ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Enhanced NMF initialization using a physical model for pollution source apportionment

Marc Plouvin[1], Abdelhakim Limem[1], Matthieu Puigt[1],
Gilles Delmaire[1], Gilles Roussel[1], and Dominique Courcot[2] *

1- LISIC, ULCO, Université Lille Nord de France, Calais, France, FR-62228
2- UCEIV, ULCO, Université Lille Nord de France, Dunkerque, France, FR-59140

Abstract. In a previous work, we proposed an informed Non-negative Matrix Factorization (NMF) with a specific parametrization which involves constraints about some known components of the factorization. In this paper we extend the above work by adding some information provided by a physical dispersion model. In particular, we derive a special structure of one of the factorizing matrices, which provides a better initialization of the NMF procedure. Experiments on simulated mixtures of particulate matter sources show that our new approach outperforms both our previous one and the state-of-the-art NMF methods.

1 Introduction

Source apportionment consists of estimating which particulate matter sources are present in the ambient air, with their relative concentrations. A source is fully characterized by a *profile* which consists of m chemical species proportions (expressed in ng/ng). Usually, several, say n, data samples are collected from a chemical sampler and can be written as mixtures of p profiles, with different concentrations (expressed in ng/m^3). Mathematically, if we respectively denote by $\overline{X}$, G, and F the non-negative $n \times m$ data matrix, $n \times p$ contribution matrix, and $p \times m$ profile matrix, the collected data read

$$\overline{X} \approx G \cdot F. \tag{1}$$

G and F are usually unknown and estimating them from $\overline{X}$ is a Blind Source Separation (BSS) problem [1], which can be solved, e.g., by Non-negative Matrix Factorization (NMF) [1, Ch. 13]. NMF was massively investigated since the pioneering works in [2, 3]. Most methods consist of minimizing a dissimilarity measure (see [4] for examples of such measures) between $\overline{X}$ and $G \cdot F$. However, [3] and its extensions are very sensitive to the matrix initialization, and moreover the convergence to a stationary point is not guaranteed, especially if some components of F or G are zero. To tackle this issue, adding assumptions [5] or initializing NMF with the output of another BSS methods [6] were introduced.

Moreover, when applied to real pollution source data, the performance of the above classical NMF methods is inconsistent [7]. Fortunately, in source apportionment, some partial knowledge on the matrices may be available. In our recent work [4], we proposed an *informed* NMF method where some known

*This work was supported partly by ArcelorMittal, and partly by the DREAL Agency.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

entries of F were integrated in NMF as constraints, thus providing an approach in between BSS and regression (i.e., where F is fully known).

In this paper, we propose to incorporate additional knowledge in the matrix G. In particular, by using a physical dispersion model of the particulate matter [8, 9], we obtain a reduced structure of the contribution matrix G which allows to exclude some sources at some sampling instances (i.e., we control the sparseness of G). This structure is then seen as constraints in the NMF procedure. Using dispersion models for inverse problems was previously investigated in, e.g., [8, 10, 11]. In particular, the authors in [10, 11] considered the convolutive propagation of a unique source in a fluid—i.e., atmosphere or water—while we here introduce a non-convolutive model for multiple sources.

Given our newly proposed structure and an inital matrix F provided by experts, we propose a new NMF initialization based on quadratic programming. This configuration yields a better separation performance when compared to blind and informed NMF, as shown in this paper on simulated mixtures of industrial and natural source profiles. The remainder of the paper is organized as follows. The considered problem, the dispersion model, and the associated definitions and assumptions are given in Section 2. We introduce the incorporation of the physical model in the NMF intialization in Section 3. The performance of our method is provided in Section 4 while the conclusion is outlined in Section 5.

2 Modelization, definitions, and assumptions

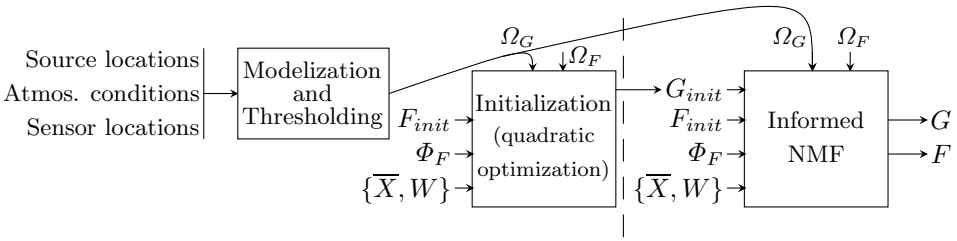

Fig. 1: Structure of the proposed informed NMF method.

In this paper, we assume that a chemical sampler—located at the position $\xi_s \triangleq (\mathrm{x}_s, \mathrm{y}_s, \mathrm{z}_s)$—regularly estimates m chemical species concentrations. They are collected in a $n \times m$ data matrix $\overline{X}$ defined in Eq. (1) together with a weight matrix W derived from the known uncertainty measures provided by a chemical expert [4]. These species may have a natural origin—they can be due to, e.g., sea salts—or an industrial one, i.e, they may come from chemical plants or steel factories. In particular, we assume to know the locations—denoted $\xi_l \triangleq (\mathrm{x}_l, \mathrm{y}_l, \mathrm{z}_l)$ with $1 \leq l \leq q$—of the q industrial sites ($q \leq p$) which emit particles in the atmosphere and which are in the neighbourhood of the sampler.

Our objective consists of accurately estimating F and G, defined in Eq. (1). Figure 1 describes the successive steps of the whole method. The stages on the left part of the plot are novel and aim to optimize the structure and the initialization of the informed NMF. The latter was introduced in [4], and its

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

iterative solution is based on a special parametrization which reads

$$F = \Phi_F \circ \Omega_F + (1 - \Omega_F) \circ \Delta F, \tag{2}$$

where Ω_F is a mask describing the positions of the known values of F, Φ_F is the matrix of known values, ΔF is the matrix of the free components of F, and $\circ$ denotes the componentwise product. The iterative updates rules [4]—recalled below for the sake of clarity—are defined for a β-divergence and read:

$$\triangle F^{k+1} \leftarrow \triangle F^k \circ (1 - \Omega_F) \circ N_{F^k}, \quad G^{k+1} \leftarrow G^k \circ \frac{(W \circ \overline{X} \circ (GF)^{\beta-1})F^T}{(W \circ (GF)^{\beta})F^T}, \tag{3}$$

where k is the current iteration index and

$$N_{F^k} = \frac{G^T.\left[W \circ (\overline{X} - G\Phi_F) \circ [G(F^k - \Phi_F)]^{\beta-1}\right]}{G^T.\,[W \circ [G(F^k - \Phi_F)]^{\beta}]}. \tag{4}$$

These updates are followed by a normalization stage. As shown in [4], the proposed method outperformed state-of-the-art NMF methods. However, it might suffer from poor results if G is sparse, because of the sensitivity of NMF to its initialization (particularly while some components are zeros). In this paper, we propose to use a physical dispersion model to inform—from known wind directions—which sources did or did not contribute to the samples. We thus derive the sparse structure of G as a binary mask denoted Ω_G (see Fig. 1). Many physical models are available in the literature, e.g., convolutive [11] or stationary Gaussian models [12]. Since the sampling period of concentration data is much longer than the transfer of particles from a source to a sensor, it would be without interest to use a convolutive model. Instead, we choose a Gaussian plume model [9] which computes an atmospheric transfer coefficient, denoted t hereafter, according to

$$t(\xi'_s, \xi'_l, u) \triangleq \frac{\exp(-\frac{(\mathrm{y}'_s - \mathrm{y}'_l)^2}{2\sigma^2_{\mathrm{y}'_l}}) \left(\exp(-\frac{(\mathrm{z}'_l - \mathrm{z}'_s)^2}{2\sigma^2_{\mathrm{z}'_l}}) + \exp(-\frac{(\mathrm{z}'_l + \mathrm{z}'_s)^2}{2\sigma^2_{\mathrm{z}'_l}})\right)}{2\pi u \sigma_{\mathrm{y}'_l} \sigma_{\mathrm{z}'_l}}, \tag{5}$$

where ξ'_s and ξ'_l are the coordinates ξ_s and ξ_l, respectively expressed in a new basis which sets the wind in the x-direction. u denotes the wind speed while $\sigma_{\mathrm{y}'_l}$ and $\sigma_{\mathrm{z}'_l}$ shape the Gaussian plume [9]. Typically, $t \in \left[0, 10^{-6}\right]$ and the highest values denote an important transfer between a source and the sensor. This model is assumed to be valid in quasi-stationary wind conditions for a source-sensor transfer. For one row in $\overline{X}$, we actually measure several, say ν, wind velocities and angles, which provide ν atmospheric transfer coefficients for each industrial source. They are gathered in a $\nu \times q$ matrix denoted T below.

3 Incorporating a special structure into NMF

3.1 Model incorporation and thresholding

We now show how we link the dispersion model with the observed data $\overline{X}$ defined in Eq. (1). Let us first recall that $\overline{X}$ gathers n measurements. However—and

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

as explained in Section 2—for each measurement, we actually get ν atmospheric transfer coefficients since the wind data is oversampled by a factor ν. If we were also oversampling the data measurement by a factor ν, for each row of $\overline{X}$, we would obtain a $\nu \times m$ data matrix X which could be written as

$$X \triangleq X^I + X^N \triangleq G^I F^I + G^N F^N, \tag{6}$$

where the superscripts ^{I} and ^{N} denote the industrial and natural activities, respectively. We also assume that for each measurement of $\overline{X}$, the flow rate of each industrial site is constant. Chemical flows can then be gathered in a $q \times m$ matrix[1] denoted Q. This implies that

$$X^I = TQ. \tag{7}$$

Actually, Q is equal to F^I, up to a normalization stage, i.e., rows of F^I are equal to normalized rows of Q. From this relationship—the demonstration is not provided for space considerations—Eq. (7) can be expressed as

$$X^I = (T \circ \mathcal{Q})F^I, \tag{8}$$

where $\mathcal{Q} \triangleq 1_{\nu \times m} Q^T$, $1_{\nu \times m}$ is a $\nu \times m$ matrix of ones, and the superscript ^{T} denotes the transposition. Eq. (8) combined with Eq. (6) shows that G^I is a function of T. In particular, low entries of T imply that the corresponding entries in G^I are low as well.

As averaging the rows in X provides one row of $\overline{X}$, the industrial concentrations in one row of G correspond to averaged concentrations of $T \circ \mathcal{Q}$ which are denoted $\overline{T} \circ \overline{\mathcal{Q}}$ below. In practice, we derive from the dispersion model a vector $\overline{T}$ which provides the contribution of one industrial source in a sample. When some values of this vector are low, this implies that some contributions are negligible and can be rounded to zero. We first gather each vector $\overline{T}$ in a $n \times q$ matrix $\overline{\mathcal{T}}$. We then derive a $n \times p$ mask Ω_G which provides the locations of the zeros in G, i.e.

$$\Omega_{Gi,j} = \begin{cases} 0 & \text{if } \overline{\mathcal{T}}_{i,j} \geq \frac{\max_j(\overline{\mathcal{T}}_{i,j})}{10^5} \text{ or } j > q \\ 1 & \text{otherwise} \end{cases} \tag{9}$$

This enables to discard sources which do not contribute to the mixture. This structure matrix is similar to Ω_F defined in Eq. (5).

3.2 NMF initialization by quadratic optimization

Let $\underline{w}_i$, $\underline{\overline{x}}_i$, and $\underline{g}_i$ be the i^{th} row of W, $\overline{X}$, and G, respectively. The i^{th} row of the contribution matrix minimizes a weighted least-square cost function under some specific constraints, i.e.,

$$J(g_i) = \left(\underline{\overline{x}}_i^T - F^T \underline{g}_i^T\right)^T \cdot D_{w_i} \cdot \left(\underline{\overline{x}}_i^T - F^T \underline{g}_i^T\right), \tag{10}$$

[1]The extension to time-varying emissions—not provided in this paper for space considerations—can be derived by noticing that Q becomes a $\nu \times q \times m$ tensor.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

where $D_{w_i} = \text{diag}(\underline{w}_i)$. This function may be written under a quadratic form

$$J(g_i) = \frac{1}{2}\underline{g}_i H \underline{g}_i^T + u^T \underline{g}_i^T, \tag{11}$$

with $H = 2FD_{w_i}F^T$ and $u^T = -2\underline{\overline{x}}_i D_{w_i} F^T$. Initializing G consists of estimating

$$\min_{g_i} J(g_i) \quad \text{s.t.} \quad g_i^T \geq 0, \sum_i \underline{g}_i = \sum_i \underline{\overline{x}}_i, \underline{\overline{x}}_i^T \geq \Phi_F^T \cdot \underline{g}_i^T, \text{and } \underline{g}_i \circ \Omega_{Gi} = 0. \tag{12}$$

The four constraints state that (i) each entry of $\underline{g}_i$ is non-negative, (ii) the contributions must be normalized because the profiles are proportions [13], (iii) the known part of the data has to be lower than the whole data, and (iv) the values of G on the constraints are zero. In practice, Eq. (12) is solved using the interior-point convex algorithm from Matlab.

4 Experimental validation

In order to measure the performance of our proposed method, we simulate 250 mixtures of $p = 3$ sources—i.e., of $q = 2$ industrial and one natural sources—with various input Signal-to-Noise Ratio (SNR) conditions. The wind direction is chosen so that only one industrial source contributes to the mixtures whereas the natural source is always active. $\nu = 48$ wind conditions are given with each observation data row. The data matrix consists of $n = 50$ samples and $m = 7$ species—listed in Table 1—with a weight matrix W. A uniform noise is added while keeping positivity of the data, as explained in [4]. The chosen dissimilarity measure between $\overline{X}$ and GF is a Frobenius norm, i.e., $\beta = 1$ in Eq. (3). Four approaches are tested in this paper, i.e., the multiplicative NMF [3], its Weighted counterpart (WNMF) [4], our Constrained WNMF [4] (CWNMF), and the proposed Model-based Constrained WNMF (MCWNMF) methods.

Fe	Ca	SO_4	Zn	Mg	Al	Cr
0	0	0	0	0	0	1
0	0	0	0	0	0	0
1	0	1	0	0	0	0

Table 1: Used matrix Ω_F for the CWNMF and the MCWNMF methods.

To assess the quality of the fit, we use the Mixing-Error Ratio (MER) which was also used in [4]. We generate several matrices $\overline{X}$, with an input SNR ranging from 15 to more than 70 dB. Figure 2 provides the MERs averaged along input SNR intervals. It shows that our proposed MCWNMF outperforms all the other methods, even for low to moderate input SNRs where the informed CWNMF provides almost the same performance as the blind WNMF approach.

5 Conclusion

In this paper, we extended our previous informed NMF method [4] by using an atmospheric dispersion model in order to inform the zero locations of one of the

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

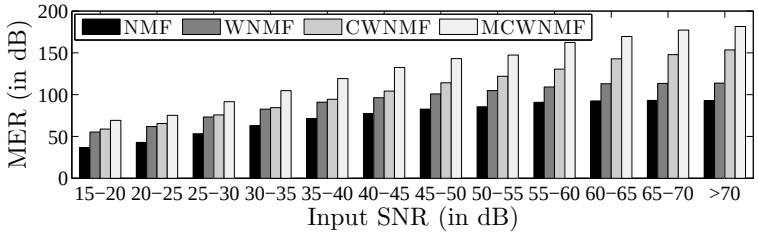

Fig. 2: MERs of the tested NMF methods, vs the input SNR.

factorized matrices. Experiments conducted with various input signal-to-noise ratio conditions showed the relevance of the proposed method. In future work, we will validate the enhancement it provides on real data.

References

[1] P. Comon and C. Jutten. *Handbook of blind source separation : independent component analysis and applications*. Academic Press, 2010.

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

[3] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[4] A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot. Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment. In *Proc. of MLSP*, 2013.

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

[6] D. Benachir, Y. Deville, S. Hosseini, M. S. Karoui, and A. Hameurlain. Hyperspectral image unmixing by non-negative matrix factorization initialized with modified independent component analysis. In *Proc. of WHISPERS*, 2013.

[7] M. Viana, T. A. J. Kuhlbusch, X. Querol, and A. Alastuey. Source apportionment of particulate matter in europe: a review of methods and results. *Journal of Aerosol Science*, 39(10):827–849, October 2008.

[8] J. M. Stockie. The mathematics of atmospheric dispersion modeling. *SIAM Rev.*, 53(2):349–372, May 2011.

[9] D. B. Turner. *Workbook of atmospheric dispersion estimates: An introduction to dispersion modelling*. Lewis Publisher, second edition, 1995.

[10] G. Delmaire and G. Roussel. Joint estimation decision methods for source localization and restoration in parametric convolution processes. Application to accidental pollutant release. *Digital Signal Processing*, 22(1):34–46, 2012.

[11] J. Ranieri, I. Dokmaníc, A. Chebira, and M. Vetterli. Sampling and reconstruction of time-varying atmospheric emissions. In *Proc. of ICASSP*, 2012.

[12] C. Leroy, D. Maro, D. Hébert, L. Solier, M. Rozet, S. Le Cavelier, and O. Connan. A study of the atmospheric dispersion of a high release of krypton-85 above a complex coastal terrain, comparison with the predictions of Gaussian models (Briggs, Doury, ADMS4). *Journal of Environmental Radioactivity*, 101(11):937–944, 2010.

[13] H. Lantéri, C. Theys, C. Richard, and C. Févotte. Split gradient method for nonnegative matrix factorization. In *Proc. of EUSIPCO*, pages 1199–1203, 2010.