Authors' final version of a paper published in "Applied Numerical Mathematics"

Non-negative matrix factorization under equality constraints - a study of industrial source identification.

A. Limem[a,b], G. Delmaire[a,b,*], M. Puigt[a,b], G. Roussel[a,b], D. Courcot[a,c]

[a]*Université Lille Nord de France, 59655 Villeneuve d'Ascq, France.*
[b]*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.
[c]*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

This work is devoted to the factorization of an observation matrix into additive factors, respectively a contribution matrix G and a profile matrix F which enable to identify many pollution sources. The search for G and F is achieved through Non-negative Matrix Factorization techniques which alternatively look for the best updates on G and F.

These methods are sensitive to noise and initialization, and—as for any blind source separation method—give results up to a scaling factor and a permutation. A Weighted Non-negative Matrix Factorization extension has also been proposed in the literature, so that different standard deviations of the data matrix components are taken into account. However, some estimated profile components may be inconsistent with practical experience. To prevent this issue, we propose an informed Non-negative Matrix Factorization, where some components of the profile matrix are set to zero or to a constant positive value. A special parametrization of the profile matrix is developed

*Corresponding author

Email addresses: `abdelhakim.limem@lisic.univ-littoral.fr` (A. Limem),
`gilles.delmaire@lisic.univ-littoral.fr` (G. Delmaire),
`matthieu.puigt@lisic.univ-littoral.fr` (M. Puigt),
`gilles.roussel@lisic.univ-littoral.fr` (G. Roussel),
`dominique.courcot@univ-littoral.fr` (D. Courcot)

in order to freeze some profile components and to let free the other ones.

The problem amounts to solve a family of quadratic sub-problems. A Maximization Minimization strategy leads to some global analytical expressions of both factors.

These techniques are used to estimate source contributions of airborne particles from both industrial and natural influences. The relevance of the proposed approach is shown on a real dataset.

1. Introduction

Approximate matrix factorization techniques are of great importance in several areas, such as pattern recognition, antenna array processing, or environmental data processing. Non-negative matrix Factorization (NMF) aims to factorize a non-negative matrix as the product between two matrices, such that each entry of the latter matrices is null or positive. It emerged under the name Positive Matrix Factorization (PMF) with Paatero and Tapper's contributions [29, 30]. It then appeared in the field of signal and image processing with the work by Lee and Seung [24]. Some of the most well known approaches have been proposed by Cichocki and Zdunek [7] and Lin [26], and are based on a projected gradient method. Recently, some weighted versions of NMF were investigated and took into account specific standard deviations associated to each data point. Inconsistencies in practical results lead to investigate additional assumptions such as sparseness constraints [20, 2], fixed row and column sums constraints [18], or orthogonality constraints [35, 5].

Over the last decade, matrix factorization has been extensively investigated in the field of atmospheric sciences, and particularly for receptor models with the aim to re-construct the impacts of emissions from different pollutant sources, e.g., airborne particles [19]. Such source apportionment studies enable to identify the main emission sources and their relative contribution at different monitoring sites. However, Viana *et al.* [33] reported that, depending on the applied NMF method, these techniques may provide highly different solutions. Moreover—and to the best of our knowledge—NMF tools fail in discriminating some natural sources (i.e., crustal matter from local or regional re-suspension and road dust), as well as sources containing sulfate and nitrate components. Moreover, the separation of mineral

3

particles emitted from both industrial dust emissions and natural sources appears as a difficult task [9]. In this paper, we propose to take into account some partial *a priori* knowledge on some source profiles in order to improve the estimation to the unknown components.

The remainder of the paper is organized as follows. In Section 2, we present the concepts of some state-of-the-art NMF methods. Section 3 introduces our proposed approach while its performance is investigated in Section 4. We conclude and discuss future directions of this work in Section 5.

2. A short review of Non-negative Matrix Factorization

NMF aims to estimate two non-negative matrices whose product approximates the observed one. The non-negativity constraints make the decomposition purely additive whereas other ones like Principal Components Analysis (PCA) and Independent Components Analysis (ICA) can provide negative components. Informed NMF is located between NMF methods—which only assume non-negativity of the matrices—on one hand, and on the other hand, regression models which assume that one matrix is exactly known. In this paper, we introduce an informed NMF method, since matrices may be partially known.

2.1. Receptor modelling

This model is very generic since it is often used in signal and image processing. In environmental field, the factorization model (1) below is best known as receptor modelling and links the data matrix and the active sources. The factorization enables to approximate the $n \times m$ data matrix X by the product of two matrices,

$$X \approx G \cdot F, \tag{1}$$

where

- X is the $n \times m$ data matrix, where n is the number of samples and m is the number of species. In environmetrics, x_{ij}—the $(i,j)^{th}$ element of X—accounts for the concentration, expressed in ng/m^3, of the jth chemical species coming from the ith sample. A row gathers concentrations of all registered species for the current sample.

- G stands for the $n \times p$ contribution matrix, where p is the pre specified number of sources and n is the number of samples. In environmetrics

4

and chemometrics, the $(i, k)^{th}$ element of G—denoted g_{ik}—is referred to as the massic contribution from Source k to Sample i, expressed in $\mu\text{g}/\text{m}^3$.

- F is a $p \times m$ matrix of profiles. Its $(k, j)^{th}$ term f_{kj} stands for a mass ratio (in $\text{ng}/\mu\text{g}$) corresponding to a percentage of the jth species with respect to the whole mass of the Source k.

The number p of sources is usually chosen such that $np + pm \ll nm$. The *a priori* choice for the appropriate value of p mostly depends on the nature of the investigated data [16] and on the expert's knowledge of the number of sources. In the absence of such information, p can be estimated in a pre-processing stage by applying some information theory techniques [11] such as AIC or BIC or extended versions.

2.2. Non-negative matrix factorization

Non-negative matrix factorization has received a lot of attention since Lee and Seung [24] published their popular multiplicative algorithm. In fact, NMF looks for an approximate factorization of the data matrix according to receptor modelling (1) under the additional non-negativity constraint. The basic NMF solves

$$\min_{G,F \geq 0} \sum_{i=1}^{n} \sum_{j=1}^{m} \left((x_{ij} - (GF)_{ij}) \right)^2 = \min_{G,F \geq 0} \| X - G \cdot F \|_{\mathfrak{f}}^2 , \qquad (2)$$

where $G, F \geq 0$ means that all components of G, F are non-negative and $\| . \|_{\mathfrak{f}}$ stands for the Frobenius norm[1]. The problem in (2) is convex with respect to G and F separately, but not for both at the same time. As a consequence, most algorithms are only able to yield local minima. Classical NMF algorithms thus alternatively search for F (respectively G) when G is fixed (respectively F). As an alternative, Kim *et al.* proposed a general framework [22], called Block Coordinate Descent, which includes several methods described hereafter. Different iterative approaches to solve (2) were proposed in the literature:

---

[1]Please note that other cost functions are also widely used, e.g., Kullback divergence [24] or flexible parametric divergences such as generalized alpha-beta divergences [6].

- Most commonly used ones are the multiplicative updates rules. Lee and Seung [24] mentioned that this update is a good trade-off between speed of convergence and ease of implementation. The update of F (respectively G) is made through a multiplicative update coefficient which depends on the data matrix X and the contribution matrix G (respectively the profile matrix F):

$$F \leftarrow F \circ \frac{(G^T X)}{(G^T G F)}, \qquad G \leftarrow G \circ \frac{(X F^T)}{(G F F^T)}, \qquad (3)$$

where $X \circ Y$ and $\frac{X}{Y}$ respectively account for component-wise product and element-wise division between two matrices. According to the update rule (3), the Frobenius norm is not increasing and becomes constant if a limit point is reached.

A limit point (G, F) is said to be stationary if it checks the Karush-Kuhn Tucker (KKT) conditions, i.e.,

$$
\begin{array}{ll}
F \geq 0, & G \geq 0, \\
G^T \cdot (G \cdot F - X) \geq 0, & (G \cdot F - X) \cdot F^T \geq 0, \\
F \circ \big(G^T \cdot (G \cdot F - X) \big) = 0, & G \circ \big(((G \cdot F - X)) \cdot F^T \big) = 0.
\end{array}
$$

Stationarity is only a necessary condition of a local minimum. Moreover, Lin [25] reported that some limit points of (3) which are not stationary may exist, especially if some components of F and G are initialized to zero.

- As an alternative, Alternating Non-negative Least Squares (ANLS) consists of alternatively estimating G (respectively F) by setting F (respectively G). The mathematic formulation of the problem at iteration $r + 1$ may be divided into two sub-problems depending on the results obtained at iteration r [21]:

$$
\begin{aligned}
F^{r+1} &= \arg \min_{F \geq 0} \left\| X - G^r F \right\|_{\mathfrak{f}}^2, \\
G^{r+1} &= \arg \min_{G \geq 0} \left\| X - G F^{r+1} \right\|_{\mathfrak{f}}^2.
\end{aligned}
\qquad (4)
$$

Each step of (4) may be seen as a set of independent non-negative least squares problems involving one column or one row of the data matrix.

Let x_i be the ith column of the matrix X and f_j^{r+1} be the jth column of the matrix F^{r+1}. Searching for F amounts to find separately each column of F,

$$f_j^{r+1} = \arg\min_{f_j \geq 0} \|x_i - G^r f_j\|_{\mathfrak{f}}^2 \quad \forall j \in \{1...m\} \quad \forall i \in \{1...m\}.$$

The resolution of this problem is generally more time consuming at each iteration than the multiplicative updates but the number of iterations is expected to be lower. This method may also be considered as a special case of the Block Coordinate Descent framework in the case of two blocks [22]. Grippo and Sciandrone [15] state that every limit point obtained from (4) is a stationary point.

- Lin [26] proposed to develop projected gradient-based techniques for bound constrained optimization. The iterative update of the matrix is done through a projected gradient approach on the feasible region:

$$F^{r+1} \leftarrow P\left[F^r - \alpha^r \quad \nabla_{F^r}\left(\|X - G^r F\|_{\mathfrak{f}}^2\right)\right],$$

where the projection P enables to map the candidate back into the bounded region if it lies outside:

$$\forall (i,j) \in \{1...n\} \times \{1...m\}, \quad (P(X))_{i,j} = \begin{cases} X_{i,j} & \text{if } m_{i,j} \leq X_{i,j} \leq M_{i,j}, \\ M_{i,j} & \text{if } X_{i,j} \geq M_{i,j}, \\ m_{i,j} & \text{if } X_{i,j} \leq m_{i,j}. \end{cases}$$

Many articles—such as [3] for example—are devoted to the research of a step α^r which checks a sufficiently fast decreasing for each step. This stage is also the most time consuming one.

Even if most work—including the above methods and our contribution—focus on approximated solutions of the factorization of X, please note the existence of recent investigations on *exact* factorization. For example, Vavasis [32] showed the NP-hardness of computing such a factorization. Recently, Arora *et al.* proposed a near-separable NMF algorithm which provably runs in polynomial time under some additional assumptions on the input matrix X [1]. Near-separability has been furtherly explored in, e.g., [4, 13].

2.3. Weighting the Non-negative Matrix Factorization

Recently, some authors have proposed an uncertainty weighted version of NMF, expressed either as a sequence of columns updates [8] or as a whole update [17]. The function to be minimized derives from the classical NMF cost function and involves a weight matrix, i.e.,

$$\min_{G,F\geq 0} \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\frac{(x_{ij} - (GF)_{ij})}{\sigma_{ij}} \right)^2,$$

where σ_{ij} stands for the standard deviation obtained from the measurement of the ith sample and the jth species, gathered in the uncertainty matrix $\Sigma = \{\sigma_{ij}\}$. A whole weight matrix W may be expressed: $W = \frac{1_{n\times m}}{\Sigma \circ \Sigma}$, where each element of the matrix $1_{n\times m}$ is equal to 1.

Ho [17] derived the Karush-Kuhn Tucker conditions in the weighted case, which are outlined as

$$F \geq 0, \qquad\qquad\qquad\qquad G \geq 0,$$
$$G^T \cdot (W \circ (G \cdot F - X)) \geq 0, \qquad (W \circ (G \cdot F - X)) \cdot F^T \geq 0,$$
$$F \circ \left(G^T \cdot (W \circ (G \cdot F - X))\right) = 0, \qquad G \circ \left((W \circ (G \cdot F - X)) \cdot F^T\right) = 0.$$

Ho also introduced a unified expression of an iterative solution of this criterion using the maximization minimization theorem [17]. The main advantage lies in the whole iterative update of each iteration, i.e.,

$$F \leftarrow F \circ \frac{G^T(W \circ X)}{G^T(W \circ (GF))}, \qquad G \leftarrow G \circ \frac{(W \circ X)F^T}{(W \circ (GF))F^T}, \qquad (5)$$

In order to prevent division by 0, a small quantity ε is added to the denominator of the relation (5), i.e.,

$$F \leftarrow F \circ \frac{G^T(W \circ X)}{G^T(W \circ (GF)) + \varepsilon 1_{p\times m}}, \qquad G \leftarrow G \circ \frac{(W \circ X)F^T}{(W \circ (GF))F^T + \varepsilon 1_{n\times p}}.$$

Using weights is particularly important in the case of physical measurements, where data are obtained with radically different uncertainties. Note that the update rules in (5) may be viewed as an extension of the classical multiplicative update described in (3) to weighted data. Indeed, by setting W to the matrix $1_{n\times m}$, the updates rules in (3) are exactly those in (5). From this remark, weighted NMF (WNMF) behaves as the classical NMF for convergence aspects. In fact, some non-stationary limit points may exist with

weighted multiplicative updates, especially in the case when some entries of F and G are initialized to zero [25].

Other criteria may also be investigated, notably the minimization of the weighted Kullback-Leibler divergence. For this case, the expression of weighted multiplicative updates is available in [17]. To conclude, the main interest for this kind of algorithm lies in the direct update of both matrices and its ease of implementation.

3. Solving a constrained problem

Practically, the profiles recovery is never completely blind. In some cases, *a priori* 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 the best of our knowledge—no state-of-the-art method supports weighted cost functions with constraints.

3.1. Introduction of equality constraints

In many cases, some known components of the source profile may be provided by experts. The aim of this section is to find a parametrization that naturally integrates this knowledge. Constraints may be formulated as follows. Let Ω be a $p \times m$ binary matrix which specifies the presence or the absence of constraints on each element of the profile matrix F,

$$\forall (i,j) \in \{1...p\} \times \{1...m\}, \quad \Omega_{ij} = \begin{cases} 1 & \text{if } F_{ij} \text{ has to be set,} \\ 0 & \text{otherwise.} \end{cases}$$

Φ is the $p \times m$ matrix of set values. By construction, Φ_{ij} is equal to zero when $\Omega_{ij} = 0$. It may thus be easily checked that

$$\Phi \circ \Omega = \Phi \ . \tag{6}$$

Using the definition of matrices (Φ, Ω), known components of F may easily be set to zero or to other positive values according to the following relationship:

$$F \circ \Omega - \Phi = 0 \ . \tag{7}$$

Let:

- f_i be the ith column of the F matrix,

- φ_i be the ith column of the Φ matrix,

- M_i be the $(l_i \times p)$ constraint matrix made up with the ith column of the matrix Ω (i.e., containing l_i constraints),

- δ_i be the vector of the constraint values extracted from Φ.

Constraints may be expressed for the ith column vector of the profile matrix as

$$M_i f_i = M_i \varphi_i = \delta_i, \qquad \forall i \in \{1...m\}. \tag{8}$$

As an example, let us consider a case with, say $p = 5$ sources and $l_i = 2$ constraints for all i. The latter are here assumed to be set on the second and the fourth position, and to be respectively equal to, say 80 and 30. The previous matrices then read:

$$M_i = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad \begin{matrix} \varphi_i^T = \begin{bmatrix} 0 & 80 & 0 & 30 & 0 \end{bmatrix}, \\ \delta_i^T = \begin{bmatrix} 80 & 30 \end{bmatrix}, \end{matrix} \qquad \forall i \in \{1...m\}.$$

Let Γ_i be the $(p \times (p - l_i))$ matrix whose columns span the supplementary space to the rows of M_i such that $\operatorname{rank} \begin{bmatrix} M_i \\ \Gamma_i^T \end{bmatrix} = p$. Γ_i checks the following normalization relation

$$\begin{cases} M_i \Gamma_i = & 0_{l_i \times (p-l_i)}, \\ \Gamma_i^T \Gamma_i = & I_{(p-l_i) \times (p-l_i)}, \end{cases} \qquad \forall i \in \{1...m\},$$

where $0_{l_i \times (p-l_i)}$ and $I_{(p-l_i) \times (p-l_i)}$ are respectively the null matrix and the identity matrix. By noticing in (8) that $(f_i - \varphi_i)$ has to lie in the space spanned by the columns of Γ_i, it turns out that

$$\Delta f_i \triangleq f_i - \varphi_i = \Gamma_i \theta_i \qquad \forall i \in \{1...m\}, \tag{9}$$

where θ_i is the $((p - l_i) \times 1)$ vector of free parameters. The expression of one column of the profile matrix may then be deduced

$$f_i = \varphi_i + \Gamma_i \theta_i \qquad \forall i \in \{1...m\}. \tag{10}$$

The expression in (9) reads in matrix form,

$$\Delta F = F - \Phi, \tag{11}$$

10

where ΔF is the matrix gathering each column Δf_i, $i = 1...m$. It may be easily stated by using (6) and (7) that

$$\Delta F \circ \Omega = 0. \tag{12}$$

From (11), we can express F with respect to ΔF and Φ as,

$$F = \Delta F + \Phi. \tag{13}$$

By combining the relationship (13) with the fact that $1 = \Omega + (1 - \Omega)$, we obtain

$$F = \Omega \circ \Phi + \Omega \circ \Delta F + (1 - \Omega) \circ \Delta F + (1 - \Omega) \circ \Phi. \tag{14}$$

Introducing the above properties (6) and (12) into (14) leads to the simplified expression of the profile matrix

$$F = \Omega \circ \Phi + (1 - \Omega) \circ \Delta F. \tag{15}$$

Since $(1 - \Omega)$, Ω, Φ, ΔF are non-negative matrices, the profile matrix verifies $F \geq \Phi$, according to Relation (15). This matrix form is also equivalent to its vectorial expression (10); both may be used in the following sections.

3.2. Constrained problem formulation

The aim of this section is to find new update rules which verify the defined above constraints. Constraints defined through (7) have to be incorporated and—similarly to (5)—multiplicative rules are investigated. The whole minimization problem may be formulated as

$$\min_{G,F} \text{Tr}\left((X - G \cdot F)^T ((X - G \cdot F) \circ W) \right)$$
$$\text{s. t.} \quad G \geq 0, \quad F \geq \Phi, \quad F = \Omega \circ \Phi + (1 - \Omega) \circ \Delta F. \tag{16}$$

The proof of the KKT conditions are detailed in Appendix A for the case $\Phi = 0$. The extension of these conditions should be easily derived when Φ is different from the null matrix. For the sake of clarity, let us define

$$\overline{\Omega} \triangleq (1 - \Omega) \text{ and } \widetilde{F} \triangleq \overline{\Omega} \circ \triangle F + \Phi.$$

The KKT conditions (see Appendix A for details) for the constrained weighted problem remain unchanged, i.e.,

$$\begin{aligned}
&\triangle F \geq 0, & &G \geq 0, \\
&\overline{\Omega} \circ \left(G^T((G\widetilde{F} - X) \circ W) \right) \geq 0, & &(W \circ (G\widetilde{F} - X))\widetilde{F}^T \geq 0, \\
&\triangle F \circ \overline{\Omega} \circ \left(G^T((G\widetilde{F} - X) \circ W) \right) = 0, & &G \circ \left((W \circ (G\widetilde{F} - X))\widetilde{F}^T \right) = 0.
\end{aligned} \tag{17}$$

The profile matrix (respectively the contribution matrix) should be split into columns (respectively rows). The problem (16) is only solved for the profile matrix since the same technique may be applied to the contribution matrix. To solve (16), the criterion may be split into independent parts which are only related to one column of the profile matrix. The problem may be formulated as a collection of quadratic optimization sub-problems,

$$
\begin{aligned}
J\left(f_i\right) = & \quad \left(x_i - G f_i\right)^T D_{w_i}\left(x_i - G f_i\right) \\
\text{s. t.} & \quad M_i f_i = \delta_i \quad \text{and} \quad f_i \geq \varphi_i, \quad \forall i \in \{1...m\},
\end{aligned}
\tag{18}
$$

where x_i stands for the ith column of the X matrix. Remember that the matrix D_{w_i} is defined by $D_{w_i} \triangleq \mathrm{diag}(w_i)$, where w_i accounts for the ith column of the weight matrix W.

The sub-problem (18) shall be investigated for two different cases.

3.2.1. The case $\delta_i = 0$

Here, we are interested in the solution of the special case when $\delta_i = 0$ which leads to $\varphi_i = 0$. Substituting $\varphi_i = 0$ into (10) yields a simplified expression, which tells that f_i is a linear combination of column vectors from Γ_i:

$$
f_i = \quad \Gamma_i \theta_i \quad \forall i \in \{1...m\},
\tag{19}
$$

where θ_i $((p - l_i) \times 1)$ stands for the free parameters vector extracted from f_i.

By substituting (19) into (18), it turns out to be a weighted non-negative least square function, depending on the free parameters vector θ_i:

$$
\begin{aligned}
K\left(\theta_i\right) = & \left(x_i - (G\Gamma_i)\theta_i\right)^T D_{w_i}\left(x_i - (G\Gamma_i)\theta_i\right) \\
\text{s. t.} & \quad \theta_i \geq 0 \quad \forall i \in \{1\ldots m\}.
\end{aligned}
\tag{20}
$$

It appears to be a WNMF (as detailed in [17]) applied to the search of the vector θ_i and to a new contribution matrix $G\Gamma_i$. Ho [17] expressed the vector at iteration $(k+1)$ as a function of the vector at iteration k by using a maximization minimization strategy. As a consequence, we can respectively replace f_i and G in his contribution by θ_i and $G\Gamma_i$. The solution may be obtained also by splitting the first part of (5) into vectorial form, and we obtain the following update rule for θ_i:

$$
\theta_i^{k+1} \leftarrow \theta_i^k \circ \frac{\Gamma_i^T G^T (D_{w_i} x_i)}{\Gamma_i^T G^T (D_{w_i} G \Gamma_i \theta_i^k)} \quad \forall i \in \{1...m\}.
\tag{21}
$$

12

Referring to (19), the expression of the ith column of the profile matrix may be deduced and reads

$$f_i^{k+1} \leftarrow f_i^k \circ \frac{\Gamma_i \Gamma_i^T G^T (D_{w_i} x_i)}{\Gamma_i \Gamma_i^T G^T (D_{w_i} G f_i^k)} \quad \forall i \in \{1...m\}.$$

Given that $\Gamma_i \Gamma_i^T = \text{diag}(1_{p\times 1} - \omega_i)$, with ω_i the ith column of the matrix Ω, it may be noticed that this operator only selects active components among the profile vector in (5). The previous expression may be summarized into matrix formulation, i.e.,

$$F \leftarrow F \circ \left((1_{p\times m} - \Omega) \circ \left[\frac{G^T (W \circ X)}{G^T (W \circ (GF))} \right] \right). \tag{22}$$

In (22), $(1_{p\times m} - \Omega)$ has to be seen as a constraint mask applied to the components of the profile matrix. As a result, (22) is equivalent to WNMF since an appropriate choice of the initial profile $(F^0 \circ \Omega = F^0)$ leads exactly to the same iterations.

3.2.2. The case $\delta_i \neq 0$

Back to (10), let us use it in the case when, δ_i is different from zero. By replacing it into (18), it appears as a non-negative least square cost function

$$K(\theta_i) = \begin{array}{ll} (x_i - G\varphi_i - (G\Gamma_i)\theta_i)^T & D_{w_i}(x_i - G\varphi_i - (G\Gamma_i)\theta_i) \\ \text{s. t.} & \theta_i \geq 0 \quad \forall i \in \{1...m\}. \end{array} \tag{23}$$

It should be noticed that (20) and (23) are similar, except that x_i in (20) has been replaced by $(x - G\varphi)_i$ in (23). From (21), it results in a new expression of the free parameters vector at iteration $(k+1)$:

$$\theta_i^{k+1} \leftarrow \theta_i^k \circ \frac{\Gamma_i^T G^T D_{w_i}(x_i - G\varphi_i)}{\Gamma_i^T G^T (D_{w_i} G \Gamma_i \theta_i^k)} \quad \forall i \in \{1...m\}.$$

Expression (11) enables to update the ith column of the profile matrix:

$$f_i^{k+1} - \varphi_i \leftarrow \left(f_i^k - \varphi_i \right) \circ \frac{\Gamma_i \Gamma_i^T G^T D_{w_i}(x_i - G\varphi_i)}{\Gamma_i \Gamma_i^T G^T (D_{w_i} G(f_i^k - \varphi_i))} \quad \forall i \in \{1...m\}.$$

The previous equation may be written in matrix form, in a similar way as in (22), by noticing that the matrix $\Gamma_i \Gamma_i^T$ acts as a component selector operator equivalent to the constraint mask $(1_{p\times m} - \Omega)$:

13

$$F - \Phi \leftarrow (F - \Phi) \circ \left((1_{p \times m} - \Omega) \circ \left[\frac{G^T(W \circ (X - G\Phi))}{G^T(W \circ (G(F - \Phi)))} \right] \right). \qquad (24)$$

This equation is the global expression which enables to update iterates of the profile matrix. It is the generalization of (22) and has to be compared to Expression (5), which stands for the unconstrained version. It can be seen as the constrained WNMF applied to $(X - G\Phi)$ instead of X (the non-negativity constraint is required on the equivalent data).

Moreover, it can be easily checked that any limit point called $\triangle F^\star$ verifies the complementary slackness condition (last left condition from (17)). However, this method suffers from the same convergence problems as the NMF multiplicative updates proposed by Lee and Seung[2] [24]. Indeed, Gonzales and Zhang pointed the existence of numerical examples showing the fail of this algorithm to reach a stationary point [14]. As our method can be reduced to the same approach by setting $W = 1_{n \times m}$ and by imposing $\Omega = \Phi = 0_{p \times m}$, the same comment here applies. Non stationarity may typically occur when some profile components are initialized to zero.

In the same way, the contribution matrix is updated according to unconstrained WNMF:

$$G \leftarrow G \circ \frac{(W \circ X)F^T}{(W \circ (GF))F^T}. \qquad (25)$$

However, an unconstrained WNMF on the contribution matrix does not ensure the non-negativity of $(X - G\Phi)$. A projection step of equivalent data is thus performed before updating the contribution matrix.

3.3. Outline of the algorithm

From the previous results, we derive the algorithm whose steps are detailed in Algorithm 1. The normalization is made after each computation step by normalizing each row of the profile matrix and by updating the corresponding column of the contribution matrix. As the normalization has no effect on the product $G \cdot F$, the value of the trace in (16) remains unchanged before or after the normalization. However, it makes the constraints unverified within iterations. From a practical point of view, if a limit point is

---

[2]Please note the existence of alternative algorithms which fix this issue [12].

Algorithm 1 Update rules of the CWNMF method.
___
    Reading Φ and Ω.

    Initialization of G and F

    Normalize F and G

    while the stopping rule is not checked **do**

        Check for the non-negativity of $(X - G\Phi)$ and project it if necessary

        Search for F at constant G according to (24)

        Search for G at constant F according to (25)

        Normalize F and G

    end while
___

reached (i.e., $\triangle F^{k+1} = \triangle F^k$), constraints are observed to be verified asymptotically. Despite the fact that—as for other NMF methods [24, 17]— we cannot guarantee the stationarity when a limit point is reached, we never faced convergence issues—i.e., non-stationarity—in the numerous experiments we conducted on our dataset. It should be noticed that limit points tending to zero should be checked with great care. Indeed, although the third KKT condition is always satisfied, the sign of the gradient in the second KKT condition may be negative.

Lastly, we did not meet any kind of issue with the initialization of the matrices, as reported in Appendix B. It must be noticed that the equality constraints limit the number of free parameters to initialize, which probably tends to make the informed NMF method more robust to initialization than blind NMF, as shown in our experiments described in Appendix B.

4. Application to industrial source identification

4.1. Introduction

Data relative to airborne particles with size lower than 10 μm collected at Dunkerque were considered. This site is a medium city located in Northern France on the coast of the North Sea (Figure 1). There is an important industrial activity, especially with the presence of an integrated steelwork complex. Sampling was performed on the roof of the Research Institute and the sampling point is directly under the influence of emissions of the industrial sector when the wind blows from direction 260 degres to 320 degres. In this work, we have selected 92 samples collected under a wider wind sector (i.e. from 250 degres to 60 degres) to consider other influences such as

marine, crustal and long-range transport origins. Concentrations of 19 elements and ions listed in Table 1 have been determined for each sample. In our previous work [8], we focused our attention on the identification of main sources contributing to atmospheric particulate background. Four particles types were evidenced: sea salts, aged sea-salts, secondary inorganic aerosols and crustal particles. Their respective profiles were in good agreement with results from literature [10, 31, 36]. We can assume that the composition of such emissions does not vary with time. In the present work, we aim to identify the industrial emission sources which additionally influence the level of atmospheric particulate at our sampling site. The integrated steelwork complex gathers processes devoted to coke making, iron ores sintering, steel production and slag treatments. A ferromanganese alloys production plant concerned by atmospheric emissions is also located next to the above industrial site. The particulate emissions have different chemical compositions from one process to another and do not always occur simultaneously. For these reasons, it is not adequate to use a single profile to track the impact of this industrial site on atmospheric particulate level. In this study, it is proposed that 5 industrial sources are expected to contribute significantly to airborne particulate matter at a local scale [23].
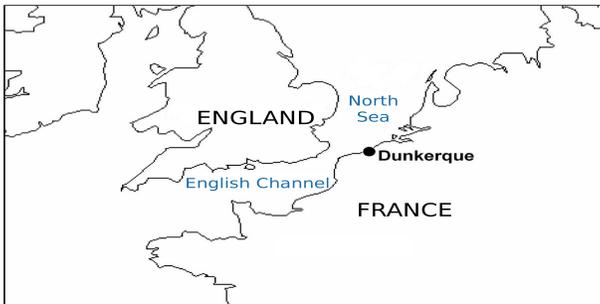

Figure 1: Location of the area under study.

4.2. Major source identification

We assume that 4 background sources are completely known whereas 5 industrial sources are partially unknown. The matrices related to the constraints—i.e., Ω and Φ— are respectively given in Tables 1 and 2. The first four rows represent background sources while the others account for industrial sources. When a value is set to 1, it means that the corresponding species is constrained. For example, copper for the sixth source is set to 0.

16

Table 1: Contraint matrix Ω.

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

Table 2: Matrix Φ of set profile values.

Al	Ca	Cr	Cu	Fe	K	Mg	Mn	Na	Ni	Pb	Sn	Ti	V	Zn	Cl$^-$	NO$_3^-$	SO$_4^{2-}$	NH$_4^+$
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	421	316	263
367	140	.2	.1	147	136	64	2.5	122	.1	.1	0	15	.3	.3	3	2.4	.5	.5
0	10	0	0	0	9	30	0	253	0	0	0	0	0	0	200	280	217	0
0	12	0	0	0	12	37	0	312	0	0	0	0	0	0	550	0	78	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

To sum up, 55 parameters are free in the profile matrix, they are located in the 5 last rows corresponding to the 5 industrial sources. No constraint is applied on the contribution matrix. As a result, it contains 92×9 free parameters. The estimation of these parameters is ordered according to sources in Figures 2-6. In each figure, the top plot shows the contribution of one source (corresponding to one column of the contribution matrix) versus samples, while the middle one gives a rate of the whole mass emitted. The bottom graph shows the source profile (corresponding to one row of the profile matrix). The fifth source denoted blast furnace / steel plant sources is represented on Figure 2 while the steel slag source is given in Figure 3. Two sintering plant sources are detailed on Figures 4 and 5 while a ferromanganese plant source is reported in Figure 6. By comparing these profiles with experimental data defined for each source [23], the profile issued from Figure 2 is referred to as iron rich particles emitted by blast furnaces or steel plant. These two sources have got very similar chemical identities. These features have also been reported for similar processes located on other sites all over the world [27, 28]. The second profile (see Figure 3) is characterized by a high proportion of calcium and iron and then various metallic impurities. Unambiguously, this profile can be attributed to fine particles which gather
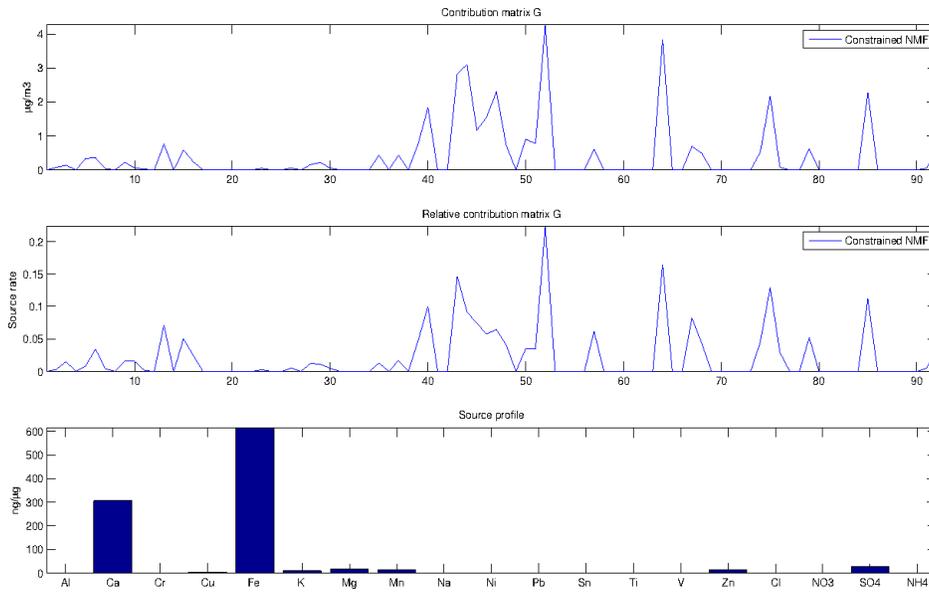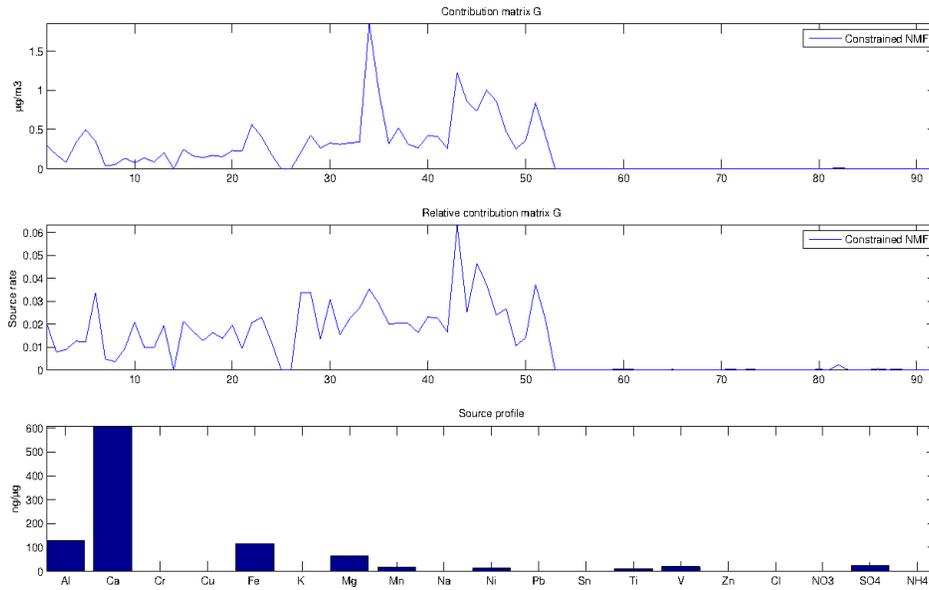
Figure 2: Blast furnaces source.

Figure 3: Steel slag source.

a set of undesirable elements in steel composition and which are separated in the form of slags. Such emissions are classified as fugitive emissions.
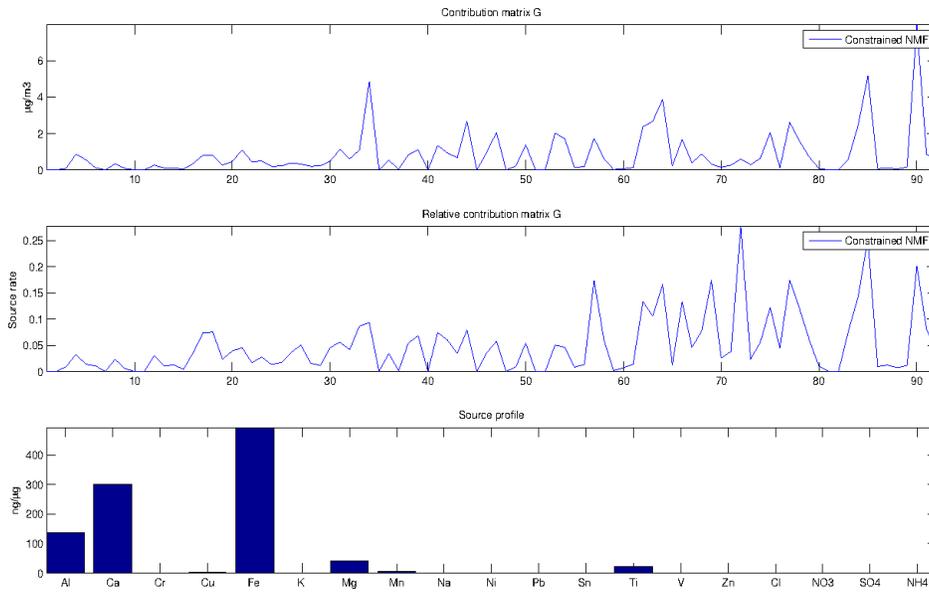
18

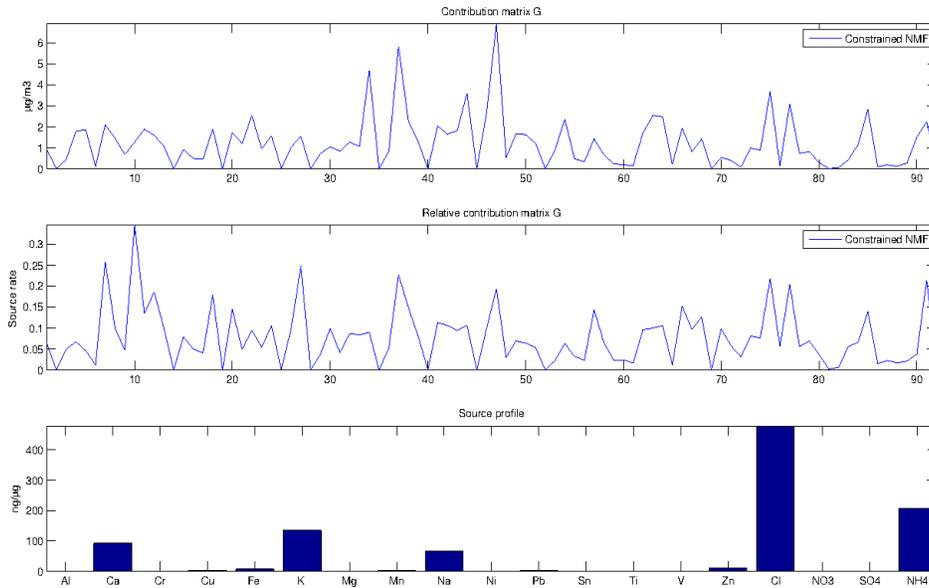Figure 4: Ores sintering plant non point source.

Figure 5: Sintering chimney point source.

The third profile (see Figure 4) highlights species in the following order: Fe> Ca> Al> Mg> Mn> K> Na. It is ascribed to ores sintering plant source
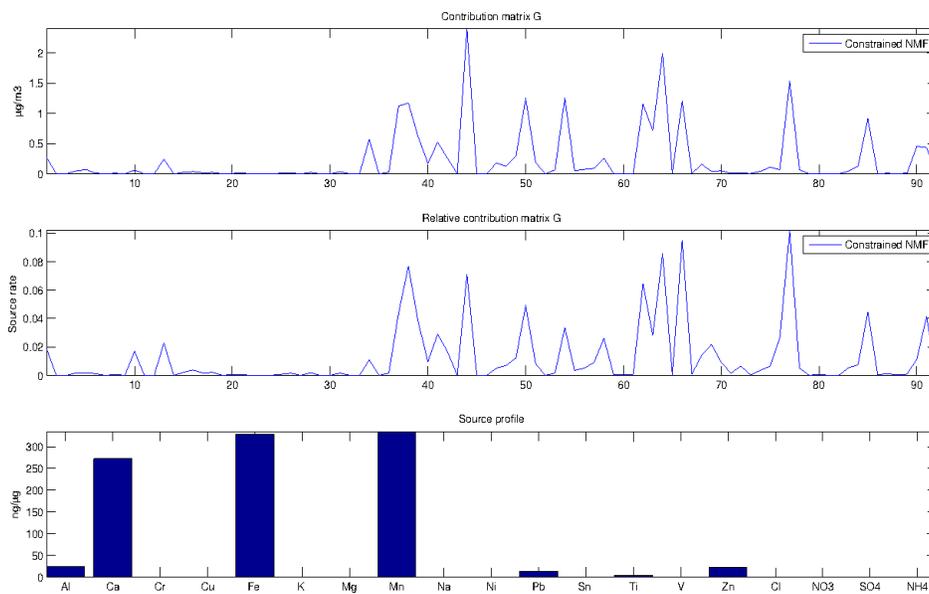
Figure 6: Ferromanganese plant source.

and such particles correspond to fugitive emissions depending on all handling performed on sintered ores. It comes from the heating of raw materials. However, a lack of potassium and sodium may be noted. The fourth source profile (Figure 5) is expected to be mainly composed with Cl^-, K, and Fe. It stands for the sintering chimney point source. But it can be noticed that there is a lack of Fe particles and also NH_4^+ is over estimated. Finally, a Ferromanganese plant source may be recognized on Figure 6, it can be noted that the expected profile is very well recovered.

A brief summary of significant sources show that 3 sources (blast furnaces source, steel slags source, sintering chimney point source) present a few peaks over 3 $\mu g/m^3$. The steel slag source and the ferromanganese plant source are minor contributors to particles pollution. Hence, these results enable to evidence the impact of point source emissions on atmospheric particulate levels and particularly provide for the first time a relative contribution of fugitive emissions from this site.

5. Conclusion

We have introduced in this article the use of basic equality constraints and have derived theoretical expressions of constrained WNMF. We provided

global multiplicative rules which take into account at the same time linear equality constraints and weights. The second interest stands in the choice of the application which aims to find natural and industrial profiles. Partial knowledge of these profiles enables to focus the search on species which are completely unknown. As a consequence, according to chemists, our results allow them to better understand the industrial profiles. This shows the relevance of our proposed approach. Upcoming prospects concern the integration of inequality constraints in the algorithms.

Acknowledgements

Appendix A. Expression of the KKT conditions in the constrained problem

Karush-Kuhn Tucker conditions are carried out without loss of generality in the case when $\Phi = 0$. It may be easily extended to the general case. For the sake of concision, let assume that $\overline{\Omega} \triangleq (1_{p \times m} - \Omega)$. The minimization of the criterion (16) leads to consider a function $H(.)$, defined by

$$H(\triangle F) = \text{Tr}\left((X - G(\triangle F \circ \overline{\Omega}))^T((X - G(\triangle F \circ \overline{\Omega})) \circ W)\right) = H_1 - 2H_2 + H_3,$$

where
$$\begin{cases} H_1 &= \text{Tr}\left(X^T(X \circ W)\right), \\ H_2 &= \text{Tr}\left((X \circ W)^T G(\triangle F \circ \overline{\Omega})\right), \\ H_3 &= \text{Tr}\left((\triangle F \circ \overline{\Omega})^T G^T((G(\triangle F \circ \overline{\Omega})) \circ W)\right). \end{cases}$$
The property $\text{Tr}(U^T(V \circ W)) = \text{Tr}((U \circ W)^T V)$ enables to write the differential of H_2, H_3 as

$$\begin{cases} \partial H_1 &= 0, \\ \partial H_2 &= \text{Tr}\left((((X \circ W)^T G) \circ \overline{\Omega}^T)\partial \triangle F\right), \\ \partial H_3 &= 2 \cdot \text{Tr}\left((((W \circ G(\overline{\Omega} \triangle F))^T.G) \circ \overline{\Omega}^T)\partial \triangle F\right). \end{cases}$$

The differentiation of H_2 and H_3 then results in

$$\begin{cases} \frac{\partial H_2}{\partial F} &= \overline{\Omega} \circ (G^T(X \circ W)), \\ \frac{\partial H_3}{\partial F} &= 2[G^T(W \circ (G(\overline{\Omega} \circ \triangle F)))] \circ \overline{\Omega}. \end{cases} \tag{A.1}$$

According to (A.1), the differentiation of the function $H(.)$ gives

$$\frac{\partial H}{\partial \triangle F} = 2\overline{\Omega} \circ \left(G^T((G[\overline{\Omega} \circ \triangle F] - X) \circ W)\right). \tag{A.2}$$

The Lagrangian function may be expressed as

$$L(G, F) = \frac{1}{2}H(\triangle F) - \mathrm{Tr}(\Gamma_G \circ G) - \mathrm{Tr}(\Gamma_F \circ \triangle F \circ \overline{\Omega}).$$

Cancelling the derivatives of this Lagrangian function with respect to $\triangle F$ yields the following relation,

$$\frac{\partial L}{\partial \triangle F} = \overline{\Omega} \circ \left(G^T((G[\overline{\Omega} \circ \triangle F] - X) \circ W)\right) - \Gamma_F = 0, \tag{A.3}$$

leading to an expression of Γ_F. Karush-Kuhn Tucker conditions for the minimization of (16) may be summed up as

$$\begin{array}{ll} \triangle F \geq 0, & G \geq 0, \\ \frac{\partial H}{\partial \triangle F} \geq 0, & \frac{\partial H}{\partial G} \geq 0, \\ \triangle F \circ \Gamma_F = 0, & G \circ \Gamma_G = 0. \end{array} \tag{A.4}$$

For the sake of clarity, let us define $\widetilde{F} \triangleq \overline{\Omega} \circ \triangle F$. By replacing the relations (A.2) and (A.3) into (A.4) and by using the same analogy with the matrix G, the final KKT conditions for the constrained weighted problem are

$$\begin{array}{ll} \triangle F \geq 0, & G \geq 0, \\ \overline{\Omega} \circ \left(G^T((G\widetilde{F} - X) \circ W)\right) \geq 0, & (W \circ (G\widetilde{F} - X))\widetilde{F}^T \geq 0, \\ \triangle F \circ \overline{\Omega} \circ \left(G^T((G\widetilde{F} - X) \circ W)\right) = 0, & G \circ \left((W \circ (G\widetilde{F} - X))\widetilde{F}^T\right) = 0. \end{array}$$

Table B.3: Used matrix Ω for the NMF method.

Fe	Ca	SO$_4$	Zn	Mg	Al	Cr
1	0	0	0	0	0	1
0	1	1	1	0	1	0
1	0	0	1	0	0	1

Table B.4: Used matrix Φ for the NMF method.

Fe	Ca	SO$_4$	Zn	Mg	Al	Cr
700	0	0	0	0	0	0
0	400	5	0	0	75	0
400	0	0	0	0	0	0

Appendix B. Sensitivity of the algorithm to the initialization

The proposed NMF algorithm has been tested in the framework of pollutant source separation. As explained in Subsection 3.3—and as it is the case for classical NMF approaches [24, 17]—the robustness of our method to initialization is not theoretically guaranteed (but was shown not to be an issue with the considered application). In this appendix, we aim to investigate this behaviour: we generate synthetic data according to industrial source profiles available in the literature [28]. Seven chemical species—listed in Table B.3—are under consideration. Three industrial sources coming from steel industry (blast furnaces source, steel slag source, ores sintering plant) shown in this order in Table B.5 are assumed to be active.

The data matrix consists of 50 samples and 7 species, with a known uncertainty measure σ_{ij}—provided by a chemical expert—associated to each data point x_{ij}. The uncertainty measure allows us to compute the weight matrix W, according to $w_{ij} = \sigma_{ij}^{-2}$. A uniform noise ranging in $[-\min\{\lambda\,\sigma_{ij}, x_{ij}\}; \lambda\,\sigma_{ij}]$ is added while keeping non-negativity of the data. Please note that λ is related to an input Signal-to-Noise Ratio (SNR). In the experiments below, the input SNR is equal to 30 dB. The actual profile matrix is given in Table B.5.

In our method, we can select which components of the profile matrix are to be set, making it highly flexible. In the tests reported in this appendix, 9 constraints are taken into account and are shown in Table B.3, where the 1s indicate the location of the constraints in the profile matrix. The matrix Φ is provided in Table B.4. Note that 70 percent of the first source profile is set, so that it only subsists 30 percent of it to be distributed over the remaining species. To assess the quality of the fit, we apply different performance

Table B.5: Real profile matrix F.

Fe	Ca	SO$_4$	Zn	Mg	Al	Cr
700	100	80	60	40	20	0
300	400	5	0	200	75	20
400	200	80	0	120	200	0

measures—similar to SNRs—on the contribution and the profile matrices.

- When applied to the columns of G, the chosen measure is named Mixing-Error Ratio (MER) [34]. We express each estimated column $\hat{g}_j$ of the matrix G as

$$\hat{g}_j = g_j^{coll} + g_j^{orth},$$

  where g_j^{coll} and g_j^{orth} are respectively collinear and orthogonal to the *true* vector g_j. The MER (in dB) associated with this estimated column is then expressed as

$$\text{MER}_j = 10 \log_{10} \left\| g_j^{coll} \right\|^2 / \left\| g_j^{orth} \right\|^2.$$

  This definition provides one MER for each source—respectively denoted as S_1, S_2, and S_3—so that p MERs are obtained.

- When applied to the lines of F—or equivalently to the columns of F^T— we measure a Signal-to-Interference Ratio (SIR) which is classically met in blind source separation [34]. The SIR is computed for each estimated source. From a practical point of view, we apply the above MER algorithm on the columns of F^T as an SIR measure.

The sensitivity of the algorithm is studied by choosing different initializations of the profile matrix F_{init}. We exclusively use initial profile matrices which satisfy the constraints defined in Tables B.3 and B.4. For each initialization, we run the algorithm during 20 thousand iterations.

We performed numerous tests in this configuration and all the obtained enhancement provided by our NMF approach was consistent. For the sake of readibility, we only gather 12 significant initial profile matrices in Table B.6. This table provides three performance indices:

1. the first three columns show the SIR obtained for each source when each line of F is set to its initialization. We see that the performance is pretty low, as it ranges from 1.15 to 24.07 dB. While the SIR can be

Table B.6: Performance of the NMF approach for different initializations. Performance criteria: SIR (in dB) and MER (in dB).

trial	SIR (on F_{init})			MER (on G)			SIR (on F)		
number	S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
1	24.07	18.56	16.50	29.89	30.06	32.44	54.61	78.18	54.81
2	17.31	18.56	16.50	29.89	30.05	32.44	54.78	81.66	54.61
3	13.96	18.56	16.50	29.89	30.06	32.44	54.64	80.19	54.78
4	24.07	3.87	16.50	29.89	30.06	32.44	54.61	78.18	54.81
5	24.07	6.11	16.50	29.89	30.06	32.44	54.61	78.18	54.81
6	24.07	18.56	3.19	29.89	30.06	32.44	54.61	78.18	54.81
7	24.07	18.56	1.15	29.89	30.06	32.44	54.61	78.18	54.81
8	17.31	18.56	1.15	29.89	30.06	32.44	54.61	78.18	54.81
9	13.96	18.56	1.15	29.89	30.06	32.44	54.61	78.18	54.81
10	13.96	6.11	1.15	29.89	30.06	32.44	54.61	78.18	54.81
11	13.96	3.87	1.15	29.89	30.06	32.44	54.61	78.18	54.81
12	10.47	3.87	1.15	29.89	30.06	32.44	54.61	78.18	54.81

really low for Source 2 and Source 3, we could not reach an SIR lower than 10 dB with Source 1 because the mass to be distributed over the free parameters is quite low, as discussed above.

2. The last three columns show the SIRs obtained after running the algorithm. We notice that, whatever the initialization, the SIR for each source keeps almost constant and is high, i.e., between 54.61 and 81.66 dB. This shows that the algorithm converges to the same stationary point.

3. Such a result is seen again in the three middle columns. They show the MERs obtained after applying the NMF algorithm. Here again—and as for the SIR—the measure obtained for each source remains almost constant over the trials. It can be noticed that SIRs—estimated over the profile matrix—are greater than MERs which estimated over the contribution matrix. It may be due to the fact that the constraints are applied on the profile matrix only.

Other experiments have been conducted with less constraints. We noticed that relaxing some of the above constraints made the algorithm more sensitive to the initialization. This was expected as less constraints means more parameters to estimate.

To conclude, these experiments show the relevance of our informed NMF method. It not only provides a good performance using the additional knowledge on the profile matrix F, but it is also less sensitive to the initialization than its blind counterpart.

References

[1] S. Arora, R. Ge, R. Kannan, A. Moitra, Computing a nonnegative matrix factorization – provably, in: Proceedings of the 44th symposium on Theory of Computing, 2012, pp. 145–162.

[2] W. Berry, M. Browne, N. Langville, V. Pauca, J. Plemmons, Algorithms and applications for approximate non-negative matrix factorization, Computational Statistics & Data Analysis 52 (1) (2007) 155–173.

[3] D. P. Bertsekas, Nonlinear Programming, 2nd ed., Athena Scientific, 1999.

[4] V. Bittorf, C. Ré, B. Recht, J. Tropp, Factoring nonnegative matrices with linear programs, in: Advances in Neural Information Processing Systems 25 (NIPS), 2012, pp. 1223–1231.

[5] S. Choi, Algorithms for orthogonal non-negative matrix factorization, International Joint Conference on Neural Networks, Hong Kong, 2008.

[6] A. Cichocki, S. Cruces, S. Amari, Generalized alpha-beta divergences and their application to robust non-negative matrix factorization, Entropy 13 (1) (2011) 134–170.

[7] A. Cichocki, R. Zdunek, NMFLAB for signal processing, Technical report, Laboratory for Advanced Brain Signal Processing, BSI RIKEN, Saitama, Japan (2006).

[8] G. Delmaire, G. Roussel, D. Hleis, F. Ledoux, Une version pondérée de la factorisation matricielle non négative pour l'identification de sources de poussières. Application au littoral de la mer du nord, RS-JESA 44 (4-5) (2010) 547–566.

[9] A. Escrig, E. Monfort, I. Celades, X. Querol, F. Amato, M. Minguillón, P. Hopke, Application of optimally scaled target factor analysis for assessing source contribution of ambient PM10, Journal of the Air Waste Management Association 59 (29) (2009) 1296–1307.

[10] M. C. Evans, S. W. Campbell, V. Bhethanabotla, 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 38 (29) (2004) 4847–4858.

[11] E. Fishler, M. Grosmann, H. Messer, Detection of signals by information theoretic criteria: general asymptotic performance analysis, IEEE Transactions on Signal Processing 50 (5) (2002) 1027–1036.

[12] N. Gillis, F. Glineur, Accelerated multiplicative updates and hierarchical als algorithms for nonnegative matrix factorization, Neural Computation 24 (4) (2012) 1085–1105.

[13] N. Gillis, S. Vavasis, Fast and robust recursive algorithms for separable nonnegative matrix factorization, IEEE Trans. on Pattern Analysis and Machine Intelligence 36 (4) (2014) 698–714.

[14] E. F. Gonzales, Y. Zhang, Accelerating the Lee-Seung algorithm for nonnegative matrix factorization, Technical report, Dept. of Computational and Applied Mathematics, Rice University (2005).

[15] L. Grippo, M. Sciandrone, On the convergence of the block nonlinear gauss-seidel method under convex constraints., Oper. Res. Lett. 26 (3) (2000) 127–136.

[16] D. Guillamet, J. Vitria, B. Schiele, Introducing a weighted non-negative matrix factorization for image classification, Pattern Recognition Letters 24 (14) (2003) 2447–2454.

[17] N.-D. Ho, Non-negative matrix factorizations algorithms and applications, Phd thesis, Université Catholique de Louvain (2008).

[18] N.-D. Ho, P. Vandooren, Non-negative matrix factorization with fixed row and column sums, Linear Algebra and its Applications 429 (5-6) (2008) 1020–1025.

[19] P. K. Hopke, K. Ito, T. Mar, W. F. Christensen, D. J. Eatough, R. C. Henry, E. Kim, F. Laden, R. Lall, T. V. Larson, H. Liu, L. Neas, J. Pinto, M. Stölzel, H. Suh, P. Paatero, G. D. Thurston, PM source apportionment and health effects: Intercomparison of source apportionment results, Journal of Exposure Science and Environmental Epidemiology 16 (3) (2006) 275–286.

[20] P. O. Hoyer, Non-negative matrix factorization with sparseness constraint, Journal of Machine Learning Research 5 (2004) 1457–1469.

[21] H. Kim, H. Park, Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method, SIAM Journal on Matrix Analysis and Applications 30 (2) (2008) 713–730.

[22] J. Kim, Y. He, H. Park, Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework, Journal of Global Optimization (2013) 1–35.

[23] H. Laversin, D. Courcot, F. Ledoux, Tracing the atmospheric particulate emissions from an iron and steel works, Chemical Engineering Transactions 10 (2006) 179–184.

[24] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.

[25] C. J. Lin, On the convergence of multiplicative update algorithms for non-negative matrix factorization, IEEE Transactions on Neural Networks 18 (6) (2007) 1589–1596.

[26] C. J. Lin, Projected gradients methods for non-negative matrix factorization, Neural Computation 19 (10) (2007) 2756–2779.

[27] S. D. Machemer, Characterization of airborne and bulk particulate from iron and steel manufacturing facilities, Environ. Sci. Techn. 38 (2) (2004) 381–389.

[28] K. Oravisjarvi, K. Timonen, T. Wiikinkoski, A. Ruuskanen, K. Heinanen, J. Ruuskanen, Source contributions to PM2.5 particles in the urban air of a town situated close to a steel works, Atmospheric Environment 37 (2) (2003) 1013–1022.

[29] P. Paatero, Least squares formulation of robust non negative factor analysis, Chemometrics 12 (5) (1997) 788–791.

[30] P. Paatero, U. Tapper, Positive matrix factorization: a non negative factor model with optimal utilization of error estimates of data values, Environmetrics 5 (2) (1994) 111–126.

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

[32] S. Vavasis, On the complexity of nonnegative matrix factorization, SIAM J. on Optimization 20 (3) (2009) 1364–1377.

[33] M. Viana, T. A. J. Kuhlbusch, X. Querol, A. Alastuey, The effects of particle-size distribution and chloride depletion of sea-salt aerosols on estimating atmospheric deposition at a coastal site, Journal of Aerosol Science 39 (10) (2008) 827–849.

[34] E. Vincent, S. Araki, P. Bofill, The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation, in: Proc. of ICA, 2009, pp. 734–741.

[35] J. Yoo, S. Choi, Non-negative matrix factorization with orthogonality constraints, Journal of computing science and engineering 4 (2) (June 2010) 97–109.

[36] Y. Zhao, Y. Gao, Acidic species and chloride depletion in coarse aerosol particles in the US east coast, Science of the total Environment 407 (1) (2008) 541–547.