Identification of atmospheric $PM_{2.5}$ sources using Non-negative Matrix Factorization modeling in urban-industrial mixed environments

A. Kfoury[1,2], F. Ledoux[1,2], A. Limem[1,3], G. Delmaire[1,3], G. Roussel[1,3], D. Courcot[1,2,*]

[1]Université Lille Nord de France, F-59000 Lille, France
[2]Université du Littoral Côte d'Opale, UCEIV EA4492, F-59140 Dunkerque, France
[3]Université du Littoral Côte d'Opale, LISIC EA4491, F-62228 Calais, France
Keywords: $PM_{2.5}$, receptor modeling, ICP-MS, source apportionment.
Author email: dominique.courcot@univ-littoral.fr

Receptor modeling constitutes a complementary method added to chemical analysis, which are essential to reveal the nature of PM sources. Between the different available modeling algorithms, many types of receptor models like multivariate models and positive matrix factorization (PMF) developed by Paatero (Paatero et *al.*, 1993) provide a flexible approach that can effectively use the information content of chemical data, and has been used worldwide as a credible modeling method (Pandolfi et *al.*, 2008).

Many problems can be encountered when using receptor modeling, the most important one of which is the identification of closely related sources such as "crustal dust" and "traffic resuspended soil dust" for example (Viana et *al.*, 2008).

In order to resolve this problem, constraints must be included in the calculations in order to increase the sensitivity of the model on certain characteristics considered exclusive to each of the sources. Our non-negative matrix factorization is a new method that allows the use of measurement uncertainties as well as constraints in the calculations (Delmaire et *al.*, 2010).

Our model calculates the profile (F) and contribution (G) matrices using the following equation:

$$X = \mathbf{GF} + E$$

Equation 1. Matrix factorization equation (X: observation matrix; E: error)

Furthermore, it optimizes the calculations by minimizing the error between estimated and observed values:

$$\{G, F\} = \arg\min_{G,F} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(x_{ij} \_ (GF)_{ij}\right)^2$$

Equation 2. Minimizing the error between estimated and observed values

The constraints are acquired from the *a priori* knowledge on the sources under study. This information is available through literature reviewing or direct source samples' analysis.

Furthermore, the constraints can be included in two forms: equality and inequality. When applied, the first forces the calculations to consider a specific value of a chemical element in a profile, whereas the inequality constraint gives an order of abundance to each element between the different sources profiles.

Recent studies have showed that the use of receptor modeling that includes measurement uncertainties as well as constraints gave reliable results in the study of particulate pollution source apportionment (Viana et *al.*, 2008), especially when using chemical data representing the composition of PM collected from pollution mixed environments.

This method has been applied in our study using data on $PM_{2.5}$ samples collected in three medium sized cities subjected to a mixture of emission sources and located in northern France, a region in which PM limit values are exceeded regularly. $PM_{2.5}$ samples were taken from the sites during two sampling campaigns realized in winter 2010 and spring 2011, then analyzed for major and trace elements, major ions and total carbon contents. The data were modeled using NMF and the results allowed us to identify site specific and regional profiles.

The application of the NMF using the chemical data acquired from the analysis of about 350 $PM_{2.5}$ samples allowed us to identify 9 to 11 sources of PM pollution as well as their contributions depending on the sampling sites. Finally, the reconstructed concentrations obtained using NMF estimated values proved to be well correlated to the observed ones attained by direct chemical analysis.

The UCEIV laboratory contributes in the work of the Institut de Recherche en Environnement Industriel (IRENI), which is financed by the Communauté Urbaine de Dunkerque, the Région Nord Pas-de-Calais, the Ministère de l'Enseignement Supérieur et de la Recherche, the CNRS and European Regional Development Fund (ERDF).

Delmaire, G., Roussel, G., Hleis, D., Ledoux, F. (2010) *Journal Europeen des Systemes Automatises* **44**, 547-566

Paatero, P., Tapper, U. (1993). *Chemometrics and Intelligent Laboratory Systems* **18**, 183–194.

Pandolfi M, Viana M., Minguillo´n M.C., Querol X., Alastuey A., Amato F., Celades I., Escrig E., Monfort E. (2008). *Atmospheric Environment* **42**, 9007-9017.

Viana M., T.A.J. Kuhlbusch T.A.J., X. Querol X., Alastuey A., Harrison R.M., Hopke P.K., Winiwarter W., Vallius M., Szidat S., Prévôt A.S.H., Hueglin C., Bloemen H., Wåhlin P., Vecchi R., Miranda A.I., Kasper-Giebl A., Maenhaut W., Hitzenberger R., 2008, *Aerosol Science* **39**, 827–849.