



**HAL**  
open science

## **Caractérisation et identification de la variabilité temporelle des sources de particules dans un environnement intérieur : approche statistique**

Rachid Ouaret, Anda Ionescu, Olivier Ramalho, Viorel Petrehus, Yves Candau

### ► **To cite this version:**

Rachid Ouaret, Anda Ionescu, Olivier Ramalho, Viorel Petrehus, Yves Candau. Caractérisation et identification de la variabilité temporelle des sources de particules dans un environnement intérieur : approche statistique. 29ème Congrès Français sur les Aérosols, Jan 2014, Paris, France. 6 p. <hal-01064389>

**HAL Id: hal-01064389**

**<https://hal.science/hal-01064389v1>**

Submitted on 16 Sep 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# CARACTÉRISATION ET IDENTIFICATION DE LA VARIABILITÉ TEMPORELLE DES SOURCES DE PARTICULES DANS UN ENVIRONNEMENT INTÉRIEUR : APPROCHE STATISTIQUE

R.Ouaret\*(1, 2), A. Ionescu(2), O.Ramvalho(1), V. Petrehus(3), Y. Candau(2)

(1) Université Paris-Est, Centre Scientifique et Technique du Bâtiment, Direction Santé Confort

(2) Centre d'Études et de Recherche en Thermique, Environnement et Systèmes (CERTES EA 3481),  
Université Paris-Est Créteil.

(3) Université Technique des Constructions de Bucarest, Département de Mathématiques, Roumanie

rachid.ouaret@cstb.fr

## TEMPORAL CHARACTERIZATION AND SOURCE IDENTIFICATION OF INDOOR AIR PARTICULATE MATTER: A STATISTICAL APPROACH

### ABSTRACT

Particle number concentration (particles 0.3-20  $\mu\text{m}$  in aerodynamic diameter) and  $\text{CO}_2$  concentration were monitored in an office occupied by one adult during 45 days. The temporal characterization and source identification of indoor particulate matter were studied using principal component analysis (PCA) and Independent Component Analysis (ICA). The resulting temporal profiles were explained by relating the calculated source contributions to the information provided by the autocorrelation function. Four particle sources were identified and characterized. The source of 3,5-8,75  $\mu\text{m}$  diameter particles was associated with occupant activity. A random emission source was found to be associated with the larger particles and one continuous emission source was associated with the fine particles.

### RÉSUMÉ

Les mesures de concentrations de  $\text{CO}_2$  et de concentrations en nombre de particules ont été effectuées simultanément dans un bureau pendant 45 jours consécutifs. L'identification et la caractérisation temporelle de sources ont été réalisées en utilisant l'analyse en composante principale et l'analyse en composantes indépendantes. La fonction d'autocorrélation nous a permis de comprendre la structure temporelle des fluctuations (polluants et sources) et de faire une typologie des émissions (intermittentes et chroniques).

### 1. Introduction

L'identification des sources de particules dans les environnements intérieurs et la quantification de leur contribution nécessite de développer des connaissances sur la variabilité temporelle de la concentration en particules dans l'air. Depuis plusieurs années, les méthodes de séparation des sources "en aveugle" ont été essentiellement développées et appliquées dans le domaine de l'air extérieur ou du traitement du signal. Étant donné que les concentrations de certains polluants sont en général plus élevées à l'intérieur qu'à l'extérieur, et que les environnements intérieurs représentent 80% de l'exposition totale aux polluants de l'air, il est devenu nécessaire de mieux comprendre les fluctuations par l'évaluation des contributions des différentes sources pour un polluant donné.

Dans ce travail, une étude sur la caractérisation des particules émises par différentes sources potentielles dans un bureau individuel a été entreprise. Plus précisément, il s'agit de caractériser la structure des profils temporels des sources de particules. Pour cela, nous avons d'abord entrepris une analyse statistique des concentrations en nombre de particules dans un environnement de type bureau, étape préalable au développement de modèles permettant de quantifier la contribution des sources. Ensuite, nous ferons appel aux méthodes appliquées en statistique exploratoire des données et en traitement du signal pour la séparation des sources, notamment l'analyse en composantes principales (ACP) et l'analyse en composantes indépendantes (ACI). A notre connaissance, ces derniers modèles n'ont pratiquement jamais été employés pour les environnements intérieurs sur des mesures avec un pas de temps fin.

L'objectif de notre travail, dont cet article n'est qu'une première étape, est de montrer en quoi certains outils statistiques les plus récents peuvent apporter des éclairages nouveaux sur la compréhension des fluctuations des sources particulières dans un environnement intérieur.

## 2. Matériels et méthode

### Description du bureau instrumenté

Cette étude est menée dans un bureau individuel situé dans les locaux du Centre Scientifique et Technique du Bâtiment à Champs sur Marne. Cet établissement se situe dans une zone périurbaine, à quelques kilomètres de grands axes routiers comme l'autoroute de l'Est (A4). Le bureau se situe dans un bâtiment relativement peu élevé, la trajectoire des vents autour de l'ensemble des bureaux devrait être peu modifiée. Le bâtiment où se situe le bureau instrumenté est entouré de deux rues principales où la circulation automobile est modérée. En particulier, des camions viennent livrer régulièrement des marchandises au magasin situé à une centaine de mètres du bureau. En outre, deux parkings sont à proximité du bureau, et 4 à 5 véhicules stationnent chaque jour dans chaque parking. Ces conditions pourraient être à l'origine de sources particulières importantes, notamment de particules fines.

### Mesure des séries temporelles

La concentration en nombre de particules et la concentration du CO<sub>2</sub> ont été enregistrées en continu avec un pas temps de 10 minutes. Les concentrations en nombre de particules sont enregistrées par un compteur optique des particules sur 15 gammes de taille entre 0,3 et 20 µm (Dust Monitor 1.108, Grimm). Les mesures de concentrations de CO<sub>2</sub> ont été effectuées en parallèle dans le même bureau par un capteur infrarouge non dispersif. La campagne de mesure a été effectuée pendant 45 jours consécutifs, du 19 février 2011 à minuit jusqu'au 4 avril 2011 à 23h50.

Les résultats de la statistique descriptive montrent que, en termes de variabilité, plus les particules sont fines, plus leur écart-type est grand et plus elles sont corrélées entre elles. En effet, les coefficients de corrélation de Pearson pour les particules entre 0,3 µm et 0,80 µm varient entre 0,75 et 0,98. Les coefficients de corrélation entre les particules > 20 µm et < 4,5 µm ne dépassent pas 0,35.

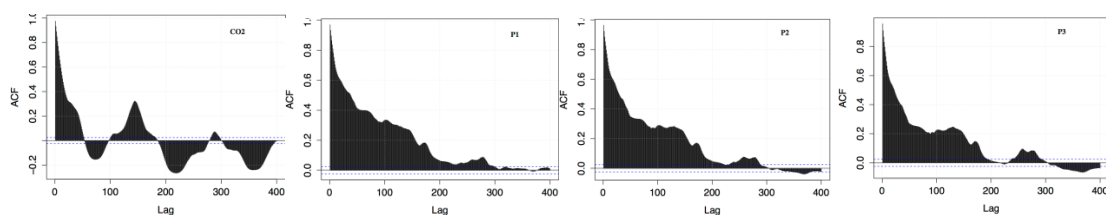
Les coefficients d'asymétrie des paramètres mesurés sont positifs, leurs queues de distribution sont étalées vers la droite ; ceci laisse présager des distributions log-normales.

Outre la mesure d'aplatissement et d'asymétrie, les distributions de probabilités ont été estimées par une méthode non paramétrique (à noyau gaussien). Une distribution log-normale a été également ajustée aux données. De façon générale, on remarque que pour les particules fines, les distributions en nombre sont caractérisées par une densité de probabilité de type log-normale ; ce n'est pas le cas pour les grosses particules, en particulier du fait que leur nombre est très souvent voisin de zéro.

### Classification et statistique des séries temporelles

L'approche statistique développée dans ce travail consiste à exploiter l'ensemble des données mesurées et à construire un jeu de variables permettant d'identifier les éventuelles structures temporelles des sources. En effet, l'analyse des profils diurnes de tous les polluants met en évidence la présence des niveaux de concentration très élevés durant les heures de travail, notamment, pour les grosses particules et le CO<sub>2</sub>, tandis que pour les particules comprises entre 0,30-0,40 µm et 0,4-0,5 µm de diamètre, les profils diurnes ne montrent aucune variabilité significative.

Dans le but de comprendre la structure de ces fluctuations, on s'intéresse, d'une part aux propriétés intrinsèques des séries (autocorrélation), et d'autre part aux spécificités de la variabilité, en faisant une classification pour trouver des journées-type.



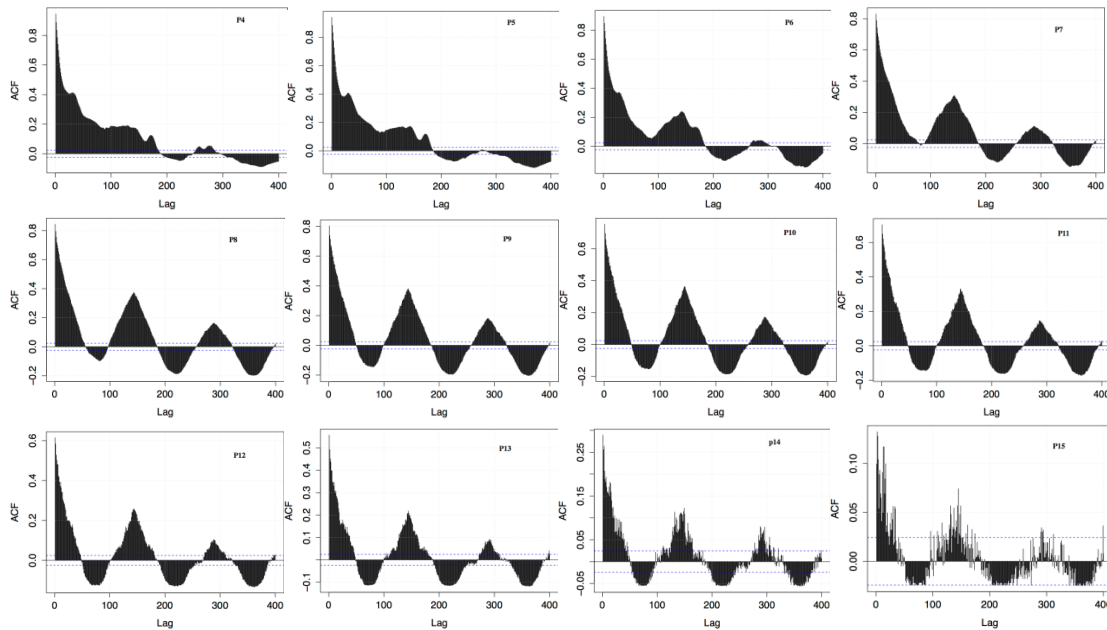


Figure 1: Fonctions d'autocorrélation des séries temporelles particulières des 15 fractions granulométriques et du  $CO_2$  ; le retard (lag) est en  $\times 10$  minutes. Les deux lignes horizontales pointillées représentent l'intervalle de confiance à 95% pour les autocorrélations d'un processus *i.i.d.* gaussien.

La Figure 1 présente les courbes de la fonction d'autocorrélation (ACF) (le corrélogramme) calculées à partir des séries temporelles de la concentration en nombre de particules et du niveau du  $CO_2$ .

On remarque que les valeurs de l'ACF de la série des concentrations de  $CO_2$  sont significativement non nulles jusqu'à ce que le corrélogramme soit tronqué à partir de 8h (480 minutes) ; le même type de comportement est observé pour les fractions 1,6-2,0  $\mu m$  jusqu'à la taille 5,0-7,5  $\mu m$ . La plupart des fonctions d'autocorrélation se comportent comme un mélange de fonctions exponentielles/sinusoïdales amorties, en particulier pour les particules de tailles moyenne et le  $CO_2$ . En effet, l'ACF de ces fractions traduit l'aspect de saisonnalité des émissions de sources.

En revanche, les corrélogrammes correspondants aux tailles de particules fines présentent des autocorrélations relativement élevées avec une décroissance très lente, le premier passage à zéro se produisant après 2 jours seulement. Les fortes valeurs de l'ACF expriment ici, à la fois la force et l'étalement de la persistance. En fait, cette persistance matérialisée par une décroissance lente des autocorrélations (corrélation à long terme) représente un mécanisme complexe d'émission pour les particules fines. Les processus d'émissions génèrent ainsi des fluctuations multi-fréquentielles, c'est-à-dire selon différentes échelles temporelles.

### Méthodes de séparation des sources

Plusieurs méthodes de séparation des sources sont utilisées en environnement extérieur. Ces méthodes sont basées sur les techniques de factorisation matricielle des mesures au récepteur en profils et contributions des sources. On rencontre deux cas de figure : (i) les profils des sources sont connus et dans ce cas on applique des méthodes de bilan de masse (CMB, donner une référence) ;(ii) les profils de sources ne sont pas connus et dans ce cas on applique des méthodes statistiques de factorisation. Parmi ces dernières méthodes, l'Analyse en Composantes Principales (ACP) (Jolliffe 2002) fournit une factorisation matricielle sous contrainte d'orthogonalité des facteurs, mais procure aussi des facteurs négatifs non interprétables physiquement. Les algorithmes de factorisation positive tels que la Factorisation Matricielle Positive (PMF) (Paatero et Tapper, 1994) ont été utilisés avec succès dans un bon nombre d'études environnementales (Gianini, 2012; Srimuruganandam, 2012). Dans l'air intérieur, l'application de la PMF a fait ses débuts vers les années 2000 et on cite quelques exemples. Ogulei *et al.* (2006) ont extrait par la PMF neuf contributions de sources de particules à partir des mesures dans une maison située en milieu urbain (à Reston, Etats-Unis), occupée par un couple durant 5 mois environ. Les principales contributions en nombre de particules sont liées à des activités de cuisines (91%) ; ce résultat corrobore les conclusions de (Zhao et al. 2007). La contribution extérieure correspondant aux particules issues du trafic ne représente que 3% du nombre total de particules dans la maison ; en outre, 4% de variance n'est pas expliquée par le modèle PMF.

### 3. Résultats sur la caractérisation temporelle des sources

Le but de cette analyse est d'identifier les principales structures temporelles des sources. Selon la littérature, les occupants jouent un rôle majeur soit directement en tant que sources de particules au travers de la resuspension des particules déposées soit indirectement par la modulation des sources présentes (imprimantes, fenêtres etc.).

Pour atteindre notre objectif, nous allons procéder à l'application de différentes méthodes statistiques de séparation des sources pour les séries temporelles. Les deux approches suivies : l'analyse en composantes principales (ACP) et l'analyse en composantes indépendantes (ACI), donneront des informations complémentaires.

L'Analyse en Composantes Principales a été appliquée aux séries temporelles des polluants (toutes les fractions de particules), chaque série étant considérée une variable. Les variables ont été standardisées (centrées-réduites) pour neutraliser l'effet d'échelle provoqué essentiellement par le nombre des particules fines ; chaque variable reçoit ainsi la même importance, quelle que soit sa moyenne et son écart-type. Les six premières composantes principales expliquent 89% de la variance totale (Figure 2 : à gauche), avec 55% de variabilité expliquée par la première composante principale. Le corrélogramme associé à la première composante principale (Figure 2 : à droite) possède une forme sinusoïdale et alterne de signe. On retrouve le même type de corrélogramme pour le CO<sub>2</sub> et pour les particules de tailles moyennes (0,62 µm à 0,2 µm). Cette structure la FAC laisse présager l'existence d'une source commune liée, qui peut être liée à l'occupation, dont le principal indicateur est le CO<sub>2</sub>.

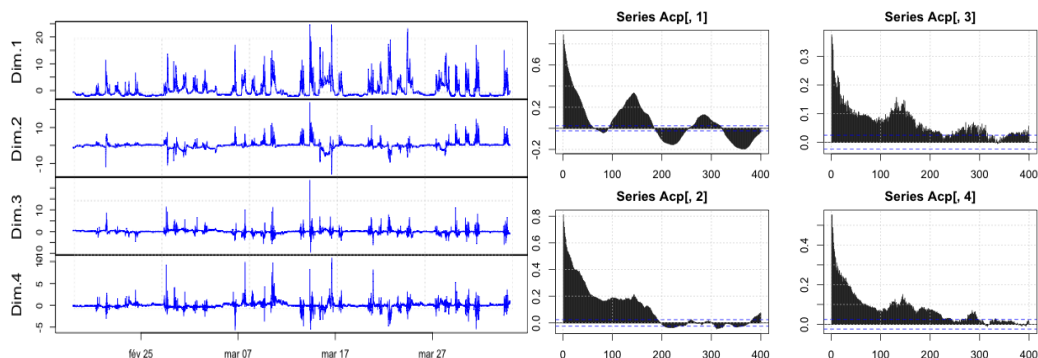


Figure 2: A gauche: Profils temporels des quatre premières composantes principales. A droite: corrélogramme associés aux composantes principales.

La présence des valeurs négatives des composantes de l'ACP et de l'ACI est non interprétable physiquement. C'est pour cela que nous proposons l'utilisation de l'une des structures intrinsèques des fluctuations, en particulier la fonction d'autocorrélation qu'on applique aux composantes (variables artificielles, combinaisons linéaires des variables d'origines) qui représentent également des séries chronologiques. En fait, les caractéristiques de la variabilité des phénomènes qui ont généré les séries temporelles sont mal connues, la seule information dont nous disposons est l'historique des séries mesurées. Par conséquent, la modélisation de la variabilité des polluants consiste à imiter le mécanisme d'émission de sources ou de génération des données sans retrouver explicitement les fonctions d'émissions, et ce uniquement à travers la chronique des séries. Justement, la fonction d'autocorrélation (ACF) nous permet d'explorer en partie les structures statistiques des séries.

La seconde étape de la modélisation est la déconvolution aveugle par l'ACI ; pour ce type d'analyse le nombre de sources (composantes) est une donnée d'entrée. Plusieurs simulations ont été faites afin d'évaluer le critère RMSE (racine carrée de l'erreur quadratique moyenne) en fonction du nombre de composante indépendantes (Figure 3). Cette erreur représente l'écart entre les variables initiales (séries temporelles) et les variables reconstituées à partir des composantes indépendantes.

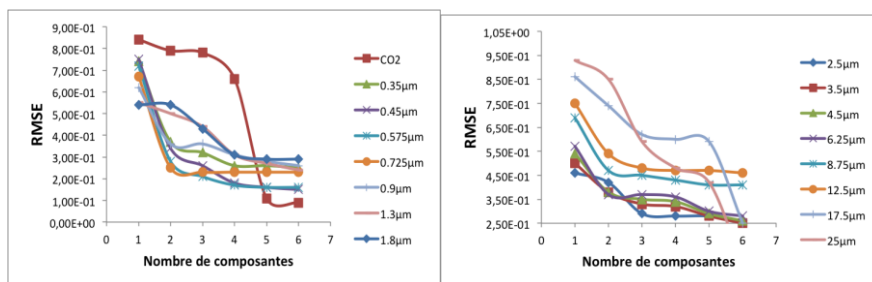


Figure 3: Evolution du RMSE en fonction du nombre de composantes indépendantes. A gauche: évaluation de la reconstitution par l'ACI des fractions de tailles médianes 0,35 –1,8 µm et du CO<sub>2</sub>. A Droite : évaluation de la reconstitution par l'ACI des fractions de tailles médianes 2,5 –25 µm.

Le choix du nombre de composantes retenues est basé sur la recherche du premier minimum local du RMSE. On remarque qu'en se basant sur ce critère, le nombre de composantes à retenir pour chaque fraction varie en fonction de la taille médiane des particules. Pour les séries 0,72 µm, 0,9 µm, 6,25 µm et 8,75 µm, seulement 2 composantes suffisent pour atteindre le minimum local. Sinon pour la plupart des fractions, 4 composantes sont nécessaires, pour cela, nous retiendrons 4 composantes indépendantes pour l'ACI.

La Figure 4 montre les fluctuations temporelles des quatre composantes indépendantes obtenues sur les 15 fractions, ainsi que les graphiques de la fonction d'autocorrélation de toutes les composantes indépendantes (c1...c4) obtenues sur les 15 fractions.

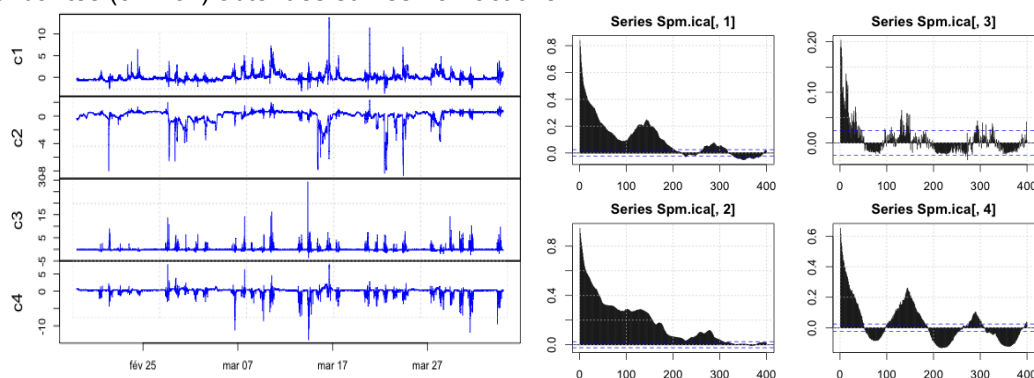


Figure 4: A droite : Profils temporels des quatre premières composantes indépendantes. A gauche: corrélogrammes associés aux composantes indépendantes)

La deuxième composante indépendante obtenue est inversée en signe, mais ceci n'est qu'un artefact, car les solutions sont déterminées à un facteur près, donc la multiplication par le facteur (-1) est envisageable. Les autocorrélations de cette composante sont significatives pour des retards importants (2 jours), montrant ainsi l'influence des observations très éloignées sur les plus récentes, d'où l'idée d'une mémoire de la série. La lente décroissance hyperbolique de la fonction d'autocorrélation indique que la variabilité des concentrations des particules fines admet les caractéristiques des séries à mémoire longue. De façon générale, une série présentant des corrélations à long terme est définie comme une série présentant une fonction d'autocorrélation de la forme :  $z(t) = t^{-\lambda}$  avec  $0 < \lambda < 1$ . En d'autres termes, la fonction d'autocorrélation suit une loi puissance. Ce type de comportement est lié aux émissions des sources chroniques.

Tout comme la première composante principale, la structure de décroissance du corrélogramme de la composante indépendante 4 indique que celle-ci présente des fluctuations du niveau de particules similaire aux fluctuations du niveau de CO<sub>2</sub>. Avec quatre composantes, l'ACI permet de mettre en évidence l'existence d'une source responsable des fluctuations des grosses particules (la composante indépendante 3), tandis qu'il faut six dimensions pour l'ACP (95% de variance expliquée) pour pouvoir faire apparaître ce type de variation au niveau des autocorrélations. Les valeurs de l'autocorrélation de la composante indépendante 3 se présentent comme des pics discrets à des faibles corrélations. Cette discontinuité est souvent liée à des phénomènes de faibles probabilités d'occurrence.

#### 4. Analyse de la contribution des sources

Dans la figure 5, on présente la contribution des sources estimée par l'ACI pour chaque composante indépendante sur chaque fraction de particules. Les quatre composantes indépendantes contribuent de manière distributive en fonction de la taille des particules. Pour les fractions les plus fines de particules (0,35  $\mu\text{m}$ , 0,45  $\mu\text{m}$ , 0,575  $\mu\text{m}$ ), la contribution de la composante 2 avoisine les 80% de l'ensemble des contributions des autres composantes. Elle représente en plus, la source la moins contributive pour les grosses particules (>20  $\mu\text{m}$ ). La composante 3 représente sur cette gamme de particules plus de 95% de l'ensemble des autres contributions de sources. Notons par ailleurs la particularité de discontinuité de la fonction d'autocorrélation pour la composante 3 : elle caractérise les émissions dues aux phénomènes aléatoires, tandis que pour la composante 2, les autocorrélations de type fonction de puissance caractérise les émissions des sources déterministes.

Pour la fraction de taille 0,9  $\mu\text{m}$ , les composantes 2 et 1 partagent à part égale plus de 93% de l'ensemble des contributions. La composante 4 caractérisé par une décroissance sinusoïdale des autocorrélations correspond à la source majoritaire pour la gamme de taille 3,5-8,75  $\mu\text{m}$ . Elle est de plus associée aux fluctuations de concentration du  $\text{CO}_2$ , elle est par conséquent reliée à la présence d'occupants dans la pièce. Il s'agit vraisemblablement de particules remises en suspension ou générées par les activités de l'occupant dans le bureau.

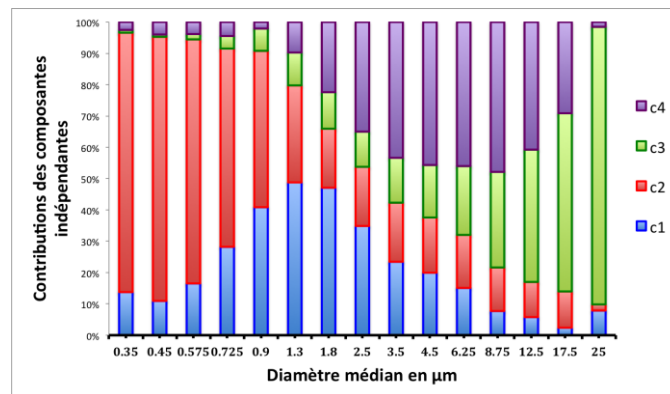


Figure 5. Contributions des composantes indépendantes sur la granulométrie des 15 variables.

#### 5. Conclusion

Cette étude a permis d'identifier les différentes typologies de sources par l'utilisation de l'analyse en composantes principales (ACP) et l'analyse en composantes indépendantes (ACI). Les sources de type aléatoire ou chronique ont été mises en évidence par l'analyse de la fonction d'autocorrélation (ACF) des fluctuations des composantes. Les composantes identifiées restent toutefois difficiles à interpréter sans l'appui de variables externes directement associées à des sources définies. En perspective, ces résultats seront utilisés pour les comparer avec les mesures dans d'autres environnements intérieurs.

#### Références

- Gianini, M. F. D., Fischer, A., et al. (2012): Comparative source apportionment of PM10 in Switzerland for 5 2008/2009 and 1998/1999 by Positive Matrix Factorisation, *Atmos. Environ.*, 54, 149-158.
- Jolliffe, I. T. (2002). *Principal Component Analysis*, Springer.
- Ogulei, D., P. K. Hopke, et al. (2006). "Analysis of indoor particle size distributions in an occupied townhouse using positive matrix factorization." *Indoor Air* 16(3): 204-215.
- Paatero, P. and Tapper, U. (1994): Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics*, 5, 111-126.
- Srimuruganandam, B. and Shiva Nagendra, S. M. (2012): Application of positive matrix factorization in characterization of PM10 and PM2.5 emission sources at urban roadside, *Chemosphere*, 88, 120-130.
- Viana, C., T.A.J. Kuhlbusch, X. Querol, et al. (2008). Source apportionment of particulate matter in Europe: A review of methods and results, *Journal of Aerosol Science*, 39(10), 827-849.
- Zhao, W., P. K. Hopke, et al. (2007). Use of an expanded receptor model for personal exposure analysis in schoolchildren with asthma. *Atmospheric Environment* 41(19): 4084-4096.