Informed Non-negative Matrix Factorization for Industrial Pollution Identification

Matthieu Puigt

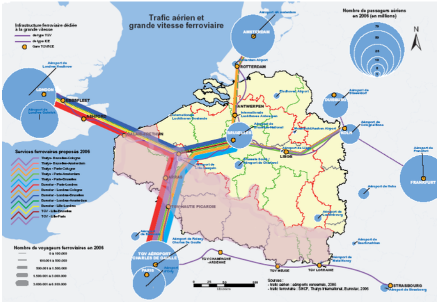
LISIC, EA 4491, ULCO, Université Lille Nord de France,
Calais, France, FR-62228
matthieu.puigt@lisic.univ-littoral.fr
http://www-lisic.univ-littoral.fr/~puigt

FORTH ICS Lecture

October 31, 2014

Introducing ULCO & LISIC (1)

The Nord–Pas de Calais area



- Northern France
- European crossroad (connects Southern and Northern Europe, borders with Belgium & the UK)
- Very rich history (major French centre of heavy industry in the 19th century, strategic situation during WWs, etc) and beautiful natural sites
- Not considered as the warmest place in France...

Introducing ULCO & LISIC (1)

The Nord–Pas de Calais area

- Northern France
- European crossroad (connects Southern and Northern Europe, borders with Belgium & the UK)
- Very rich history (major French centre of heavy industry in the 19th century, strategic situation during WWs, etc) and beautiful natural sites
- Not considered as the warmest place in France...

Introducing ULCO & LISIC (1)

The Nord–Pas de Calais area

Last August

- Northern France
- European crossroad (connects Southern and Northern Europe, borders with Belgium & the UK)
- Very rich history (major French centre of heavy industry in the 19th century, strategic situation during WWs, etc) and beautiful natural sites
- Not considered as the warmest place in France...

Introducing ULCO & LISIC (2)

Université du Littoral Côte d'Opale (ULCO)

- `http://www.univ-littoral.fr/`
- "Proximity" University born in 1993
- 4 sites ($\simeq$ 40 km between Calais and the other sites):
  - Boulogne (fishing industry)
  - Calais (chemistry, ferry port, shuttle)
  - Dunkerque (industry port—metallurgy, petrol, etc—and energy—nuclear plants)
  - Saint-Omer (marshes used for agriculture, industry—crystal, paper, cardboard)
- Research activities with applications in environment (air, ground, water)

Introducing ULCO & LISIC (3)

Laboratoire d'Informatique, Signal, Image de la Côte d'Opale (LISIC)

- Created in Calais in 2010 (fusion of 2 ULCO labs) : 39 permanent faculty members, $\simeq$10 Ph.D. students, $\simeq$3 post-docs
- Four research teams with theoretical computer science and signal processing researchers:
  1. IMAP (Images and Learning)
  2. OSMOSE (Evolutionnary modelization, optimization, simulation)
  3. MODEL (Multi-Modelisation et Software Evolution)
  4. SPeciFI (Peception systems and Information Fusion)
- Several research projects in collaboration with industry (ArcelorMittal, Innocold, etc), research institutes (CNRS, IFREMER, etc) or public institutions (DREAL, Région Nord–Pas de Calais, etc)

Framework

- Funded by the integrated steel and mining company ArcelorMittal (2010–2013) and by the Nord–Pas de Calais DREAL Agency (2013–2015)
- Team work: involved A. Limem, M. Plouvin, G. Delmaire, M. Puigt, and G. Roussel (LISIC), and A. Kfoury, F. Ledoux, and D. Courcot (UCEIV–ULCO)

The talk will focus on the work published in:

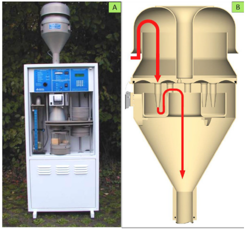
- A. Limem, G. Delmaire, M. Puigt, G. Roussel, D. Courcot: *Non-negative matrix factorization under equality constraints—a study of industrial source identification*, Applied Numerical Mathematics, vol. 85, pp. 1–15, Nov. 2014
- A. Limem, M. Puigt, G. Delmaire, G. Roussel, D. Courcot: *Bound constrained weighted NMF for industrial source apportionment*, in Proc. MLSP, 2014, Reims, France
- M. Plouvin, A. Limem, M. Puigt, G. Delmaire, G. Roussel, D. Courcot: *Enhanced NMF initialization using a physical model for pollution source apportionment*, in Proc. ESANN, 2014, Bruges, Belgium
- A. Limem, G. Delmaire, M. Puigt, G. Roussel, D. Courcot: *Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment*, in Proc. MLSP, 2013, Southampton, UK

Outline of the talk

1. Problem statement
   - Problem statement
   - The big picture
   - A short history of NMF
2. Adding expert's knowledge into NMF
   - Parameterization of the equality constraints
   - Derived optained rules
   - Adding more constraints
   - Performance of the proposed approaches
3. Adding a physical model to inform NMF
   - Motivation
   - propagation models
   - Modelization of pollutant source propagation
   - Incorporation of a special structure into NMF
   - NMF initialization by quadratic optimization
   - Performance of the model-based approaches
4. Tests with real data
   - Experimental conditions
   - Results
5. Conclusion

Outline of the talk

Problem statement (1)



- Particulate matter sensing
- n data samples are analyzed by chemists.
- Observed data are set in a $n \times m$ matrix X of m chemical species concentrations (in ng/m^3)
- Observed data are mixtures of chemical "profiles" (up to **outliers**)

$$X \simeq G \cdot F$$

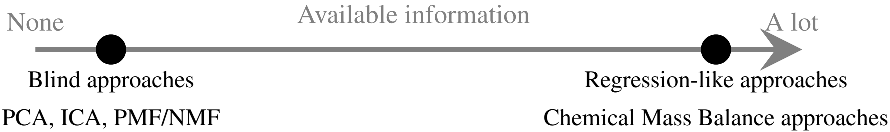
  - ➥ G is the $n \times p$ contribution matrix (ng/m^3)
  - ➥ F is the $p \times m$ profile matrix (chemical species relative proportions) of sources (ng/ng)

➪ How to estimate G and F from X?

Problem statement (2)

What chemists use to do

$$X \simeq G \cdot F$$

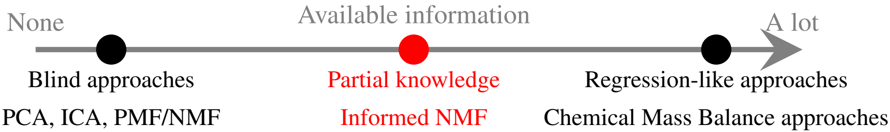


- Data are positive and profiles are potentially correlated
- ➪ Positive Matrix Factorization / Non-negative Matrix Factorization (PMF/NMF) better-suited
- But inconsistent performance (Viana *et al.*, 2008)

Problem statement (2)

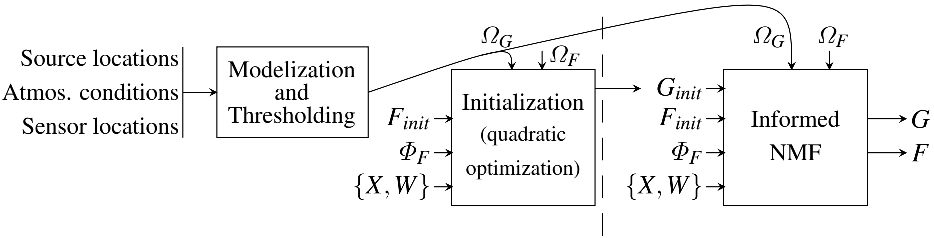
What chemists use to do

$$X \simeq G \cdot F$$



- Data are positive and profiles are potentially correlated
- ➯ Positive Matrix Factorization / Non-negative Matrix Factorization (PMF/NMF) better-suited
- But inconsistent performance (Viana *et al.*, 2008)
- Additionnal information:
  1. Some *known* or *bounded* entries of F
  2. Rows of F are normalized
  3. Uncertainty matrix Σ associated with the observed data (classical in PMF but not in NMF)
  4. Known location of the industrial sources and of the chemical sampler, and sensed wind directions and speeds

The big picture



- Ω_G and Ω_F: structure information on the matrices
- Φ_F: value information on the matrices

A short history of NMF into 2 slides (1)

Goal: estimate G and F which minimize the discrepancy between X and $G\dot{F}$:

$$\{\hat{G}, \hat{F}\} = \arg\min_{G,F} \mathcal{D}(X||G \cdot F),$$

where $\mathcal{D}$ may be the Frobenius norm of $X - GF$ or a divergence.

1. **PMF** (Paatero and Tapper, 1994): estimation of both G and F at once, using non-negative least-square techniques (multiple local minima, limited to small-sized problems)
2. **NMF** (Lee and Seung, 1999): alternating methods with multiplicative updates (uniqueness of the solution not guaranteed)
   - Weighted NMF (Guillamet *et al.*, 2003, Ho, 2008): add a weight matrix into the factorization (different confidence in the data)
   - NMF using parametric divergences (Cichocki *et al.*, 2011, Févotte and Idier, 2011)
   - Many others (e.g., adding sparseness—Hoyer, 2004)
3. **ANLS** (Alternating Non-neg. Least Squares—Kim & Park, 2008): alternating approach (as NMF) based on non-negative LS (as PMF), high computational cost at each iteration (but less iterations than NMF).
4. **BCD** (Block Coordinate Descent—Kim *et al.*, 2013): generalize the ANLS approaches and divides the problems in more than 2 blocks.

A short history of NMF into 2 slides (1)

NMF (with Frobenius norm)

- Optimizing

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

➩ Update rules:

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

A short history of NMF into 2 slides (1)

Goal: estimate G and F which minimize the discrepancy between X and $G\dot{F}$:

$$\{\hat{G}, \hat{F}\} = \arg\min_{G,F} \mathcal{D}(X||G \cdot F),$$

where $\mathcal{D}$ may be the Frobenius norm of $X - GF$ or a divergence.

1. **PMF** (Paatero and Tapper, 1994): estimation of both G and F at once, using non-negative least-square techniques (multiple local minima, limited to small-sized problems)
2. **NMF** (Lee and Seung, 1999): alternating methods with multiplicative updates (uniqueness of the solution not guaranteed)
   - Weighted NMF (Guillamet *et al.*, 2003, Ho, 2008): add a weight matrix into the factorization (different confidence in the data)
   - NMF using parametric divergences (Cichocki *et al.*, 2011, Févotte and Idier, 2011)
   - Many others (e.g., adding sparseness—Hoyer, 2004)
3. **ANLS** (Alternating Non-neg. Least Squares—Kim & Park, 2008): alternating approach (as NMF) based on non-negative LS (as PMF), high computational cost at each iteration (but less iterations than NMF).
4. **BCD** (Block Coordinate Descent—Kim *et al.*, 2013): generalize the ANLS approaches and divides the problems in more than 2 blocks.

A short history of NMF into 2 slides (1)

Weighted NMF—WNMF—with Frobenius norm

- Optimizing

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

where σ_{ij} is the (i, j)-th element of Σ.

- Update rules:

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

where $W \triangleq \frac{1_{n \times m}}{\Sigma \circ \Sigma}$.

A short history of NMF into 2 slides (1)

Goal: estimate G and F which minimize the discrepancy between X and $G\dot{F}$:

$$\{\hat{G},\hat{F}\} = \arg\min_{G,F} \mathcal{D}(X||G\cdot F),$$

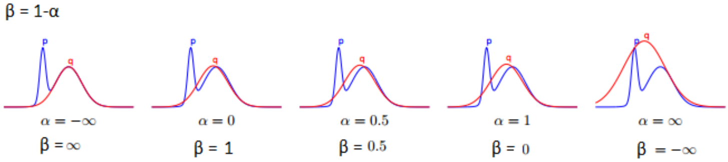
where $\mathcal{D}$ may be the Frobenius norm of $X - GF$ or a divergence.

1. **PMF** (Paatero and Tapper, 1994): estimation of both G and F at once, using non-negative least-square techniques (multiple local minima, limited to small-sized problems)
2. **NMF** (Lee and Seung, 1999): alternating methods with multiplicative updates (uniqueness of the solution not guaranteed)
   - Weighted NMF (Guillamet *et al.*, 2003, Ho, 2008): add a weight matrix into the factorization (different confidence in the data)
   - NMF using parametric divergences (Cichocki *et al.*, 2011, Févotte and Idier, 2011)
   - Many others (e.g., adding sparseness—Hoyer, 2004)
3. **ANLS** (Alternating Non-neg. Least Squares—Kim & Park, 2008): alternating approach (as NMF) based on non-negative LS (as PMF), high computational cost at each iteration (but less iterations than NMF).
4. **BCD** (Block Coordinate Descent—Kim *et al.*, 2013): generalize the ANLS approaches and divides the problems in more than 2 blocks.

A short history of NMF into 2 slides (1)

Parametric divergences

- Parametric (α-, β- or) αβ-divergences encompass various divergences (KL, Bregman, Itakura-Saito, etc) and the Frobenius norm
- Non-symmetrical, may be robust to outliers



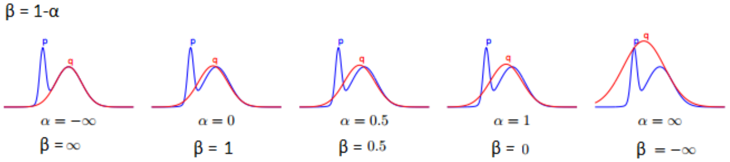
A short history of NMF into 2 slides (1)

Parametric divergences

- Parametric (α-, β- or) αβ-divergences encompass various divergences (KL, Bregman, Itakura-Saito, etc) and the Frobenius norm
- Non-symmetrical, may be robust to outliers
- Extensions to weighted divergence, e.g., for $(\alpha, \beta, \alpha+\beta) \neq 0$:

$$\mathcal{D}^{\alpha,\beta}(p_{ij}||q_{ij}) = \frac{-1}{\alpha\beta}\left(p_{ij}^{\alpha}q_{ij}^{\beta} - \frac{\alpha}{\alpha+\beta}p_{ij}^{\alpha+\beta} - \frac{\beta}{\alpha+\beta}q_{ij}^{\alpha+\beta}\right),$$

$$\mathcal{D}_{W}^{\alpha,\beta}(p_{ij}||q_{ij}) = \frac{-1}{\alpha\beta}\sigma_{ij}^{-(\alpha+\beta)}\left(p_{ij}^{\alpha}q_{ij}^{\beta} - \frac{\alpha}{\alpha+\beta}p_{ij}^{\alpha+\beta} - \frac{\beta}{\alpha+\beta}q_{ij}^{\alpha+\beta}\right).$$



A short history of NMF into 2 slides (1)

Goal: estimate G and F which minimize the discrepancy between X and $G\dot{F}$:

$$\{\hat{G}, \hat{F}\} = \arg\min_{G,F} \mathcal{D}(X||G \cdot F),$$

where $\mathcal{D}$ may be the Frobenius norm of $X - GF$ or a divergence.

1. **PMF** (Paatero and Tapper, 1994): estimation of both G and F at once, using non-negative least-square techniques (multiple local minima, limited to small-sized problems)
2. **NMF** (Lee and Seung, 1999): alternating methods with multiplicative updates (uniqueness of the solution not guaranteed)
   - Weighted NMF (Guillamet *et al.*, 2003, Ho, 2008): add a weight matrix into the factorization (different confidence in the data)
   - NMF using parametric divergences (Cichocki *et al.*, 2011, Févotte and Idier, 2011)
   - Many others (e.g., adding sparseness—Hoyer, 2004)
3. **ANLS** (Alternating Non-neg. Least Squares—Kim & Park, 2008): alternating approach (as NMF) based on non-negative LS (as PMF), high computational cost at each iteration (but less iterations than NMF).
4. **BCD** (Block Coordinate Descent—Kim *et al.*, 2013): generalize the ANLS approaches and divides the problems in more than 2 blocks.

A short history of NMF into 2 slides (2)

5. **Projected Gradients** (Lin, 2007): gradient method which projects the data onto the domain of admissible values
6. **Sum-to-one constraint** (row constraints—Lantéry *et al.*, 2010—or column constraints—Miao and Qi, 2007)
7. **Non-negative PCA** (Plumbley and Oja, 2004) and **Orthogonal NMF** (Yoo and Choi, 2010)
8. **Non-negative ICA** (Plumbley, 2002, 2003)

The above methods are performing an *approximate* factorization

Exact factorization (Vavasis, 2009)

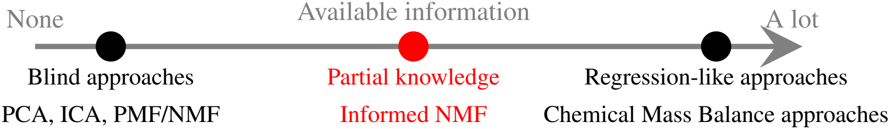
- Not guaranteed in general
- Only reachable with sparse assumptions aka "near separability" (Donoho and Stodden, 2003)

➪ Recent approaches (Arora *et al.*, 2012, Gillis *et al.*, 2012–) provably perform exact NMF

Outline of the talk

Adding expert's knowledge into NMF (1)

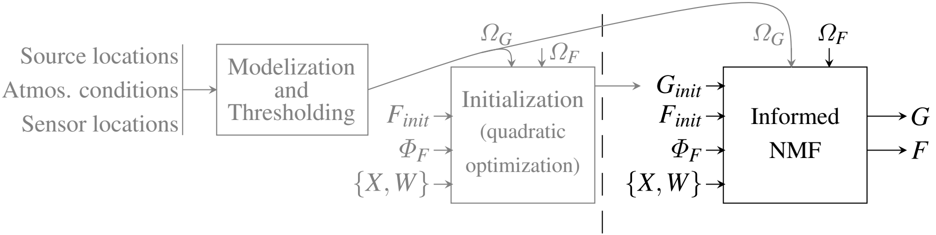
$$X \simeq G \cdot F$$



- As explained before, we have some information on F
  - some known entries (expert's knowledge),
  - some bounded entries (expert's knowledge),
  - row sums are equal to 1
- Possibility to add a weight matrix (expert's knowledge)
- Proposed approaches are extensions of both Weighted and parametric divergence NMF methods
- As shown before, some existing approaches already have some interesting properties we are looking for (but none have all of them)
- Our main contribution: **parameterization** of the NMF problem
- Let us first focus on approaches with known entries

Adding expert's knowledge into NMF (2)

Going into the big picture



Parameterization of the equality constraints

Parameterization of the equality constraints

$$\Omega_{ij}^E \triangleq \begin{cases} 1 & \text{if } F_{ij} \text{ has to be set,} \\ 0 & \text{otherwise.} \end{cases} \qquad \phi_{ij}^E \triangleq \begin{cases} f_{ij} & \text{if } \Omega_{ij} = 1, \\ 0 & \text{otherwise.} \end{cases} \qquad \Phi^E \triangleq F \circ \Omega^E.$$

F then reads:

$$F = \underbrace{\Omega^E \circ \Phi^E}_{\text{set}} + \underbrace{\overline{\Omega}^E \circ \Delta F}_{\text{free}}. \qquad F \geq \Phi^E \qquad \Delta F \geq 0$$

Parameterization of the equality constraints

Example with $p = 3$ sources, $m = 5$ species, and 3 constraints:

$$\Omega^E = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\Phi^E = \Omega^E \circ F = \begin{bmatrix} 0 & 80 & 0 & 0 & 0 \\ 0 & 0 & 0 & 30 & 0 \\ 0 & 30 & 0 & 0 & 0 \end{bmatrix}.$$

Parameterization of the equality constraints

Parameterization of the equality constraints

$$\Omega_{ij}^E \triangleq \begin{cases} 1 & \text{if } F_{ij} \text{ has to be set,} \\ 0 & \text{otherwise.} \end{cases} \qquad \phi_{ij}^E \triangleq \begin{cases} f_{ij} & \text{if } \Omega_{ij} = 1, \\ 0 & \text{otherwise.} \end{cases} \qquad \Phi^E \triangleq F \circ \Omega^E.$$

F then reads:

$$F = \underbrace{\Omega^E \circ \Phi^E}_{\text{set}} + \underbrace{\overline{\Omega}^E \circ \Delta F}_{\text{free}}. \qquad F \geq \Phi^E \qquad \Delta F \geq 0$$

The relationship between X and GF reads:

$$X \simeq G \cdot (\Omega^E \circ \Phi^E) + G \cdot (\overline{\Omega}^E \circ \Delta F),$$

$$X - G \cdot (\Omega^E \circ \Phi^E) \simeq G \cdot (\overline{\Omega}^E \circ \Delta F),$$

Parameterization of the equality constraints

Parameterization of the equality constraints

$$\Omega^E_{ij} \triangleq \begin{cases} 1 & \text{if } F_{ij} \text{ has to be set,} \\ 0 & \text{otherwise.} \end{cases} \qquad \phi^E_{ij} \triangleq \begin{cases} f_{ij} & \text{if } \Omega_{ij} = 1, \\ 0 & \text{otherwise.} \end{cases} \qquad \Phi^E \triangleq F \circ \Omega^E.$$

F then reads:

$$F = \underbrace{\Omega^E \circ \Phi^E}_{\text{set}} + \underbrace{\overline{\Omega}^E \circ \Delta F}_{\text{free}}. \qquad F \geq \Phi^E \qquad \Delta F \geq 0$$

The relationship between X and GF reads:

$$X \simeq G \cdot (\Omega^E \circ \Phi^E) + G \cdot (\overline{\Omega}^E \circ \Delta F),$$

$$X - G \cdot (\Omega^E \circ \Phi^E) \simeq G \cdot (\overline{\Omega}^E \circ \Delta F),$$

and the divergences between known and unknown matrices read

$$\mathcal{D}_W^{\alpha\beta}(X || G \cdot \Phi^E + G \cdot \Delta F),$$

$$\mathcal{D}_W^{\alpha\beta}(X - G \cdot \Phi^E || G \cdot \Delta F).$$

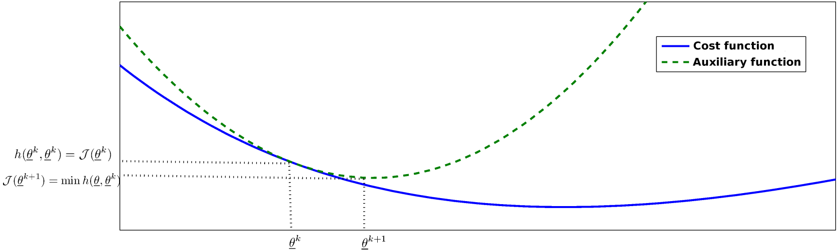
How does the historical NMF work?

- Alternating strategies:
  1. assume G to be known and update F
  2. assume F to be known and update G
- Based on multiplicative updates
- Deriving the update rules from the optimization problem is usually based on a Majoration-Minimization (MM) strategy
  - I won't go through the details

How does the historical NMF work?

- We find an auxiliary function (usually derived from the norm or divergence expression to minimize) of the cost function
- The auxiliary function is convex (often quadratic wrt the parameters), and above or equal (at the current point) to the cost function
- We then estimate the minimum from this auxiliary function

➪ Provide the next point and the update rule



How does the historical NMF work?

- Alternating strategies:
  1. assume G to be known and update F
  2. assume F to be known and update G
- Based on multiplicative updates
- Deriving the update rules from the optimization problem is usually based on a Majoration-Minimization (MM) strategy
  - I won't go through the details

αβ-WNMF algorithm

1. Initialize G and F
2. At each iteration k, repeat
   - $F^{k+1} \leftarrow F^k \circ \left[\frac{(G^k)^T \left(W \circ X^\alpha \circ \left(G^k F^k \right)^{\beta-1} \right)}{(G^k)^T (W \circ (G^k F^k)^{\alpha+\beta-1})} \right]^{\left(\frac{1}{\alpha}\right)}$
   - $G^{k+1} \leftarrow G^k \circ \left[\frac{\left(W \circ \left(X^\alpha \circ \left(G^k F^{k+1} \right)^{\beta-1} \right) \right) (F^{k+1})^T}{(W \circ (G^k F^{k+1})^{\alpha+\beta-1}) (F^{k+1})^T} \right]^{\left(\frac{1}{\alpha}\right)}$

until $F^{k+1} = F^k$ and $G^{k+1} = G^k$

Update rules of the informed methods (1)

- We only add information on F ➪ the update rules for G remain the same
- I won't go through the details of the optimization procedure

Update rules of the informed methods (1)

- We only add information on F ➪ the update rules for G remain the same
- I won't go through the details of the optimization procedure

update rule for αβ-WNMF

$$F^{k+1} \leftarrow F^{k} \circ R_F^{\alpha,\beta}$$

with

$$R_F^{\alpha,\beta} = \left[\frac{(G^k)^T \left(W \circ X^{\alpha} \circ \left(G^k F^k \right)^{\beta-1} \right)}{(G^k)^T (W \circ (G^k F^k)^{\alpha+\beta-1})} \right]^{\left(\frac{1}{\alpha}\right)}$$

update rule for αβ-CWNMFs

$F = \Phi^E + \Delta F \circ \overline{\Omega}^E$

➪ $F^{k+1} \leftarrow \Phi^E + \triangle F^k \circ \overline{\Omega}^E \circ R_F^{\alpha,\beta}$ with $R_F^{\alpha,\beta}$ which depends on the chosen divergence

Update rules of the informed methods (1)

- We only add information on F ➪ the update rules for G remain the same
- I won't go through the details of the optimization procedure

update rule for αβ-WNMF

$$F^{k+1} \leftarrow F^k \circ R_F^{\alpha,\beta}$$

with

$$R_F^{\alpha,\beta} = \left[\frac{(G^k)^T \left(W \circ X^{\alpha} \circ \left(G^k F^k\right)^{\beta-1} \right)}{(G^k)^T (W \circ (G^k F^k)^{\alpha+\beta-1})} \right]^{\left(\frac{1}{\alpha}\right)}$$

update rule for αβ-CWNMFs

$$F = \Phi^E + \Delta F \circ \overline{\Omega}^E$$

➪ $F^{k+1} \leftarrow \Phi^E + \triangle F^k \circ \overline{\Omega}^E \circ R_F^{\alpha,\beta}$ with $R_F^{\alpha,\beta}$ which depends on the chosen divergence

1. Approach using residuals (αβ-CWNMF-R)

$$R_F^{\alpha,\beta} = M_F^{\alpha,\beta} = \left[\frac{G^T \left(W \circ \left((X - G\Phi^E)_+ \right)^{\alpha} \circ \left((G(F - \Phi^E))_+ \right)^{\beta-1} \right)}{G^T \left(W \circ ((G(F - \Phi^E))_+)^{(\alpha+\beta-1)} \right)} \right]^{\left(\frac{1}{\alpha}\right)}.$$

2. Approach without residuals (αβ-CWNMF)

$$R_F^{\alpha,\beta} = N_F^{\alpha,\beta} = \left[\frac{G^T \left(W \circ X^{\alpha+\beta-1} \circ \left((X - G\Phi^E)_+ \right)^{1-\beta} \circ \left((G(F - \Phi^E))_+ \right)^{\beta-1} \right)}{G^T \left(W \circ X^{\alpha+\beta-1} \circ \left((X - G\Phi^E)_+ \right)^{1-\alpha-\beta} \circ ((G(F - \Phi^E))_+)^{(\alpha+\beta-1)} \right)} \right]^{\left(\frac{1}{\alpha}\right)}.$$

Update rules of the informed methods (2)

- Get an eye on:

$$M_F^{\alpha,\beta} = \left[\frac{G^T\left(W \circ \left((X - G\Phi^E)^+\right)^{\alpha} \circ \left((G(F - \Phi^E))^+\right)^{\beta-1}\right)}{G^T\left(W \circ ((G(F-\Phi^E))^+)^{(\alpha+\beta-1)}\right)} \right]^{\left(\frac{1}{\alpha}\right)}.$$

and

$$N_F^{\alpha,\beta} = \left[\frac{G^T\left(W \circ X^{\alpha+\beta-1} \circ \left((X - G\Phi^E)^+\right)^{1-\beta} \circ \left((G(F - \Phi^E))^+\right)^{\beta-1}\right)}{G^T\left(W \circ X^{\alpha+\beta-1} \circ \left((X - G\Phi^E)^+\right)^{1-\alpha-\beta} \circ ((G(F-\Phi^E))^+)^{(\alpha+\beta-1)}\right)} \right]^{\left(\frac{1}{\alpha}\right)}.$$

- If we define

$$W' = W \circ X^{\alpha+\beta-1} \circ \left((X - G\Phi^E)^+\right)^{1-\alpha-\beta}$$

$N_F^{\alpha,\beta}$ has the same expression as $M_F^{\alpha,\beta}$, except that the values of W' vary within iterations

- That is, the divergence without residuals provides rules which iteratively update the weights within iterations

Adding the remaining constraints

- As rows of F are proportions, their sum is equal to 1.
- Moreover, some entries of F which are not exactly known are bounded by experts.
- How to consider all these constraints in addition to the previous one?

➯ Sequential approach which iteratively:

1. Estimate G and F with the above method
2. Normalize the rows of F (we will see how later)
3. Project data onto their admissible domain

or

1. Estimate G and F with the above method
2. Project data onto their admissible domain
3. Normalize the rows of F (we will see how later)

Normalization strategies

- Normalizing is quite classical in many problems (e.g., remote sensing)

1) Normalize all the rows of F ($\alpha\beta$-N_1-CWNMF)

$$\tilde{F}^{k+1} \leftarrow \frac{\Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ R_F^{\alpha\beta}}{[\Phi^E + \Delta F^k \circ \overline{\Omega}^E \circ R_F^{\alpha\beta}] \cdot 1_{mm}}.$$

✔ Steepest descent direction in the optimization
✔ No effect of the normalization on the product $G \cdot F$
✗ Equality constraints lost (not an issue within iterations)

2) Normalize the free parameters only ($\alpha\beta$-N_2-CWNMF)

$$\tilde{F}^{k+1} \leftarrow \Phi^E + \frac{\Delta F^k \circ \overline{\Omega}^E \circ R_F^{\alpha\beta}}{(\Delta F^k \circ \overline{\Omega}^E \circ R_F^{\alpha\beta}) \cdot 1_{mm}} \circ (1_{pm} - \Phi^E \cdot 1_{mm}).$$

✔ Equality constraints kept
✗ Descent direction not optimal

Bound constraints

Parameterization of the bound constraints

$$\Omega^E \circ \Omega^I = 0 \qquad \Omega^I_{ij} = \begin{cases} 1 & \text{if } F_{ij} \text{ is bounded,} \\ 0 & \text{otherwise.} \end{cases} \qquad \Phi^{I-} \leq F \circ \Omega^I \leq \Phi^{I+},$$

Projection of the data onto their admissible domain

In practice, we define the projection operator $\mathcal{P}_{\Omega^I}$ as:

$$\mathcal{P}_{\Omega^I}(U) \triangleq \begin{cases} \Omega^I \circ \Phi^{I-} & \text{if } \Omega^I \circ U \leq \Phi^{I-}, \\ \Omega^I \circ \Phi^{I+} & \text{if } \Omega^I \circ U \geq \Phi^{I+}, \\ \Omega^I \circ U & \text{otherwise.} \end{cases}$$

Finally, we get 8 possible methods:

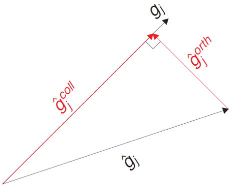
1. Projections iteratively applied after the above normalizations ($\alpha\beta$-N_1B-CWNMF(-R), $\alpha\beta$-N_2B-CWNMF(-R))
2. Alternative: first project and then normalize ($\alpha\beta$-BN_1-CWNMF(-R), $\alpha\beta$-BN_2-CWNMF(-R))

Performance of the proposed approaches (1)

First series of tests

- Simulations ($n = 50$ samples, $m = 7$ chemical species)
- $p = 3$ sources (2 industrial + 1 natural)
- Additive uniform noise (keeping the data positive)
- 5 equality constraints

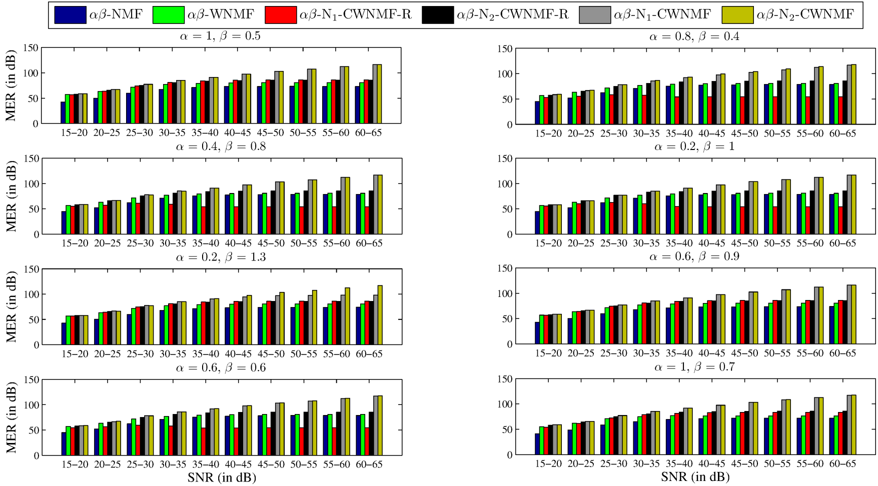
Fe	*Ca*	SO_4^{2-}	*Zn*	*Mg*	*Al*	*Cr*
X	X	X	60	X	X	0
X	X	5	X	X	X	20
X	200	X	X	X	X	X



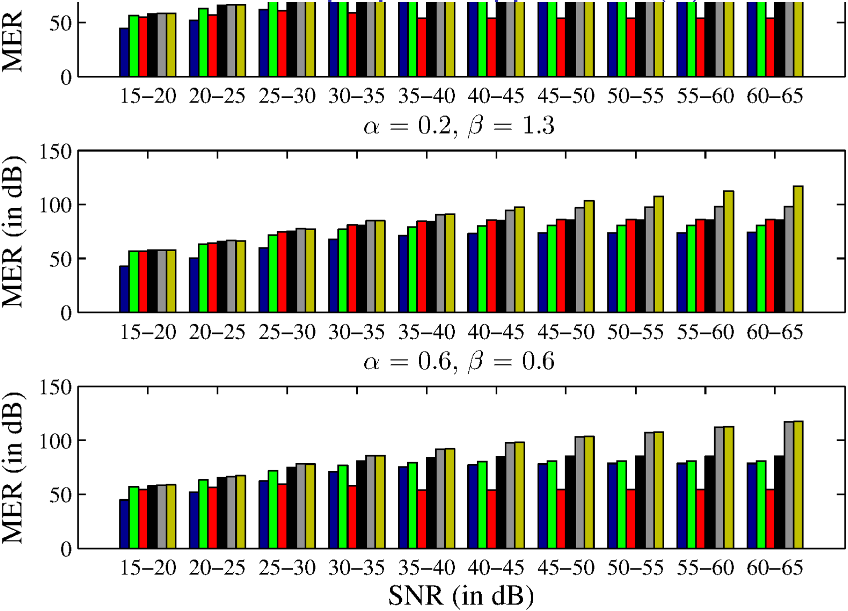
- Tested methods: Blind (αβ-NMF, αβ-WNMF) and informed methods (αβ-N_1-CWNMF, αβ-N_2-CWNMF, αβ-N_1-CWNMF-R, αβ-N_2-CWNMF-R)
- Measure of performance: Mixing-Error-Ratio (MER) expressed in dB

$$\hat{g}_j = \hat{g}_j^{coll} + \hat{g}_j^{orth}, \quad \mathrm{MER}_j = 10\log_{10}\frac{\left\|\hat{g}_j^{coll}\right\|^2}{\left\|\hat{g}_j^{orth}\right\|^2}, \quad \mathrm{MER} = \sum_j \mathrm{MER}_j$$

Performance of the proposed approaches (2)

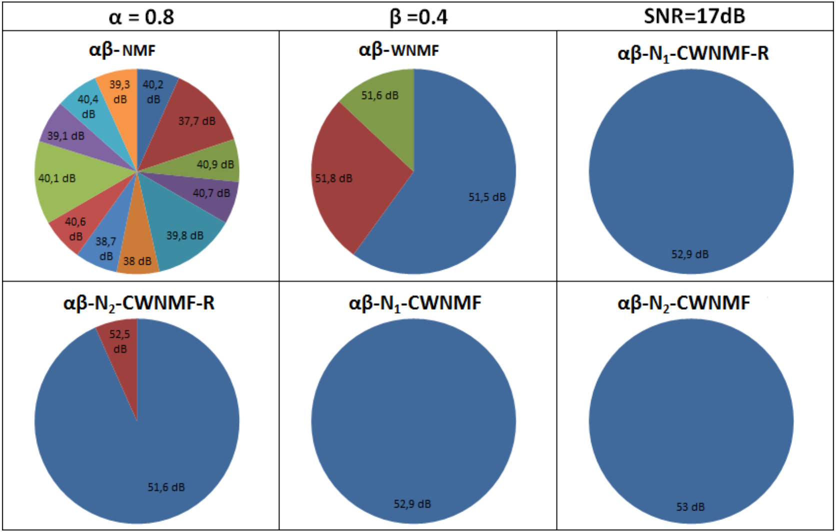


Performance of the proposed approaches (2)



Performance of the proposed approaches (3)

- Performance obtained with 15 different initializations
- Some initialized matrices far from the theoretical values



Performance of the proposed approaches (4)

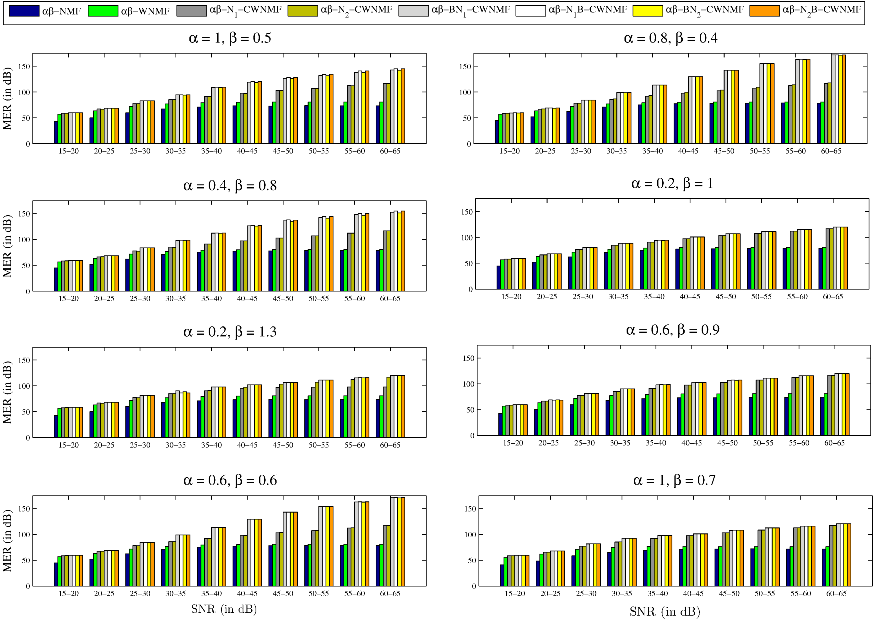
Second series of tests

- Same sources and observations as before
- Tested methods: methods (without residuals) with and without bound constraints
- 6 constraints (3 known values + 3 bounds)

	Fe	*Ca*	SO_4	*Zn*	*Mg*	*Al*	*Cr*
Source 1	X	X	X	40/80	X	X	0
Source 2	280/320	X	5	X	X	X	20
Source 3	X	180/220	X	X	X	X	X

- Measure of performance: MER

Performance of the proposed approaches (5)



Conclusion of the first part

- We proposed several informed NMF methods
- Special parameterization of equality constraints
- Sequential approach to deal with all the constraints
- 8 resulting approaches:
  - 2 divergence expressions (with or without residuals)
  - 2 normalization procedures
  - Choice projection before or after normalization
- Added information provide:
  - a better performance (except for the lowest SNRs)
  - robustness to the initialization
- The approach with residuals is less performant
- The 4 remaining variants with bound and normalization provide the same performance on these tests

Outline of the talk

Adding a physical model to inform NMF

- So far, we focused on informing NMF with knowledge on F
- The resulting approaches may be considered as in between blind separation and regression
- However, it is also possible to get some information about G
- Let us see how!

Adding a physical knowledge to inform NMF (2)

Motivation

- Particulate matter sensing
- Industrial pollutant sensing ⇔ wind directions
- NMF sensitive to initialization:
  1. multiplicative updates ➩ zero entries!
  2. possibility to initialize NMF with the output of another BSS method (Benachir *et al.*, 2013)
  3. Physical model informs NMF about active sources

  ➩ Structure used as constraints in G
- It results in an improved initialization of the matrices G and F

Adding a physical knowledge to inform NMF (2)

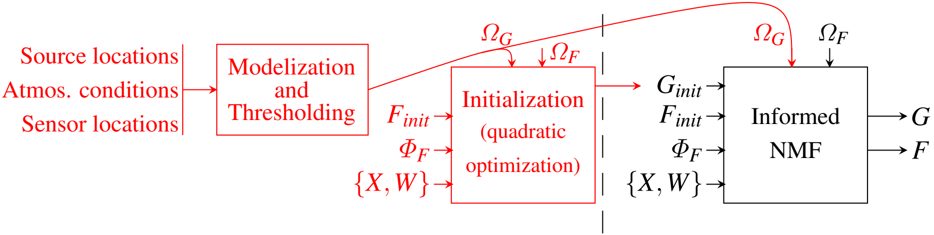
Motivation

- Particulate matter sensing
- Industrial pollutant sensing ⇔ wind directions
- NMF sensitive to initialization:
  1. multiplicative updates ➪ zero entries!
  2. possibility to initialize NMF with the output of another BSS method (Benachir *et al.*, 2013)
  3. Physical model informs NMF about active sources

  ➪ Structure used as constraints in G
- It results in an improved initialization of the matrices G and F

Adding a physical knowledge to inform NMF (3)

Going into the big picture



Physical propagation models

- Many existing physical models (Stockie, 2011)
- Classical models:
  - Computational Fluid Dynamics Model: based on discretized fluid mechanics equations
    - ✔ performant
    - ✗ very costly
  - Integral models: based on simplified fluid mechanics equations
    - ✔ less costly than the CDF model
    - ✗ not applicable for distances between sources and sensor below 10 km
    - ✗ convolutive models not well-suited with the chemical time sampling
  - Gaussian Plume model: simplest model, stationary wind and stationary source emission
    - ✔ simplest model, applicable to short and long distances
    - ✗ not really accurate

What can we do with?

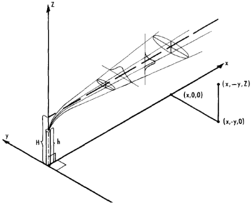
- Using a physical model for inverse problems (Ranieri *et al.*, 2012, Delmaire & Roussel, 2012)
- We propose using a model to detect some "inactive" sources

Modelization of pollutant source propagation (1)

- $\mathcal{A}$ (resp. $\overline{\mathcal{A}}$) set of the indices of the q anthropic sources (resp. $p-q$ natural sources)
- $\xi_s \triangleq (\mathrm{x}_s, \mathrm{y}_s, \mathrm{z}_s)$: location of the chemical sampler
- $\xi_l \triangleq (\mathrm{x}_i, \mathrm{y}_l, \mathrm{z}_l)$ with $1 \leq l \leq q$: location of the q industrial sites
- ξ_s' and ξ_l' in the new basis which sets the wind direction on the x axis
- transfer computed with a Gaussian plume model (Turner, 1994):

$$t(\xi_s', \xi_l', u) \triangleq \frac{1}{2\pi u \sigma_{\mathrm{y}_l'} \sigma_{\mathrm{z}_l'}} \exp\left(-\frac{(\mathrm{y}_s' - \mathrm{y}_l')^2}{2\sigma_{\mathrm{y}_l'}^2}\right) \left(\exp\left(-\frac{(\mathrm{z}_l' - \mathrm{z}_s')^2}{2\sigma_{\mathrm{z}_l'}^2}\right) + \exp\left(-\frac{(\mathrm{z}_l' + \mathrm{z}_s')^2}{2\sigma_{\mathrm{z}_l'}^2}\right) \right)$$

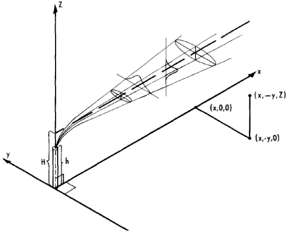
- A high transfer indicates a large contribution of the source on the measure.



Gaussian Plume (source: Turner, 1994)

Sensor & potential industrial source locations

Modelization of pollutant source propagation (2)



Gaussian Plume (source: Turner, 1995)

Sensor & potential industrial source locations

- Experimental conditions: Chemical sampling rate (T_e) vs. wind sampling rate (T_e/ν)
- For each wind measure, a vector of transfers is available.
- For a specific chemical sample, $\mathcal{T}$ is a $\nu \times q$ matrix of tranfers.

Incorporation of a special structure into NMF

- Observed data $\mathcal{X}$ collected at wind sampling rate may read

$$\mathcal{X} \triangleq \mathcal{X}^{\mathcal{A}} + \mathcal{X}^{\overline{\mathcal{A}}} \simeq \mathcal{G}^{\mathcal{A}} \cdot F^{\mathcal{A}} + \mathcal{G}^{\overline{\mathcal{A}}} \cdot F^{\overline{\mathcal{A}}} = \left[\mathcal{G}^{\mathcal{A}}, \mathcal{G}^{\overline{\mathcal{A}}}\right] \cdot \left[(F^{\mathcal{A}})^T, (F^{\overline{\mathcal{A}}})^T\right]^T,$$

where $X^{\mathcal{A}}$ and and $X^{\overline{\mathcal{A}}}$ are the $\nu \times m$ anthropic and natural parts of $\mathcal{X}$, respectively.

- $\mathcal{G}^{\mathcal{A}}$ is also a function of $\mathcal{Q}$, the $\nu \times q$ source overall flow rate.

$$\mathcal{X}^{\mathcal{A}} = (\mathcal{T} \circ \mathcal{Q}) \cdot F^{\mathcal{A}}$$

Assumption: F^I and $\mathcal{Q}$ are independent from sample number

➪ $X^{\mathcal{A}} = (T \circ Q) \cdot F^{\mathcal{A}}$, where T and Q are the $n \times m$ average transfer and flow source rate matrices, respectively, over a concentration sampling rate, and

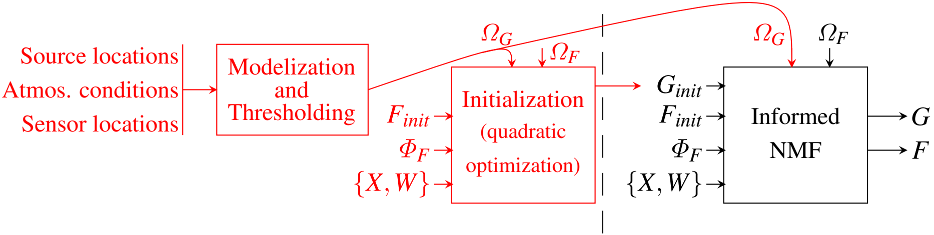
$$G^{\mathcal{A}} = (T \circ Q)$$

- Thresholding step provides a structure matrix Ω_G:

$$\Omega_{G i,j} = \begin{cases} 0 & \text{if } T_{i,j} \geq \frac{\max_j(T_{i,j})}{10^6} \text{ or } j > q & (G_{i,j} \text{ let free in the NMF}) \\ 1 & \text{otherwise} & (G_{i,j} = 0) \end{cases}$$

NMF initialization by quadratic optimization (1)

Going into the big picture



NMF initialization by quadratic optimization (2)

- Initialization of our informed NMF methods:
  - A tentative initial F is provided by chemical experts
  - G is estimated as the "best" matrice such as X is close to $G \cdot F$
  - ➪ done with Non-negative Least squares optimization:

$$\forall i \in \{1, \ldots, n\} \quad \min_{\underline{g}_i^T} \mathcal{J}(\underline{g}_i^T) = \min \left(\underline{x}_i^T - F^T \underline{g}_i^T\right)^T \cdot D_{w_i} \cdot \left(\underline{x}_i^T - F^T \underline{g}_i^T\right)$$

$$\text{such that} \begin{cases} \underline{g}_i \geq \underline{0}, \\ \sum_j g_{ij} = \sum_j x_{ij}, \\ \underline{x}_i^T \geq (\Phi^E)^T \cdot \underline{g}_i^T. \end{cases}$$

    - $\underline{x}_i^T$ is the sample measurement vector at time stamp i.
    - $\underline{g}_i$ stands for the i-th row of G
    - $D_{w_i} \triangleq \text{diag}(\underline{w}_i^T)$ is the diagonal matrix of weights asociated with $\underline{x}_i^T$

NMF initialization by quadratic optimization (2)

- Initialization of our informed NMF methods:
  - A tentative initial F is provided by chemical experts
  - G is estimated as the "best" matrice such as X is close to $G \cdot F$
  - ➪ done with Non-negative Least squares optimization:

$$\forall i \in \{1, \ldots, n\} \quad \min_{\underline{g}_i^T} \mathcal{J}(\underline{g}_i^T) = \min \left(\underline{x}_i^T - F^T \underline{g}_i^T\right)^T \cdot D_{w_i} \cdot \left(\underline{x}_i^T - F^T \underline{g}_i^T\right)$$

$$\text{such that} \begin{cases} \underline{g}_i \geq \underline{0}, \\ \sum_j g_{ij} = \sum_j x_{ij}, \\ \underline{x}_i^T \geq (\Phi^E)^T \cdot \underline{g}_i^T. \end{cases}$$

    - $\underline{x}_i^T$ is the sample measurement vector at time stamp i.
    - $\underline{g}_i$ stands for the i-th row of G
    - $D_{w_i} \triangleq \text{diag}(\underline{w}_i^T)$ is the diagonal matrix of weights asociated with $\underline{x}_i^T$

- Now, we add an extra constraint to initialize G, derived from the physical model, i.e.,

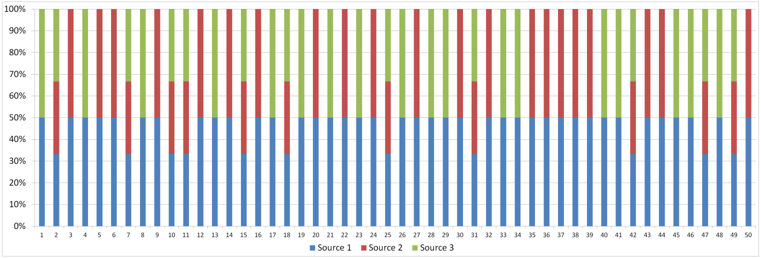
$$\forall i \in \{1, \ldots, n\} \quad \min_{\underline{g}_i^T} \mathcal{J}(\underline{g}_i^T) \quad \text{s.t.} \quad \begin{cases} \underline{g}_i \geq \underline{0}, \\ \sum_j g_{ij} = \sum_j x_{ij}, \\ \underline{x}_i^T \geq (\Phi^E)^T \cdot \underline{g}_i^T, \\ \underline{g}_i^T \circ (\Omega_G^T)_i = 0. \end{cases}$$

Performance of the model-based αβ-N_x-MCWNMFs (1)

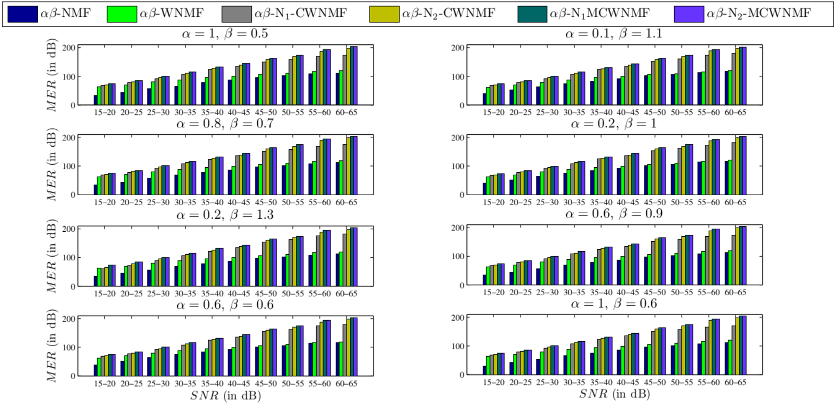
- Simulations ($n = 50$ samples, $m = 7$ chemical species, $p = 3$ sources, $q = 2$ anthropic sources)
- Additive uniform noise (keeping the data positive)
- 2 equality constraints

Fe	*Ca*	SO_4^{2-}	*Zn*	*Mg*	*Al*	*Cr*
X	X	X	60	X	X	X
X	X	X	X	X	X	20
X	X	X	X	X	X	X

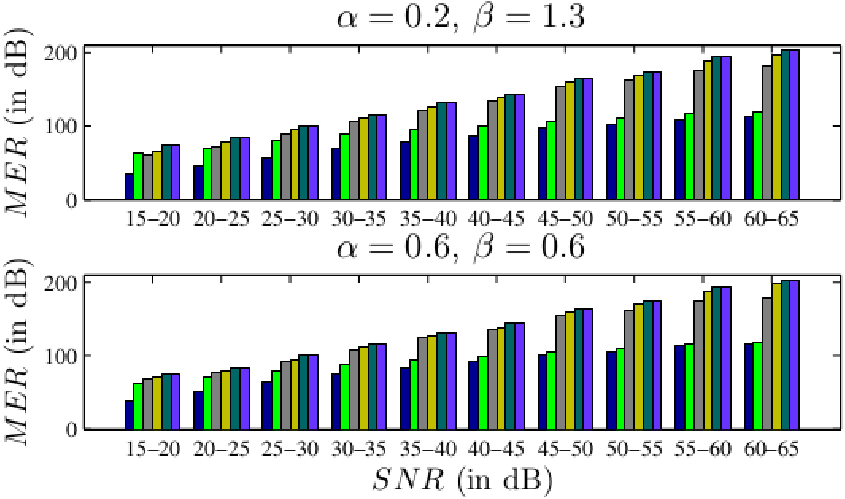
- *G* is created so that:
  - Source 1, the natural source is always contributing to the data (with contributions varying with samples)
  - Sources 2 and 3, the anthropic sources are not always sensed by the chemical sampler, depending on the wind direction



Performance of the model-based αβ-N_x-MCWNMFs (2)



Performance of the model-based αβ-N_x-MCWNMFs (2)



Conclusion of this part

- We proposed a simple yet efficient way to structure the contribution matrix G
  - controlled by a Gaussian Plume model
- Added information provide:
  - a better performance (in particular for the lowest SNRs)
  - robustness to the initialization (not shown here for time consideration)

Outline of the talk

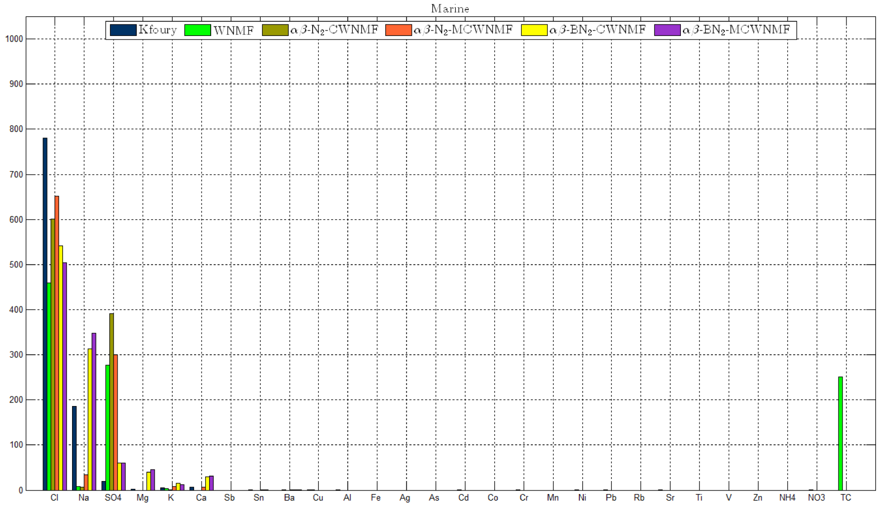
Experimental conditions (1)

- Data collected during Winter 2010 and Spring 2011 by the UCEIV Lab, under the responsability of Prof. Courcot and Dr. Ledoux
- Chemical analysis of the samples also realized by UCEIV
- $n = 164$ samples, $m = 28$ chemical species, $p = 11$ sources
- One chemical sample each $T_e = 12$ h and $\nu = 12$ wind samples per chemical sample.
- A previous Ph.D. thesis for the same data (with different tools)

Experimental conditions (2)

- $q = 6$ anthropic sources
  - 2 from the ArcelorMittal Aglomeration unit
  - 1 from another company (Ascométal)
  - 2 from traffic (one for the exhaust, one for the brake wears)
  - 1 for biomass combustion (urban heating)
- $p - q = 5$ natural sources
- 81 set constraints + 41 bound constraints
- Tested approaches:
  - Kfoury (2013 – previous work on these data)
  - $\alpha\beta$-WNMF,
  - $\alpha\beta$-N_2-CWNMF,
  - $\alpha\beta$-N_2-MCWNMF,
  - $\alpha\beta$-BN_2-CWNMF,
  - $\alpha\beta$-BN_2-MCWNMF,
- I will show you a few results:
  - Main species are on the left part of the plots
  - Minor ones are on the right part

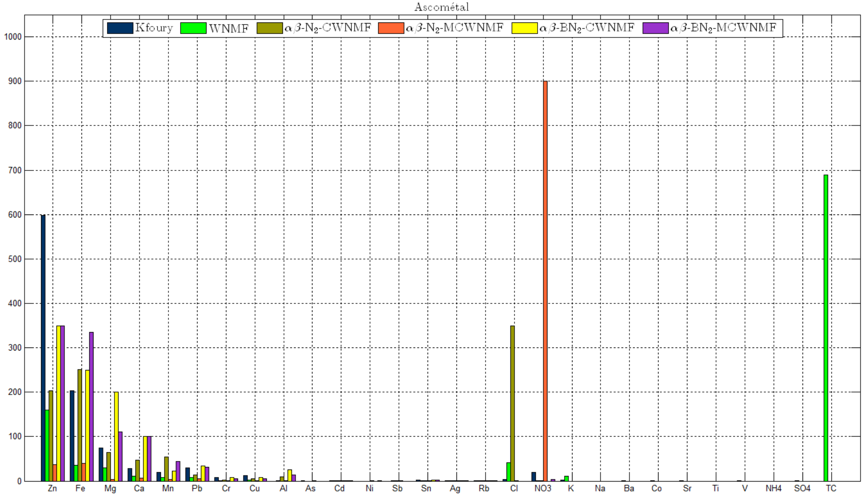
Results obtained for the Marine Source



- **Theory:** $\frac{Cl^-}{Na} \simeq 2$ and No total carbon

➩ only αβ-BN_2-CWNMF and αβ-BN_2-MCWNMF are consistent with theory

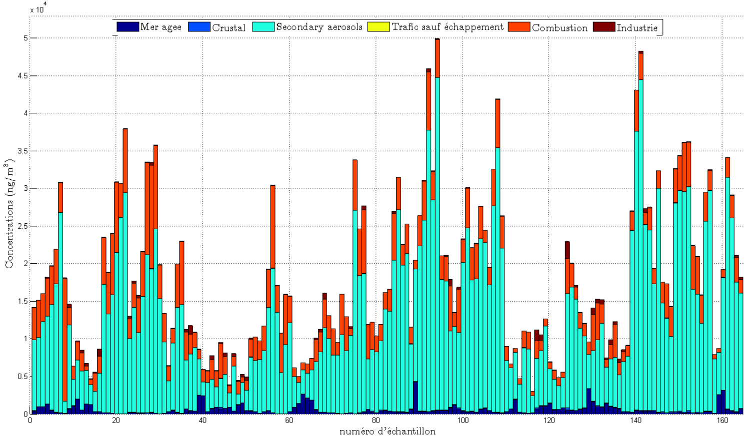
Results obtained for the Ascométal Source



- **Theory:** Mainly *Zn* in this source.
- Too much with Kfoury, too many minor species with αβ-WNMF, αβ-N_2-CWNMF and αβ-N_2-MCWNMF
- ⇨ only αβ-BN_2-CWNMF and αβ-BN_2-MCWNMF are consistent with theory

Analysis of the results

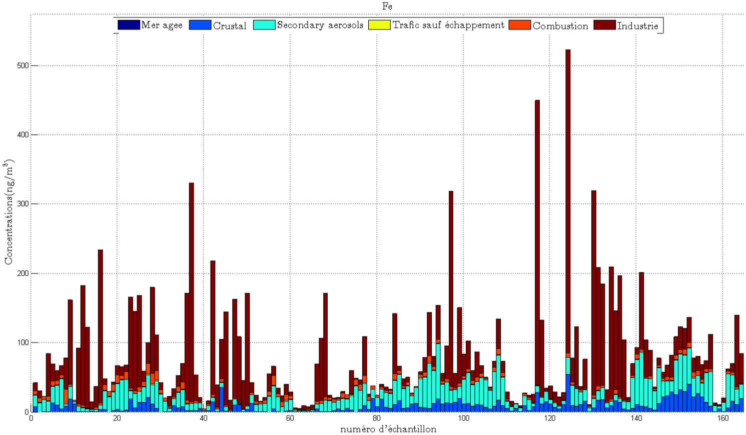
- Informed methods $\alpha\beta$-BN$_2$-CWNMF and $\alpha\beta$-BN$_2$-MCWNMF provide the best performance according to chemits' expertise
- Interest of the combination of several constraints
- We can also extract pollutant contributions from X



Source contribution in observed data

Analysis of the results

- Informed methods $\alpha\beta$-BN$_2$-CWNMF and $\alpha\beta$-BN$_2$-MCWNMF provide the best performance according to chemits' expertise
- Interest of the combination of several constraints
- We can also extract pollutant contributions from X



Fe contribution in observed data

Conclusion and future work

- We proposed several informed NMF methods
  - Information from expert's knowledge (exact or bounded values)
  - Normalization of the rows of F
  - Information from a physical propagation model
- Was shown to outperform state-of-the-art methods in the tests
- Many open problems:
  - The choice of the parameters in parametric divergences provide consistancy for major species or minor ones (but not for both at the same time)
  - ➪ update rules with parameters chosen for each species
  - Multiplicative methods are easy to implement but slow and issues of convergence
  - ➪ informing and extending modern NMF approaches (e.g., HANLS)
  - Sequential use of three kinds of constraints on F
  - ➪ new parameterization which can take into account (at least) two of them
  - Extending the physical model to the case of non-ponctual sources
  - Adding a chemical model to the physical model (chemical modification of some sources with time)

Thank you for your attention

Questions?

References I

S. Arora, R. Ge, R. Kannan, and A. Moitra.
Computing a nonnegative matrix factorization – provably.
In *Proc. of the 44th symp. on Theory of Computing*, pages 145–162, 2012.

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*, Gainesville, Florida, USA, June 2013.

A. Cichocki, S. Cruces, and S. Amari.
Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization.
Entropy, 13:134–170, 2011.

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.

D. Donoho and V. Stodden.
When does non-negative matrix factorization give a correct decomposition into parts?
In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems 16*, page None. MIT Press, Cambridge, MA, 2003.

C. Févotte and J. Idier.
Algorithms for nonnegative matrix factorization with the beta-divergence.
Neural computation, 23(9):2421–2456, September 2011.

N. Gillis.
Sparse and unique nonnegative matrix factorization through data preprocessing.
Journal of Machine Learning Research, 13:3349–3386, November 2012.

References II

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

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

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

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

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

A. Kfoury.
Origin and physicochemical behaviour of atmospheric PM2.5 in cities located in the littoral area of the Nord-Pas-de-Calais region, France.
Phd thesis, Université du Littoral Côte d'Opale, 2013.

H. Kim and 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):713–730, 2008.

References III

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

H. Lantéri, C. Theys, C. Richard, and C. Févotte.
Split gradient method for nonnegative matrix factorization.
In *Proc. of EUSIPCO*, 2010.

D. Lee and H. Seung.
Learning the parts of objects by non negative matrix factorization.
Nature, 401(6755):788–791, 1999.

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.

A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot.
Non-negative matrix factorization under equality constraints—a study of industrial source identification.
Applied Numerical Mathematics, 85:1–15, November 2014.

A. Limem, G. Delmaire, G. Roussel, and D. Courcot.
Kullback-Leibler NMF under linear equality constraints. application to pollution source apportionment.
In *Proc. of ISSPA*, pages 752–757, 2012.

A. Limem, M. Puigt, G. Delmaire, G. Roussel, and D. Courcot.
Bound constrained weighted nmf for industrial source apportionment.
In *Proc. of MLSP*, Reims, France, September 21–24, 2014.

References IV

C.-J. Lin.
Projected gradients methods for non-negative matrix factorization.
Neural Computation, 19:2756–2779, 2007.

L. Miao and H. Qi.
Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization.
IEEE Trans. Geoscience and Remote Sensing, 45(3):765–777, 2007.

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

M. Plouvin, A. Limem, M. Puigt, G. Delmaire, G. Roussel, and D. Courcot.
Enhanced NMF initialization using a physical model for pollution source apportionment.
In *Proc. of ESANN*, pages 261–266, 2014.

M. D. Plumbley.
Conditions for nonnegative independent component analysis.
Signal Processing Letters, IEEE, 9(6):177–180, 2002.

M. D. Plumbley.
Algorithms for nonnegative independent component analysis.
IEEE Transactions on Neural Networks, 14(3):534–543, 2003.

M. D. Plumbley and E. Oja.
A "nonnegative PCA" algorithm for independent component analysis.
IEEE Transactions on Neural Networks, 15(1):66–76, 2004.

References V

J. Ranieri, Y. Lu, A. Chebira, and M. Vetterli.
Sampling and reconstructing diffusion fields with localized sources.
In *Proc. of ICASSP*, Prague, Czech Republic, 2011.

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

B. Turner.
Workbook of Atmospheric Dispersion Estimates: An Introduction to Dispersion Modeling, Second Edition, chapter Estimates of atmospheric dispersion.
Lewis Publishers, 1994.

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

M. Viana, T. A. J. Kuhlbusch, X. Querol, and 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):827–849, 2008.

J. Yoo and S. Choi.
Nonnegative matrix factorization with orthogonality constraints.
Journal of computing science and engineering, 4(2):97–109, June 2010.