

Identification et quantification des contributions de sources de poussières fines en milieu urbain

Marie Chavent ¹
en collaboration avec: Brigitte Patouille¹ and Hervé Guéguan²

¹IMB, Université Bordeaux

²CENBG, Gradignan

Particules en suspension ou PM (Particulate Matter)

- Métrologie :
 - PM10 : de diamètre inférieur à $10\mu\text{m}$
 - PM2.5 : de diamètre inférieur à $2.5\mu\text{m}$ (fines)
 - PM1 : de diamètre inférieur à $1\mu\text{m}$ (ultrafines)
- Partie inhalable des poussières en suspension dans l'air
- Un indicateur de pollution atmosphérique

Particules en suspension ou PM (Particulate Matter)

Quelle est la contribution à l'empoussièrement en PM_{2.5} de chaque émetteur possible comme par exemple :

- les installations de combustion urbaine (incinérateurs...),
- les chauffages résidentiels,
- le transport routier (diesels, pneumatiques...),
- le transport aérien,
- ...

Appel à propositions du programme Primequal-Predit

- Programme de recherche scientifique piloté par le MEDAD et l'ADEME
- Objectifs du programme :
 - “Fournir les bases scientifiques et les outils nécessaires aux décideurs et aux gestionnaires pour surveiller et améliorer la **qualité de l'air** et réduire les risques pour la santé et l'environnement”,
 - “Renforcer l'**interdisciplinarité** en favorisant les collaborations entre des équipes de différentes disciplines”.

Projet “Mise au point d’un outil de détermination de la contribution des sources de poussières fines”

Projet réalisé **en collaboration** (AIRAQ-Arcane-IMB) et en trois étapes :

Etape 1 : campagne de prélèvement des filtres à PM2.5 sur le site d’Anglet (**AIRAQ**),

Etape 2 : caractérisation (par technique PIXE) des filtres par leurs concentration en différents éléments chimiques (Fe, Cu, ...) (**ARCANE**),

Etape 3 : détermination de l’origine et la contribution des sources de PM2.5 (par des outils statistiques) (**IMB**)

Etape 1 : campagne de prélèvement



Station fixe d'Anglet à proximité :

- d'une voie de circulation importante (Bayonne-Anglet-Biarritz)
- de trois agglomérations
- d'une zone industrielle au nord-est (aciérie, cimenterie)
- océan Atlantique à l'ouest

Etape 1 : campagne de prélèvement

Prélèvements de filtres à particules PM2.5:

- Périodes de 12 heures :
 - 7h-19h : “jour”
 - 19h-7h : “nuit”
- Deux campagnes de 4 semaines
 - 3 Juin-2 Juillet : “été”
 - 23 Nov- 25 Déc : “hiver”

Soit environ 64 filtres par saison.

Etape 2 : caractérisation chimique des filtres

Technique PIXE :

- analyse chimique élémentaire des filtres par utilisation de faisceaux d'ions de haute énergie produits par un accélérateur de particules
- 15 éléments dosés
(Al, Si, P, S, Cl, K, Ca, Ti, Mn, Fe, Ni, Cu, Zn, Br, Pb)
- pas de préparation d'échantillon, conservation après la mesure, analyse multi-élémentaire quantitative en quelques dizaines de minutes et système automatisé.

Etape 2 : caractérisation chimique des filtres

Deux tableaux de données (été et hiver)

Date	<i>C</i>	<i>Al</i>	<i>Si</i>	...	<i>K</i>	<i>Ca</i>	...	<i>Br</i>	<i>Pb</i>
23-11-05 day	7300	92	75	...	163	35	...	7	10
23-11-05 night	9600	135	90	...	211	23	...	7	77
24-11-05 day	11000	175	137	...	241	69	...	8	19
24-11-05 night	5300	36	31	...	94	44	...	9	7
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24-12-05 day	21000	< 2	< 1	...	266	< 1	...	7	18
24-12-05 night	18100	18	< 1	...	307	< 1	...	7	19
25-12-05 day	23300	37	22	...	311	12	...	7	14
25-12-05 night	36100	< 2	< 1	...	277	< 1	...	10	19

Table: Extrait d'un tableau de données

Etape 3 : Identification et quantification des sources

A partir des deux tableaux de données :

- Partie 1 : **Identifier** les sources ayant produit l'empoussièrement des filtres:
Analyse Factorielle (ACP + rotation)
- Partie 2 : **Identifier** et **quantifier** la part de chaque source à l'empoussièrement des filtres
Modèle source-récepteur (PMF + normalisation)

La matrice de concentration $X_{n \times p}$

	Al_2O_3	SiO_2	...	K	Ca	...	Br	Pb	C_{org}
23-11-05 day	250	160	...	163	35	...	7	10	4645.3
23-11-05 night	365	193	...	211	23	...	7	77	6564.2
24-11-05 day	475	292	...	241	69	...	8	19	7017.1
24-11-05 night	96	66	...	94	44	...	9	7	2805.5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24-12-05 day	3	1	...	266	0.5	...	7	18	16240.7
24-12-05 night	49	1	...	307	0.5	...	7	19	15878.3
25-12-05 day	101	46	...	311	12	...	7	14	20636.8
25-12-05 night	3	1	...	277	0.5	...	10	19	30885

→ Campagne hiver: $n = 61$ filtres et $p = 14$ composants chimiques.

La matrice de concentration $X_{n \times p}$

Les pré-traitements :

- les valeurs inférieures aux limites de détection (LD) remplacées par $LD/2$,
- Les éléments Al , Si , S et Fe ont été remplacés par les composés Al_2O_3 , SiO_2 , SO_4 , Fe_2O_3
- Deux éléments supprimés et la colonne C_{org} donnant la concentration des éléments non mesurés a été ajoutée :

$$C_{org} = C - (Al_2O_3 + SiO_2 + P + SO_4 + Fe_2O_3 + Cl + K + Ca + Mn + Cu + Zn + Br + Pb)$$

Les données météo

- Données de la station météo de Biarritz (2,5 km du site de prélèvement): températures et direction du vent (en degrés) toutes les heures.
- Matrice des données météo après pré-traitement :

	N	N-E	E	S-E	S	S-O	O	N-O
23-11-05 day	17%	0%	25%	8%	8%	0%	0%	42%
23-11-05 night	25%	17%	33%	25%	0%	0%	0%	0%
24-11-05 day	0%	0%	8%	17%	50%	25%	0%	0%
24-11-05 night	0%	0%	0%	0%	8%	9%	58%	25%
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Partie 1 : Identification des sources par analyse factorielle

1. Recherche de groupes d'éléments chimiques corrélés :
 - classification de variables,
 - Analyse en composantes principales (ACP) et rotation varimax.
2. Association (si possible) d'un groupe à une source à l'aide d'éléments traceurs :
 - *Zn, Pb* : source industrielle
 - *Fe, Cu* : source automobile
 - *S* : source combustion
 - *Al, Si* : source terrigène
 - *Cl* : source embruns

ACP et rotation varimax

Le modèle d'analyse en facteur s'écrit de manière simplifiée pour $q \leq r \leq p$:

$$\tilde{\mathbf{X}} = \mathbf{F}\mathbf{A}' + \mathbf{E}$$

où

- $\tilde{\mathbf{X}}$ est la matrice \mathbf{X} centrée-réduite,
- $\mathbf{F}_{(n,q)}$ est une matrice de **scores factoriels**,
- $\mathbf{A}_{(p,q)}$ est une matrice de **saturation**s,
- \mathbf{E} est le reste de l'approximation de $\tilde{\mathbf{X}}$ par $\mathbf{F}\mathbf{A}'$.

ACP et rotation varimax

Estimation du modèle par ACP :

- Décomposition en valeurs singulières de $\mathbf{Z} = \frac{1}{n}\tilde{\mathbf{X}}$

$$\mathbf{Z} = \mathbf{U}\Lambda^{1/2}\mathbf{V}'$$

avec

- Λ la matrice diagonale des valeurs propres non nulles de $\mathbf{Z}'\mathbf{Z}$ et $\mathbf{Z}\mathbf{Z}'$,
- \mathbf{U} est la matrice des vecteurs propres de $\mathbf{Z}\mathbf{Z}'$,
- \mathbf{V} est la matrice des vecteurs propres de $\mathbf{Z}'\mathbf{Z}$.

ACP et rotation varimax

Estimation du modèle par ACP :

- On en déduit

$$\tilde{\mathbf{X}} = \underbrace{n\mathbf{U}^{(q)}}_{\mathbf{F}} \underbrace{\Lambda^{(q)1/2}\mathbf{V}^{(q)'}}_{\mathbf{A}'} + \mathbf{E}$$

- **F**, la matrice des **composantes principales standardisées** ($\mathbf{f}_k = \frac{\psi_k}{\sqrt{\lambda_k}}$),
- **A**, la matrice des **corrélations variables-facteurs** ($a_{jk} = \text{corr}(\mathbf{x}_j, \mathbf{f}_k)$).

ACP et rotation varimax

- On a aussi

$$\tilde{\mathbf{X}} = \underbrace{\mathbf{F}\mathbf{T}}_{\tilde{\mathbf{F}}} \underbrace{\mathbf{T}'\mathbf{A}'}_{\tilde{\mathbf{A}}} + \mathbf{E}$$

où

- \mathbf{T} est une matrice orthonormale de dimension (q, q) ,
- $\tilde{\mathbf{F}}$ est matrice des facteurs après rotation,
- $\tilde{\mathbf{A}}$ matrice des saturations après rotation.
- On montre que $\tilde{a}_{jk} = \text{corr}(\mathbf{x}_j, \tilde{\mathbf{f}}_k)$
- Définir \mathbf{T} pour les coefficients de $\tilde{\mathbf{A}}$ soient proches de 0 ou de 1

Application à la matrice des concentrations

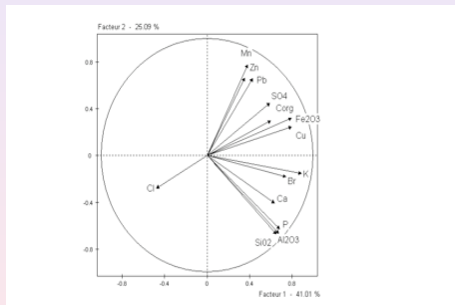
- Code SAS :

```
PROC FACTOR data=hiverorg method=prin nfactors=5  
outstat=load out=fact rotate=varimax;  
var Al2O3 SiO2 P SO4 Cl K Ca Mn Fe2O3 Cu Zn Br Pb Corg;  
run;
```

- Choix de $q = 5$ facteurs :
 - Pas d'association claire d'une source à un facteur au delà de $q = 5$
 - 90,93% de variance expliquée

Matrice A des corrélations avant rotation

	f_1	f_2	f_3	f_4	f_5
Al_2O_3	0,672	-0,663	0,221	-0,193	0,023
SiO_2	0,649	-0,669	0,254	-0,203	-0,058
P	0,682	-0,629	0,223	-0,238	0,006
SO_4	0,589	0,449	-0,413	-0,22	0,175
Cl	-0,474	-0,281	0,192	0,681	0,344
K	0,888	-0,154	-0,209	0,004	0,284
Ca	0,638	-0,405	0,099	0,399	-0,195
Mn	0,384	0,776	0,183	0,093	-0,247
Fe_2O_3	0,793	0,319	0,038	0,269	-0,36
Cu	0,796	0,248	-0,098	0,232	-0,36
Zn	0,352	0,663	0,589	-0,072	0,246
Br	0,746	-0,182	-0,126	0,352	0,363
Pb	0,428	0,659	0,519	-0,088	0,296
C_{org}	0,6	0,297	-0,613	-0,025	0,222



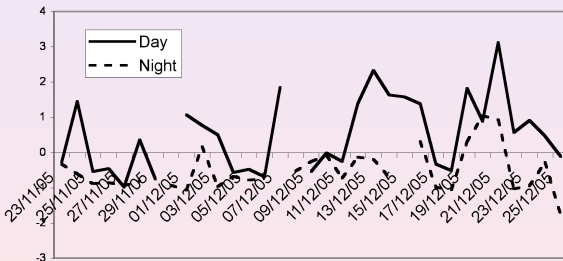
Matrice \tilde{A} des corrélations après rotation

	\tilde{f}_1	\tilde{f}_2	\tilde{f}_3	\tilde{f}_4	\tilde{f}_5
Al_2O_3	0.981	0.087	-0.042	0.070	-0.038
SiO_2	0.979	0.012	-0.055	0.104	-0.074
P	0.972	0.090	-0.017	0.071	-0.092
SO_4	-0.028	0.765	0.247	0.180	-0.345
Cl	-0.153	-0.274	-0.136	-0.181	0.879
K	0.597	0.716	0.111	0.233	0.031
Ca	0.608	0.091	-0.113	0.560	0.272
Mn	-0.279	0.119	0.604	0.582	-0.238
Fe_2O_3	0.198	0.282	0.289	0.848	-0.112
Cu	0.213	0.359	0.161	0.816	-0.149
Zn	-0.029	0.053	0.977	0.129	-0.044
Br	0.490	0.615	0.097	0.281	0.392
Pb	0.004	0.163	0.969	0.126	-0.054
C_{org}	-0.018	0.893	0.021	0.222	-0.160

Source ?	
\tilde{f}_1	Terrigène
\tilde{f}_2	Combustion
\tilde{f}_3	Industrie
\tilde{f}_4	Automobile
\tilde{f}_5	Embruns

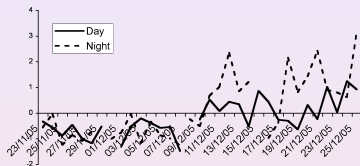
Confirmation des associations sources-facteurs

(a) Scores après rotation du facteur 4 associé à la source automobile

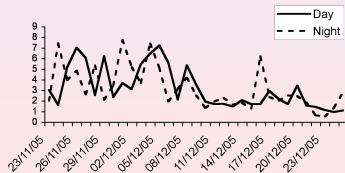


Confirmation des associations sources-facteurs

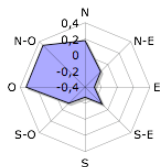
(b) Scores après rotation du facteur 2 associé à la source combustion



(c) Evolution des températures



(c) Corrélations des scores après rotation du facteur 5 associé à la source embruns aux 8 directions du vent



Remarques et limites

- Phase exploratoire,
- Problème si :
 - deux sources corrélées,
 - un composant est émis par deux sources,
 - un élément chimique traceur d'une source non mesuré,
- **Pas de quantification** des contributions des sources.

Partie 2 : Identification et quantification des sources par un modèle de type source-récepteur

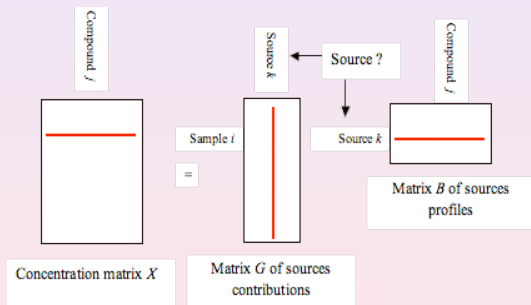
Basé sur l'équation de la balance des masses $x_{ij} = \sum_{k=1}^q g_{ik} b_{jk}$
avec:

- x_{ij} la concentration du composant chimique j dans le filtre i ,
- g_{ik} la concentration en particules de la source k dans le i ème filtre (**contributions des sources** aux filtres);
- b_{jk} la proportion (fraction de masse) de composant chimique j émis par source k (**profils des sources** supposés constants).

Le modèle source-recepteur

Forme matricielle :

$$\mathbf{X} = \mathbf{GB}' + \mathbf{E}$$



PMF et normalisation

Pour approximer **G** (contributions des sources) et **B** (compositions des sources) à partir de **X** et d'un nombre q de sources fixé, on a appliqué les deux étapes suivantes :

1. **Méthode PMF** (Positive Matrix Factorization) : factoriser **X** en un produit **HC'** de rang q sous contrainte de positivité des coefficients.
2. **Normalisation** (scaling): choix d'une solution **G** and **B** en normalisant les colonnes de **H** et **C** par des coefficients définis pour satisfaire certaines contraintes physiques du modèle.

Positive Matrix Factorization

La méthode PMF développée par Paatero and Tapper (1994) minimise :

$$Q(\mathbf{H}, \mathbf{C}) = \sum_{i=1}^n \sum_{j=1}^p \left(\frac{x_{ij} - \sum_{k=1}^q h_{ik} c_{jk}}{\sigma_{ij}} \right)^2$$

avec σ_{ij} une mesure d'incertitude sur l'observation x_{ij}
→ choix de σ_{ij} ?

Positive Matrix Factorization

- Choix utilisé dans le programme PMF de l'US Environment Protection Agency¹ est :

$$\sigma_{ij} = \begin{cases} 2LD & \text{if } x_{ij} \leq LD, \\ \sqrt{(\theta x_{ij})^2 + LD^2} & \text{if } x_{ij} > LD, \end{cases}$$

→ choix de θ ...

- Choix utilisé ici : $\sigma_{ij} = s_j$ l'écart-type empirique.
→ la minimization de l'erreur quadratique non pondérée approxime mieux les colonnes des composants chimiques de plus forte variance

¹E.P.A. PMF 1.1 Users's guide, <http://www.epa.gov/heads/products/pmf/pmf.htm>

Normalisation

• $\mathbf{X} = \mathbf{HC}' + \mathbf{E}$ où \mathbf{HC} est une solution de PMF.

•
$$\sum_{k=1}^q h_{ik} c_{jk} = \sum_{k=1}^q h_{ik} \frac{\beta_k}{\beta_k} c_{jk},$$

• $\mathbf{X} = \mathbf{GB}' + \mathbf{E}$ avec $g_{ik} = h_{ik} \beta_k$ et $b_{jk} = \frac{c_{jk}}{\beta_k}$.

→ Définir les constantes de normalisation β_k pour que \mathbf{G} et \mathbf{B} puissent être des contributions et des profils de sources.

Normalisation

Soit γ_i la concentration en particules dans le filtre i :

$$\gamma_i = \sum_{k=1}^q g_{ik} + e_i = \sum_{k=1}^q \beta_k h_{ik} + e_i$$

d'où β_k solution des moindres carrés.

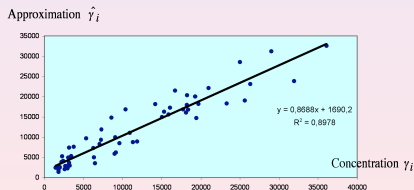
Remarque : si $\forall i, \sum_{j=1}^p x_{ij} = \gamma_i$, on devrait avoir

$$\sum_{j=1}^p b_{kj} = \sum_{j=1}^p \frac{c_{jk}}{\beta_k} = 1 \rightarrow \beta_k = \sum_{j=1}^p c_{jk}.$$

Application à la matrice des concentrations

- Choix de $q = 5$ sources (ACP)
- PMF : Programme matlab avec entrées-sorties Excel (Jianhang et Laosheng, 2004)
- Qualité de l'approximation :

Ajustement de γ par $\hat{\gamma}$:



La matrice des profils des sources **B**

	Source1	Source2	Source3	Source4	Source5
Al_2O_3	41.6	0.0	0.0	0.0	0.0
SiO_2	18.5	0.0	0.0	0.0	0.0
P	6.2	0.0	0.7	0.0	0.6
SO_4	10.1	15.3	59.6	12.2	22.6
Cl	0.0	0.0	0.0	0.0	74.5
K	12.9	0.0	3.6	1.5	0.1
Ca	2.4	1.6	0.0	0.0	1.4
Mn	0.0	0.2	0.3	0.0	0.0
Fe_2O_3	6.7	15.0	12.7	0.2	0.0
Cu	0.3	0.7	0.4	0.0	0.0
Zn	0.7	0.0	16.3	0.0	0.3
Br	0.2	0.1	0.2	0.0	0.5
Pb	0.3	0.0	6.2	0.0	0.0
C_{org}	0.0	67.1	0.0	85.9	0.0

- ne permet pas d'identifier clairement les sources,
- calcul des contribution relatives des sources aux composants chimiques.

Identification des sources

Contributions relatives des sources aux composants chimiques :

	Source1	Source2	Source3	Source4	Source5
Al_2O_3	100.0	0.0	0.0	0.0	0.0
SiO_2	100.0	0.0	0.0	0.0	0.0
P	81.5	0.5	3.9	8.2	6.0
SO_4	4.5	9.5	10.7	67.9	7.5
Cl	0.0	0.0	0.0	0.0	100.0
K	38.8	0.0	4.4	56.7	0.2
Ca	42.0	39.6	0.0	0.0	18.4
Mn	0.0	54.9	33.1	8.5	3.5
Fe_2O_3	19.0	59.2	14.4	7.4	0.0
Cu	18.5	56.8	9.1	15.6	0.0
Zn	9.0	0.5	87.5	0.0	3.1
Br	19.4	12.1	5.7	33.4	29.4
Pb	10.7	0.0	81.4	7.9	0.0
C_{org}	0.0	8.0	0.0	92.0	0.0

Sources associées :

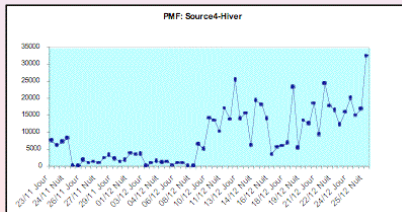
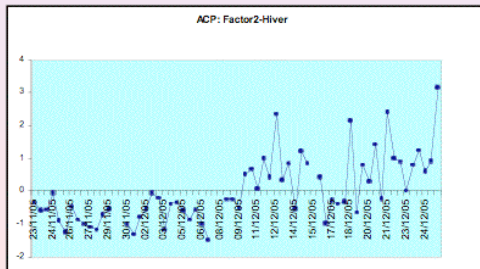
Source1	Terrigène
Source2	Automobile
Source3	Industrie
Source4	Combustion
Source5	Embruns

→ 92% du C_{org} émis par la source 4.

Confirmation des associations

Corrélations entre les colonnes de \mathbf{G} et $\tilde{\mathbf{F}}$:

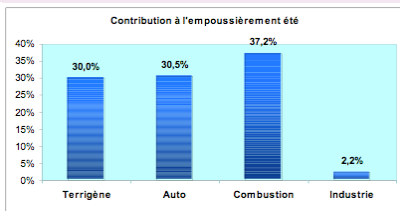
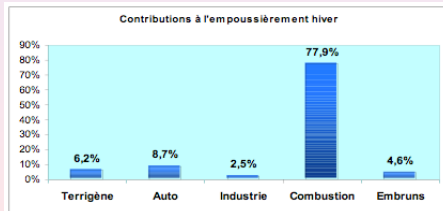
	Source 1	Source 2	Source 3	Source 4	Source 5
Factor 1	0.98	-0.18	-0.11	-0.02	-0.18
Factor 2	0.11	0.12	0.06	0.95	-0.30
Factor 3	-0.05	-0.09	0.98	0.02	-0.15
Factor 4	0.12	0.96	0.10	0.11	-0.22
Factor 5	-0.02	-0.13	-0.10	-0.27	0.88



Quantification de la contribution des sources

A partir de la matrice **G** des contributions :

- Graphiques d'évolution des quantités des sources dans les filtres,
- Graphique de la contribution de chaque source à l'empoussièrément des filtres sur la période d'analyse :



- Chavent M., Guégan H., Kuentz V., Patouille B. , Saracco J., 2007. Air pollution sources apportionment in a french urban site, Case Studies in Business, Industry and Government Statistics journal, Vol 1, Issue 2, 119-129.
- Hopke, P.K., 1991. Receptor Modeling for Air Quality Management. Elsevier, Amsterdam.
- Jianhang, L., Laosheng, W., 2004. Technical details and programming guide for a general two-way positive matrix factorization algorithm. Journal of Chemometrics 18, 519-525.
- Paatero, P., Tapper, U., 1994. Positive Matrix Factorization: a non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5, 111-126.