Kullback-leibler NMF under linear equality constraints. Application to pollution source apportionment

Abdelhakim Limem, Gilles Delmaire, Gilles Roussel, Dominique Courcot

KULLBACK-LEIBLER NMF UNDER LINEAR EQUALITY CONSTRAINTS. APPLICATION TO POLLUTION SOURCE APPORTIONMENT

Limem Abdelhakim, Delmaire Gilles, Roussel Gilles, Courcot Dominique

Université Lille Nord de France, 59655 Villeneuve d'Ascq, France.
Laboratoire d'Informatique Signal Image de la Côte d'Opale, EA 4491
Université du Littoral Côte d'Opale Maison de la Recherche Blaise Pascal.
50, rue Ferdinand Buisson BP 699 62228 Calais Cedex, France.

Unité de Chimie Environnementale et Interactions sur le Vivant EA 4492
Université du Littoral Côte d'Opale
Maison de la Recherche en Environnement Industriel de Dunkerque.
145 Avenue Maurice Schumann 59140 Dunkerque, France.

ABSTRACT

Non negative matrix factorisation (NMF) coupled to divergence measure has been investigated in the frame of an application to polluant source identification. It relies on receptor modelling which considers the data matrix as the result of cumulative effects of p sources.
NMF aims at finding a contribution matrix G and a profile matrix F by minimizing a specific cost function. The focus is made here on the Kullback-Leibler divergence (KL) cost function. Linear equality constraints are incorporated into parts of the decomposition and general mutiplicative like expressions, which take into account these constraints, are derived.
This method is applied in the frame of source apportionment of particulate matter.

1. INTRODUCTION

Non-negative matrix factorization (NMF) is a well known algorithm used in blind sources separation [11].
The method is based on the approximation of non-negative data matrix X by the product of two non-negative matrices G and F. Equation (1) called receptor model expresses the error between the data and the approximated product:

$$X = GF \circ E \tag{1}$$

where the operator $\circ$ is element wise product and :

- X stands for the *n* x *m* data matrix. In the case of environmental studies, X involves chemical species concentrations for all samples. They are expressed in ng/m^3. Each element x_{ij} accounts for the concentration of the jth chemical specie coming from the ith sample.
- G stands for the *n* x *p* contribution matrix. The general term g_{ik} is referred to the mass contributions from source k to sample i. They are expressed in μg/m^3.
- F is a *p* x *m* matrix of profiles. The general term f_{kj} corresponds to the percentage of the j^{th} chemical specie coming from source k.
- E is the *n* x *m* error matrix in ng/m^3, where E is assumed to be log-normal distributed matrix.

Non-negative matrix factorizations depend on the cost function used in order to measure the dissimilarity between the initial data X and the product GF. Lee and Seung [10] has studied two of the most popular cost functions, the Frobenius norm and the Kullback-Leibler divergence.

In environmetrics, measurements are corrupted with a wide range of uncertainties. Ho [4] proposed an efficient way to incorporate a weight matrix which enables to lower the effects of large uncertainties. Up to now, few works have been done in the field of constrained NMF. The specificity of G and F enables to include for example orthogonality constraints [7] or sparseness constraints [6]. Concerning our approach, we propose to define linear equality constraints directly on some components of the profile matrix and to take them into account in a Kullback-Leibler divergence cost function. The aim of this paper is to derive general rules for constrained weighted multiplicative NMF (KL-CWNMF). This matrix factorization is applied in the field of source apportionment in order to identify airborne particulate matter sources and their relative contribution.

2. KULLBACK-LEIBLER NMF

Divergence is similar to distance in the sense that it checks positivity. Unfortunatly, triangle inequality and symmetry property are generally not satisfied.

2.1. Kullback-Leibler divergence

Kullback-Leibler is widely used in information theory and probability. It is used as an information gain, relative entropy or an information divergence [8]. At the beginning, it measures the difference between two probability distributions and it has been generalized to evaluate the difference between two non negative vectors p and q.

$D(p \parallel q) = \sum \left(p \circ log(\frac{p}{q}) - p + q\right)_i$
where i is the index of the resulting vector inside brackets and the operator $\circ$ is element wise product whereas $\frac{p}{q}$ is element wise division.
Some basic properties are recalled:
$D(p \parallel q) = 0 \Longrightarrow p = q$
$D(p \parallel q) \neq D(q \parallel p)$
It turns out from these properties that a specific divergence has to be minimized.

2.2. Classical NMF

Matrix factorization is often used to search parts based decomposition. It have been extensievely studied in the literature, pca and singular value decomposition may be some of the classical examples. NMF requires in the contrary, the positivity of all of its components. It emerged in the 90s under the name PMF, then it has been more widely used with the work of Lee and Seung [10] in 1999 where they defined two classical cost functions: The Frobenius norm and the Kullback-Leibler divergence. They were the first ones to propose the update rules called multiplicative NMF updates. Other technics may be encountered such as ALS or projected gradient [11].

2.3. The weighted KL-NMF

To our knowledge, the weighted NMF associated to KL divergence has only been developped by Ho [4]. Main results are reported below.
The data matrix factorization leads to minimize the appropriate divergence $D_W(X \parallel GF)$ with respect to GF such as:

$$D_W(X \parallel GF) = \sum_{i,j} [W \circ (X \circ \log \frac{[X]}{[GF]} - X + GF)]_{i,j} \quad (2)$$

where, $X \circ Y$ and $\frac{X}{Y}$ account for respectively component-wise product and element-wise division between two matrices.

W is the weight matrix defined by $W = \frac{1_{n\times m}}{\sum \circ \sum}$ where $\sum$ is the uncertainty matrix associated to X.

The proof below is only devoted to the search of the profile matrix F. The divergence is split into partial divergences with one column of F,W and X respectively denoted f,w and x. The divergence in (2) is the sum of partial divergences of vectors x and Gf:

C(f)= $D_w(x \parallel Gf) = \sum_i w_i(x_i \log x_i - x_i + \sum_j G_{i,j} f_j - x_i \log \sum_j G_{i,j} f_j)$ (3)

This partial divergence is approximated by the following auxiliary function which is a majorant function of the cost $C(f)$ (3):

H(f,f^k) $= \sum_i w_i[x_i \log x_i - x_i + \sum_j G_{i,j} f_j$
$- x_i \sum_j \frac{G_{i,j} f_j^k}{\sum_l G_{i,l} f_l^k} (\log G_{i,j} f_j - \log \frac{G_{i,j} f_j^k}{\sum_l G_{i,l} f_l^k})]$ (4)

The majorization-minimization theorem implies that f^{k+1} checks :

$$C(f^k) \geq \min(H(f, f^k)) = H(f^{k+1}, f^k) \geq C(f^{k+1}) \quad (5)$$

So, minimizing H instead of C ensures that the cost function C is decreasing according to iterations. This minimization with respect to f leads to solve:

$\frac{\partial H}{\partial f_j} = \sum_i w_i G_{i,j} - \frac{f_j^k}{f_j} \sum_i w_i x_i \frac{G_{i,j}}{\sum_l G_{i,l} f_l^k} = 0$

and the minimum is given by:
$f^{k+1} = \frac{f^k}{G^T W} \circ \left(G^T \frac{(X \circ W)}{(Gf^k)}\right)$
The whole update may be found by putting together all the columns of the profile matrix. The same thing may be done for the contribution matrix G [4]:

$$F = \frac{F}{G^T W} \circ \left(G^T \frac{(X \circ W)}{(GF)}\right) \quad (6)$$

$$G = \frac{G}{W F^T} \circ \left(\frac{(W \circ X)}{(GF)} F^T\right) \quad (7)$$

2.4. Solving an equality constrained NMF problem

Practically, profiles recovery is never completely blind. In some cases, information on components are available and some values can be set to zero if some species are absent from a source profile. This kind of knowledge should be included as constraints in our algorithms. Up to now and to our knowledge, no contribution supports weighted criteria with constraints.

2.4.1. Introduction of equality constraints

Linear equality constraints on the profile matrix are solely taken into account. They simply reflect the presence or absence of some compounds in a source profile. In the contrary, no knowledge on the contribution matrix G is provided.

The global formulation of equality constraints is done through two matrices Ω $(p \times m)$ and Φ $(p \times m)$:

$$F \circ \Omega - \Phi = 0 \quad (8)$$

Ω is a binary matrix reflecting the presence or absence of constraints on the source and the species :

$$\Omega_{ij} = \begin{cases} 1 & \text{if } F_{ij} \text{ has to be set.} \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

Φ is the matrix of set values. Some profiles may be set to zero or to a positive value.

Let f_i^k be the i^{th} column of the F matrix and φ_i be the i^{th} column of the Φ matrix. Let also M_i $(l_i \times p)$ be the constraint matrix issuing from the i^{th} column of the Ω matrix containing l_i constraints. It checks the following relation :

$$M_i f_i - \delta_i = 0 \quad (10)$$

where δ_i is the extraction of set values issued from Φ : $M_i \varphi_i - \delta_i = 0$
and φ_i is the i^{th} column of the Φ matrix.
For example, a 5 sources and two constraints case where the 2^{nd} and the 4^{th} component are set to values :

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} f_i \end{bmatrix} = \begin{bmatrix} 80 \\ 30 \end{bmatrix} \quad (11)$$

Let $Span(\Gamma_i)$ be the supplementary space to the rows of M_i such that rank $\begin{bmatrix} M_i \\ \Gamma_i^T \end{bmatrix} = p$.
Γ_i is $(p \times (p - l_i))$, it checks the following normalization relation :

$$\begin{cases} M_i \Gamma_i = & 0_{l_i \times (p-l_i)} \\ \Gamma_i^T \Gamma_i = & I_{(p-l_i)\times(p-l_i)} \end{cases} \quad (12)$$

Let us consider the vectorial form of the model restricting to the search of one column of the profile matrix. Equations (10) and (12) show that a column of the profile matrix may be expressed :

$$f_i = \varphi_i + \Gamma_i \theta_i \quad (13)$$

where θ_i is $(p - l_i) \times 1$ vector of free parameters.
This is the general form used for solving constrained problems in next sections. We shall use also the following formalism ΔF for matrix or Δf_i for the associated column vector :

$$\Delta f_i = f_i - \varphi_i \qquad or \qquad \Delta F = F - \Phi \quad (14)$$

2.4.2. KL-CWNMF formalism

We try in this section to extend the approach to the case of Constrained weighted NMF. In the case of Frobenius norm, the work has already been investigated by Delmaire [1]. Usually, the Kullback-Leibler divergence is related to the data matrix and the unknown matrix factorization. The second part of the divergence is generally devoted to unknown variables.

$$\min_F D_W(X||GF) \Rightarrow X \approx GF$$

Equation (14) enables to write an approximated formulation of $X - G\Phi$ so that another minimization may be investigated:

$$\min_{\Delta F} D_W(X - G\Phi||G\Delta F) \Rightarrow X - G\Phi \approx G\Delta F$$

It may be seen that both minimizations lead to an approximated factorization. We chose to develop the second one under the principle that the unknowns remains in the second part of the divergence. We focus on a column of the data since the divergence $D_W(X - G\Phi||G\Delta F)$ may be split into partial divergences.
For sake of simplicity, the column index i is dropped in next equations.

Let x be a column of the data matrix, φ one column of the Φ matrix, $\triangle f$ one column of the $\triangle F$ matrix where $\triangle F = F - \Phi$.

Let $U = G\Gamma$ and
$A_i = (x_i - (G\varphi)_i)\log(x_i - (G\varphi)_i)$
$B_i = x_i - (G\varphi)_i$ $\qquad C_i = \sum_j U_{i,j}\theta_j$
and $D_i = (x_i - (G\varphi)_i)\log(\sum_j U_{i,j}\theta_j)$
The partial divergence of one column vector may be expressed :

$$D_w(x - G\varphi \parallel U\theta) = \sum_i w_i \left[A_i - B_i + C_i - D_i\right] \quad (15)$$

Let:
$E_i = (x_i - (G\varphi)_i)\sum_j \frac{U_{i,j}\theta_j^k}{\sum_l U_{i,l}\theta_l^k}(\log U_{i,j}\theta_j - \log \frac{U_{i,j}\theta_j^k}{\sum_l U_{i,l}\theta_l^k})$

Using the property that $(-\log X)$ is convex, the majorization-minimization theorem enables to define a majorant function that checks the property (5):

$$H(\theta, \theta^k) = \sum_i w_i \left[A_i - B_i + C_i - E_i\right] \quad (16)$$

Free parameters are gathered in the θ vector so that the minimization has to be made with respect to θ. Derivating with respect to θ_j leads to:
$\frac{\partial H}{\partial \theta_j} = \sum_i w_i U_{i,j} - \frac{\theta_j^k}{\theta_j}\sum_i w_i (x_i - (G\varphi)_i)\frac{U_{i,j}}{\sum_l U_{i,l}\theta_l^k} = 0$
Putting together the unknown parameters:

$$\frac{\theta_j^k}{\theta_j^{k+1}} = \frac{\sum_i w_i U_{i,j}}{\sum_i w_i (x_i - (G\varphi)_i)\frac{U_{i,j}}{\sum_l U_{i,l}\theta_l^k}}$$

Rearranging into a vectorial form:

$$\theta^{k+1} = \frac{\theta^k}{U^T w} \circ \left(U^T \frac{(x - G\varphi)\circ w)}{(U\theta^k)}\right)$$

Computing $\triangle f$ for the i^{th} column of the profile matrix, denoted from now Δf_i, leads to multiply each terme by Γ_i (the column index i is from now taken into account):

$$\triangle f_i^{k+1} = \frac{\triangle f_i^k}{\Gamma_i \Gamma_i^T G^T w_i} \circ \left(\Gamma_i \Gamma_i^T G^T \frac{(x - G\varphi)\circ w)}{(G\triangle f_i^k)}\right) \quad (17)$$

Given that $\Gamma_i \Gamma_i^T = diag(1_{p\times 1} - \omega_i)$, with ω_i the i^{th} column of the Ω matrix, it may be noticed that this operator only selects active components among the profile vector in (6). The previous expression may be summarized into matrix formulation :

$$\triangle F^{k+1} \leftarrow \frac{\triangle F^k}{G^T W} \circ (1-\Omega) \circ G^T \frac{W \circ (X - G\Phi)}{G \triangle F^k} \tag{18}$$

and the following update of the profile matrix F is:

$$F - \Phi \leftarrow \frac{F - \Phi}{G^T W} \circ (1-\Omega) \circ \left[G^T \frac{W \circ (X - G\Phi)}{G(F - \Phi)} \right] \tag{19}$$

In order to prevent divisions by 0, the previous expression may be modified according to the following one :

$$F - \Phi \leftarrow \frac{F - \Phi}{G^T W} \circ (1-\Omega) \circ \left[G^T \frac{W \circ (X - G\Phi)}{G(F - \Phi + \epsilon\Omega)} \right] \tag{20}$$

It may be noticed that $(X - G\Phi)$ are considered as equivalent data, as a result, it has to be non negative, otherwise the negative component has to be replaced by 0. Practically, the case is very rarely encountered. It turns out also, that if the initial value $\triangle F^0$ is positive, then $\triangle F^k$ is always non negative.

2.5. Summary of the algorithm

Non negative matrix factorization is an iterative algorithm based on the successive estimation of the contribution matrix G with a fixed profile matrix F and then the estimation of F with a fixed G. The iterative procedure is outlined below:

• While the stoppping rule is not checked
{
Check for positivity of $(X - G\Phi)$ and project it if necessary
Search for F at constant G
Search for G at constant F
Normalizing F and G }

At the end of every iteration, a normalization step is applied to the rows of the profile matrix F. The contibution matrix G is then updated to keep unchanged the product GF.

3. INDUSTRIAL SOURCES IDENTIFICATION BY NON NEGATIVE MATRIX FACTORISATION

The application proposed in this work concerns a series of airborne particulate matter sampled from a coastal city in Northern France close to an integrated steelworks (Figure1).

Fig. 1. Location of the area under study: Dunkerque - North of France

A number of 92 valid samples corresponding to airborne particles with size lower than 10μm were considered. Sampling was performed under various meteorological conditions so that particles could be collected under the influence of several emission sources. The chemical composition of samples was determined focusing on metal elements (Al, Ca, Cr, Cu, Fe, K, Mg, Mn, Na, Ni, Pb, Sn, Ti, V, Zn) and ionic species (Cl^-, NO_3^-, SO_4^{2-}, NH_4^+). A previous work dedicated to the identification of potential sources [1] enabled to identify source profiles contributing mainly to the atmospheric particulate background: sea salts, aged sea-salts, secondary inorganic aerosols and crustal particles. Their respective profiles were in good agreement with results from literature [2] [12] [5]. 5 industrial expected sources are considered: blast furnaces, steel slag, ores sintering plant, sintering chimney and ferromanganese plant.

To sum up, X is made with 92 samples and 19 chemical species, G is a 92 samples and 9 sources matrix and F is a 9 sources and 19 chemical species matrix.
Moreover the $\sum$ matrix involves uncertainties provided by chemical analysis methods.
Our first tests concern the Kullback-Leibler weighted NMF (KL-WNMF) without equality constraints.

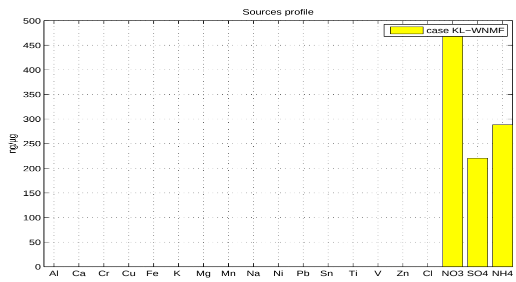

Fig. 2. Secondary inorganic aerosols

The KL-WNMF applied to our data allows us to recognize some sources such as sea salts, aged sea salts and the secondary inorganic aerosols. The latter are typically identified by the presence of NO_3^- SO_4^{2-} and NH_4^+ (Figure 2).

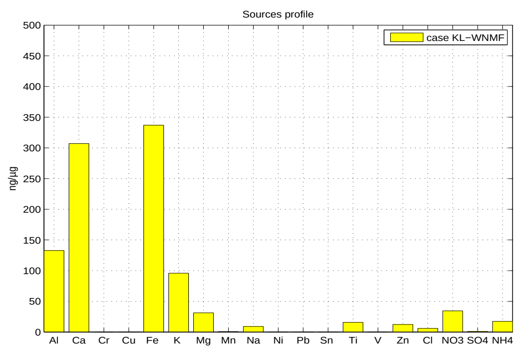

Fig. 3. Intermediate profile between crustal source and blast furnaces-steel plant source.

However, results of the algorithm are ambiguous in the case of extra profiles coming from industrial emissions. According to experts, some profiles do not show the source separation. For instance, Figure 3 shows a profile in which the separation of crustal particles (natural emissions) and blast furnaces-steel plant source (industrial emissions) could not be obtained. The contribution of the industrial source is evidenced in this case by the amount of Fe, Ca and Zn much higher than the one expected for the crustal source. The same ambiguity is noticed with other profiles.

The use of constraints appears as an interesting way to get better fits for profiles compared to characteristics known for emission source samples. The next step of our tests is to add some equality constraints to our algorithm. As previously mentioned, the four sources contributing the particulate background are in agreement with literature results so that we have considered that their profiles could be considered as known.

The Ω matrix specifies which elements are fixed to some constraints. In this case, corresponding cells are set to 1 (table 1).

Table 1. Ω matrix

Al	Ca	Cr	Cu	Fe	K	Mg	Mn	Na	Ni	Pb	Sn	Ti	V	Zn	Cl^-	NO_3^-	SO_4^{2-}	NH_4^+
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	1	0	1	1	1	1	0	1	1	0	1
0	0	0	1	0	1	0	0	1	0	1	0	0	0	1	1	1	0	1
0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	1	1	1	1
0	0	0	0	0	0	1	0	0	1	0	1	0	1	0	0	1	0	0
0	0	0	0	0	1	1	0	1	1	0	1	0	1	0	1	1	1	1

The zeros in the Ω matrix corespond to elements for which estimates are computed by the algorithm.

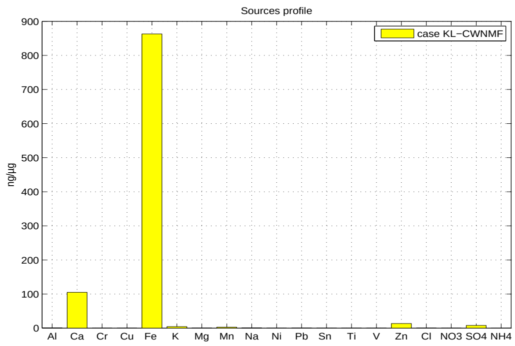

Fig. 4. KL-CWNMF: Blast furnaces - steel plant profile

According to chemical composition data available for reference samples of industrial particulate emissions [9] [3], each source may be recognized from the relative abundance of elements and ions appearing in profiles. The first industrial source profile given in figure 4 is characterized by a high amount of iron and in lower proportions calcium, zinc and manganese. Such features are encountered in the case of particles emitted by blast furnaces or by the steel plant, these two sources showing similar profiles.

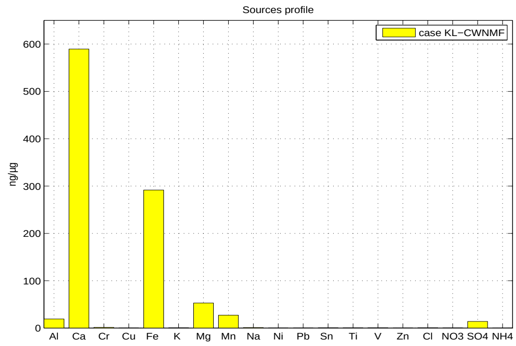

Fig. 5. KL-CWNMF: Steel slag profile

The profile of figure 5 is characterized by the presence of a large amount of calcium and the presence of iron, aluminium, manganese and magnesium. This profile can be ascribed to particulate matter from steel slags emitted in the form of fugitive emissions.

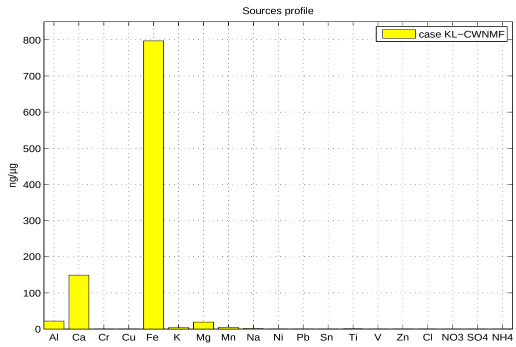

Fig. 6. KL-CWNMF: Ores sintering profile

The profile described on figure 6 shows a large amount of Fe, Ca, Al and Mg and small quantities of Mn, K and Na. This can be attributed to fugitive emissions due to handling of the sintered ores inside the steelworks site.

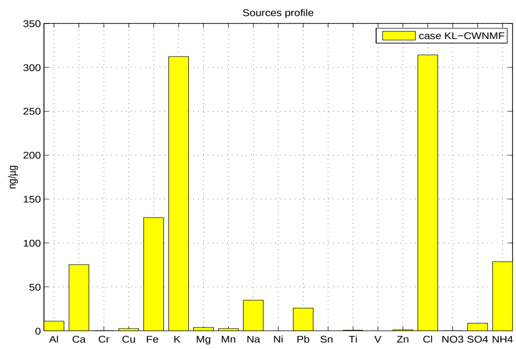

Fig. 7. KL-CWNMF: Sintering chimney profile

The profile of figure 7 corresponds to the sintering chimney point source. It is characterized by the presence of Cl^-, K, Fe, Ca, NH_4^+ and Pb. From figures 6 and 7, it can be noticed that compositions of particulate from the ores sintering either as fugitive emissions or point source emissions unit are different. The separation of both sources appears to be clearly obtained.

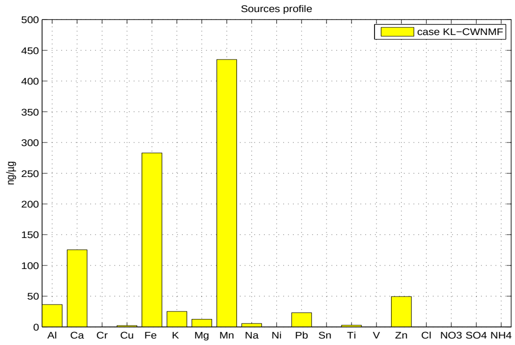

Fig. 8. KL-CWNMF: Ferromanganese alloys profile

Moreover, the ferromanganese alloys emissions may be easily recognized (figure 8) with the presence of the Mn, Fe, Ca, Zn, Al and Pb.

Finally, identification without equality constraints provides ambiguous results due to mixture of multiple sources while constrained NMF algorithm enables to identify sources correctly.

4. CONCLUSION

This article is devoted to the introduction of simple linear equality constraints into the NMF algorithm. Particularly, the focus is only made on the weighted KL divergence. Then, general mutiplicative rules are derived (20) which enable to directly update the profile matrix and the contribution matrix (KL-CWNMF). This technic is used in the frame of particulate matter source identification. Tests are made on the available data with and without constraints. It turns out clearly that constraints enable a better identification of source profiles which are in good agreement with chemical literature.

5. REFERENCES

[1] Delmaire G., Roussel G., Hleis D. and Courcot D. Factorisation matricielle non négative sous contraintes. Application à l'identification de sources industrielles. *CIFA*, 2-4 juin, Nancy France, 2010.

[2] Evans M. C. S. W. Campbell and V. Bhethanabotla and N. D. Poor. Effect of sea salt calcium carbonate interactions with nitric acid on the direct dry deposition of nitrogen to Tampa Bay, Florida. *Atmospheric Environment*. vol. 38, n 29, pp 4847-4858, 2004.

[3] Hleis D. Evaluation de la contribution d'émissions sidrurgiques à la teneur en particules en suspension dans l'atmosphre à une échelle locale, Thesis, Université du Littoral Côte d'Opale, 2010.

[4] Ho N-D. Non-negative matrix factorization algorithms and applications. Thesis from Ecole Polytechnique de Louvain, 2008

[5] Hopke P.K. Theory and Application of Atmospheric Source Apportionment, in Developments in Environmental Science, Volume 9, Edited by A.H. Legge, Elsevier, Ltd., Burlington, MA, pp. 1-33, 2009.

[6] Hoyer P. Non-negative matrix factorization with sparseness constraint *Journal of Machine Learning Research*. 1457-1469, November, 2004.

[7] Jiho Y. and Seungjin C. Nonnegative Matrix Factorization with Orthogonality Constraints *Journal of computing science and engineering*. vol. 4, no. 2, pp.97-109, June, 2010.

[8] Kullback S. Information Theory and Statistics. *New York .Courier Dover Publications*, 1997.

[9] Laversin H. Traceurs et formes chimiques du fer dans les particules émises dans l'atmosphère depuis un site sidérurgique: Etude spectroscopique et caractérisation de composés de référence et de particules collectées dans l'environnement. *Thesis from Université du Littoral Côte d'Opale*, 2006.

[10] Lee Daniel D. and Sebastian Seung H. Learning the parts of objects by non negative matrix factorization. *Nature*. vol. 401, n 6755, pp. 788-791, 1999.

[11] Lin. Chih Jen. Projected Gradients Methods for Non-Negative Matrix Factorization. *Neural Computation*. vol. 19, n 10, pp. 2756-2779, 2007.

[12] Ten-Harkel M. J. The effects of particle-size distribution and chloride depletion of sea-salt aerosols on estimating atmospheric deposition at a coastal site. *Atmospheric Environment*. vol. 31, n 3, pp 417-427, 1997.

[13] Viana M., Kuhlbusch T. A. J., Querol X. and Alastuey A.. Source apportionment of particulate matter in Europe: a review of methods and results. *Journal of Aerosol Science*. vol 39, n 10 , pp 827-849, 2008.

Acknowledgments

ArcelorMittal Atlantique et Lorraine Steel producer is gratefully acknowledged for financial supports.