



Université Lille Nord de France
Pôle de Recherche
et d'Enseignement Supérieur

Ecole doctorale régionale Sciences Pour l'Ingénieur Lille Nord-de-France - 072



Titre : *Méthodes de NMF Bayésiennes informées pour la séparation de sources*

Financement prévu : Université Libanaise ou ARCUS

Cofinancement éventuel :

(Co)-Directeur de thèse : Gilles ROUSSEL

E-mail : gilles.rousseau@univ-littoral.fr

Encadrants : Gilles Delmaire

E-mail : gilles.delmaire@univ-littoral.fr

Laboratoire : LISIC (Laboratoire d'Informatique Signal et Image de la Côte d'Opale, EA 4491)

Equipe : Specifi

Descriptif :

Le démixage de sources pour la pollution de l'air peut être formulé comme un problème de NMF en décomposant la matrice d'observation X en le produit de deux matrices non négatives G et F , respectivement la matrice de contributions et de profils. Généralement, les données chimiques sont entachées d'une part de données aberrantes. En dépit de l'intérêt de la communauté pour les méthodes de NMF, elles souffrent d'un manque de robustesse à un faible nombre de données aberrantes et aux conditions initiales et elles fournissent habituellement de multiples minimas.

En rupture par rapport aux méthodes traditionnelles de factorisations matricielles cette thèse est orientée d'une part vers les NMF informées utilisant les connaissances expertes sous l'angle global des méthodes Bayésiennes, d'autre part en ayant à l'esprit le fait de rendre la NMF robuste même lorsque le nombre d'échantillons est relativement faible.

Nous proposons donc dans le cadre de cette thèse d'étendre ces notions de factorisations collectives à des situations où la matrice des sources contient uniquement une partie commune, conduisant à considérer des couplages plus flexibles entre les différentes matrices à co-factoriser tout en permettant l'ajout de connaissances expertes telles que proposées dans nos précédentes approches de NMF informée [1-4]. Les approches envisagées utiliseront la statistique Bayésienne [5] afin de mieux cerner les profils a posteriori.

D'un point de vue applicatif, ces méthodes sont appliquées à la pollution de l'air dans l'environnement marin et terrestre. L'enjeu principal consiste en un diagnostic fin de la matrice des sources de pollution. Cette matrice rassemble les signatures chimiques des différentes sources en présence. L'apport de la nouvelle méthodologie permettra de prendre en compte plusieurs sites urbains ou ruraux de manière simultanée en considérant que certaines sources sont communes à l'ensemble des sites.

Du point de vue applicatif, ce projet vise à étudier la contribution des différentes industries sur différents sites de topologie différentes. L'exploitation simultanée des deux sites de mesure (ou plus) nécessitera donc la mise en œuvre de nouvelles méthodes dites de co-factorisation.



Université Lille Nord de France
Pôle de Recherche
et d'Enseignement Supérieur

Fort de notre expérience dans le Nord-pas de Calais, nous souhaitons étendre notre expertise sur des sites de topologies différentes au Liban, situé au nord de Beyrouth. Ces deux sites pourront être traités ensemble pour séparer l'influence des différentes sources.

Mots clés : Factorisation matricielle ou tensorielle, modèle explicatif, toxicologie.