



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 : Modélisation explicative des données par factorisation en fonctionnelles. Application à la séparation des déterminants d'effets toxiques de particules atmosphériques.

Financement prévu : PMCO

Cofinancement éventuel : ULCO

(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 : L'objectif de cette étude est d'élaborer et d'évaluer le potentiel d'une utilisation particulière de la factorisation matricielle ou tensorielle, permettant d'établir des relations de causes à effets entre déterminants et les effets mesurés par les paramètres de toxicité, et ce par une méthode non descriptive.

Nous proposons ici une nouvelle approche de la factorisation matricielle pour l'analyse des données et la **construction d'un modèle explicatif en séparant les déterminants qui influent (linéairement et/ou non linéairement) sur les données observées**. Lorsque des hypothèses de séparabilité sont établies, il convient d'estimer une matrice F contenant une famille de noyaux fonctionnels générateurs, non linéaires, potentiellement dépendant des déterminants, et une matrice G contenant les facteurs d'incidence appliqués aux atomes générateurs. En suivant cette approche, l'objectif est alors d'estimer les deux matrices par une factorisation semi-non négative (les effets cocktail de certains réactifs se compensent ou se renforcent) à partir d'un algorithme itératif, où il s'agit de minimiser une divergence entre la matrice des données X et le produit GF associé à une régularisation favorisant la parcimonie de la matrice G . L'objectif de parcimonie de G vise à minimiser le nombre de déterminants pertinents. Cette approche sera mise en perspective par rapport à d'autres méthodes plus classiques en *machine learning*.

L'application concerne l'étude toxicologique par l'évaluation des déterminants de la toxicité pulmonaire des particules fines et ultrafines. Généralement, la susceptibilité des cellules pulmonaires aux éléments perturbateurs comme les particules est particulièrement difficile à expliquer car la composition de celles-ci est chimiquement complexe, et la manière dont ces particules atteignent les cellules (fraction bio-accessible), la teneur en métaux ou en composés organiques, les propriétés physiques (i.e. granulométrie) sont autant de facteurs d'influence dont il est difficile d'identifier la causalité [18]. Dans le cadre de ses travaux, le laboratoire UCEIV met en œuvre des protocoles expérimentaux afin d'échantillonner des observations issues de différents sites caractéristiques d'exposition aux particules et d'évaluer les effets toxicologiques sur les cellules pulmonaires (paramètres inflammatoires, stress oxydant, quantification de la métabolisation, génotoxicité, ...).

Le modèle explicatif ainsi construit pourra être ensuite associé à une étape préalable de séparation de



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

sources de pollution en particules des sources présentent dans un environnement donné (cette étape est déjà maîtrisée par des algorithmes développés au laboratoire) et permettrait donc de prédire à terme une partie des effets toxicologiques du cocktail de sources en présence dans cet environnement sur la population proche

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