

Stage de Master 2 – Data science

Données spatio-temporelles de concentrations horaires de 8 polluants atmosphériques en France métropolitaine : quels modèles d'exploration et de synthèse ?

Centre Léon Bérard & EA 3083 ERIC

Contexte de la mission :

En France, l'étude menée en 2016 par Santé Publique France a estimé à 48 000 le nombre de décès prématurés attribuables à la pollution de l'air (Pascal, 2016). Le Centre International de Recherche sur le Cancer (CIRC) a classé la pollution de l'air extérieur dans son ensemble et les particules fines en particulier, comme cancérogènes pour l'homme, avec des indications suffisantes pour un lien avec le cancer du poumon et des associations positives avec le cancer de la vessie (IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, 2012).

C'est dans ce contexte que travaille depuis des années le Département Prévention Cancer Environnement du Centre Léon Bérard avec différents partenaires (e.g. Laboratoire de Mécanique des Fluides et d'Acoustique (UMR CNRS 5509 LMFA – École Centrale), Laboratoire Environnement Ville Société (UMR CNRS 5600 EVS – Lyon2/Lyon3) ou encore le Laboratoire d'Informatique (UMR CNRS 5205 LIRIS – Lyon2), l'INERIS ou encore Santé Publique France) sur des études épidémiologiques sur les facteurs de risques environnementaux liés à différents cancers hormonaux (e.g. sein, testicule, etc.).

À titre d'illustration, pour le cancer du sein, le plus fréquent chez la femme, les causes environnementales et notamment l'exposition chronique à la pollution de l'air sont suspectées d'avoir un impact sur son incidence. En effet le dépistage, le vieillissement de la population et les facteurs de risque liés au mode de vie (alcool, alimentation, etc.) ne permettent pas d'expliquer à eux seuls les variations géographiques des taux d'incidence, et incitent, dès lors, à investiguer ces causes environnementales.

Dans le cadre d'une étude sur le risque de cancer du sein associée à l'exposition à la pollution de l'air, une simulation des concentrations de 8 polluants atmosphérique sur le territoire français a été réalisée via le modèle CHIMERE (modèle chimie-transport développé par l'INERIS). Nous disposons donc des concentrations horaires pour 8 polluants pour une résolution fine ($0,0625^\circ \times 0,125^\circ$ soit environ 8×8 km) sur 21 ans 1990-2010.

Ce jeu de données permet d'analyser les expositions individuelles tout au long de la période pour les sujets d'études. La précision temporelle et la couverture spatiale permettent l'étude individuelle sur la période d'étude en suivant les parcours de vie : succession des lieux de résidence et des lieux de travail mais aussi les mobilités quotidiennes et prochainement l'ensemble des autres activités quotidiennes (loisirs, visites, etc.) ainsi que des mobilités induites.

Néanmoins, ces données demeurent pour l'instant sous-exploitées notamment quant aux modèles spatio-temporels intrinsèques particulièrement aux niveaux les plus fins.

Dans ce contexte, le stage a pour objectif de contribuer, au sein d'un partenariat inédit entre le Centre Léon Bérard et le laboratoire ERIC, spécialisé en exploration de données temporelles, d'appréhender les modèles temporels et spatiaux sous-jacents permettant la synthèse et dès lors l'exploitation routinière de ces données massives dans le cadre des présents et futurs projets épidémiologiques.

Descriptif de la mission :

Vous serez en charge de la mise en place d'une méthodologie pour explorer des données spatio-temporelles complexes dans le but de proposer une/des synthèses de ces données singulières ou plurielles. Les données présentent plusieurs caractéristiques d'intérêt, ce sont de données multivariées avec une structure spatio-temporelle et la présence de phénomènes extrêmes. Il est envisagé que la personne recrutée exploite le plus grand nombre de ces caractéristiques dans les analyses. Une liste non exhaustive de missions du stage suit :

- Exploration statique des données horaires infra- et inter-polluant pour la détermination de profils types de pollution
- Analyse spatiale des polluants afin de déterminer les profils d'interactions
- Modélisation de la structure spatio-temporelle (par exemple en utilisant modèles additifs ou de données fonctionnelles)
- Caractériser les épisodes de concentration liés aux interactions de plusieurs polluants

Profil recherché :

- Étudiant·e de M2 à orientation data science
- Rigueur scientifique, capacité d'innovation et de synthèse.
- Capacité à lire et comprendre la littérature internationale (maîtrise de l'anglais scientifique)
- Maîtrise d'au moins un logiciel d'analyse de données (R, Python)
- Autonomie, bonnes qualités relationnelles et capacité à travailler en équipe pluridisciplinaire

Durée du stage : 6 mois, à partir de mars 2021 (ajustables en fonction du planning universitaire de la personne retenue)

Gratification : selon les règles en vigueur

Encadrement :

- Jairo Cugliari – MCU Mathématiques – Laboratoire ERIC / Université Lyon 2
- Lény Grassot – PhD Géomatique – Centre Léon Bérard, Département Prévention Cancer Environnement / INSERM UA8 Radiations : Défense, Santé et Environnement
- Thomas Coudon – PhD Expologie environnementale – Centre Léon Bérard, Département Prévention Cancer Environnement / INSERM UA8 Radiations : Défense, Santé et Environnement / CNRS UMR 5509 LMFA – École Centrale de Lyon

Lieu du stage : Université Lumière Lyon 2 (Bron) et Département Prévention Cancer Environnement, Centre Léon Bérard (Lyon 8ème).

Envoi des candidatures / Renseignements :

Curriculum Vitae et Lettre de motivation avant le 10 janvier 2021

A Jairo.Cugliari@univ-lyon2.fr & Leny.Grassot@lyon.unicancer.fr