cours des différentes étapes de leur mise en œuvre** ([chapitre 20](#)).

2. PRINCIPES DE FONCTIONNEMENT ET PREPARATION DU JEU DE DONNEES

La plupart des « modèles récepteurs » reposent sur une analyse factorielle multivariée des moindres carrés pondérés. Ces méthodologies visent à résoudre un système d'équations intégrant une matrice de données chimiques X ($[n \times m]$, où n est le nombre d'échantillons et m le nombre d'espèces chimiques disponibles), en déterminant le nombre de facteurs (p) expliquant les variations de concentrations observées, ainsi que la matrice des contributions G ($n \times p$) et celle des profils chimiques F ($p \times m$) de ces facteurs, selon le **principe de conservation de masse**, tel que :

$$X = G \times F + E$$

Où E est la matrice résiduelle (partie de la matrice X non expliquée par le modèle)

$$x_{ij} = \sum_{k=1}^p g_{ik} \cdot f_{kj} + e_{ij}$$

Avec :

- x_{ij} : concentration de l'espèce j le jour i ,
- g_{ik} : contribution du facteur k le jour i ,
- f_{kj} : fraction du facteur k provenant de l'espèce j ,
- e_{ij} : fraction de l'espèce j le jour i non expliquée par le modèle (résidu).

Ce type de système d'équations possède une infinité de solutions, parmi lesquelles il convient de retenir celles permettant de minimiser une fonction « Q » qui correspond au rapport du résidu (e_{ij}) de l'espèce j le jour i sur l'incertitude (σ_{ij}) associée à la concentration de cette espèce :

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

Ainsi, il est nécessaire de fournir les incertitudes de mesure en tant que données d'entrée et **il est primordial de déterminer, au préalable, ces incertitudes de la manière la plus juste et la plus cohérente** (e.g., d'une technique d'analyse chimique à une autre) **possible** ([chapitre 7](#)). En particulier, la sous-estimation des incertitudes associées aux plus faibles concentrations d'une ou plusieurs espèce(s) chimique(s) peut conduire à des biais importants dans la résolution du système d'équation proposé. Ainsi, les valeurs inférieures à la limite de détection (LD), ainsi que les données manquantes, sont fortement pondérées, en leur attribuant individuellement de larges incertitudes, de façon à minimiser leur influence dans la solution finale.

Afin de rendre compte de l'ensemble des principales sources et mécanismes de formation des polluants, il convient d'utiliser **un jeu de données aussi dense que possible**, en termes de nombre d'échantillons et/ou d'espèces chimiques mesurées et en fonction des objectifs et des modalités de mise en œuvre de l'étude ([chapitres 2 et 3](#)).

Dans tous les cas, **une validation de l'ensemble du jeu de données utilisé**, intégrant une première évaluation des sources de particules potentiellement significatives (e.g., analyse des

concentrations et des variations temporelles des espèces majeures et des traceurs de sources présumés), **doit être réalisée préalablement à la mise en œuvre du modèle récepteur** ([chapitres 5 et 6](#)).

A l'autre bout de la chaîne de traitement de données, **la validation des solutions retenues passe notamment par une vérification de leur stabilité** malgré, par exemple, la modification aléatoire d'une partie des données d'entrée ([chapitre 19](#)).

3. SPECIFICITES DE LA PMF

Si les approches de type CMB étaient fréquemment utilisées par le passé, les études récentes privilégient généralement les approches de type PMF ([introduction](#)). Ces dernières présentent notamment l'avantage de ne pas dépendre des hypothèses de départ concernant le nombre et la nature des sources à quantifier. Pour ce faire, le modèle prend en compte l'évolution de la concentration des espèces chimiques mesurées et **regroupe dans un même facteur les fractions d'espèces chimiques évoluant de la même manière d'un échantillon à l'autre**. Chaque facteur obtenu doit ensuite être identifié et attribué à une source (et/ou catégorie de particules) en se basant sur les traceurs chimiques caractérisant chaque profil de facteur résolu et leur répartition dans celui-ci. **Ce type d'approche exige néanmoins une longue série temporelle de données chimiques** (e.g., typiquement une ou plusieurs années) pour des prélèvements journaliers un jour sur trois sur la base des espèces chimiques classiquement analysées dans ce type d'étude). Compte-tenu de l'importance des coûts relatifs à l'obtention de ces jeux de données, les études mettant en œuvre la PMF ne peuvent être que ponctuelles dans l'espace et/ou dans le temps.

Enfin, **l'utilisation de ce type d'outil statistique revêt une part de subjectivité** ([chapitre 9](#)), notamment liée au choix des hypothèses de départ et/ou de la solution finalement retenue ([chapitres 10 et 11](#)). Il existe différentes approches pour réduire l'ambiguïté rotationnelle responsable de la multitude de solutions possibles, et ainsi d'aboutir à une solution la plus réaliste possible (limitant également les mélanges de facteurs), telle **l'incorporation dans le modèle de contraintes externes supplémentaires en se basant sur des connaissances a priori de certains profils de sources** ([chapitre 12](#)). Ces contraintes peuvent inclure des profils de sources connus, des contributions à des sources connues, ou encore des variables météorologiques. Il convient d'être prudent dans le choix du type/nombre de contraintes à imposer au modèle, afin **d'éviter de l'orienter de facto vers une solution attendue initialement par l'utilisateur**.