

Offre de stage de Master 2 année 2025

Equipes :

Statistique pour le Vivant et l'Homme : <https://svh.imag.fr/>

Epidémiologie Environnementale : <https://iab-grenoble.fr/fr/recherche/equipes/epidemiologie-environnementale-appliquee-au-developpement-et-a-la-sante-respiratoire-edes>

Encadrant·e·s :

Adeline Leclercq-Samson, Vincent Brault, Johanna Lepeule

Lieu : Laboratoire Jean Kuntzmann, 150 place du Torrent, 38400 Saint Martin d'Hères

Durée : 5 à 6 mois

[English version below](#)

Titre : Analyse de l'association longitudinale entre l'exposition aux facteurs environnementaux et le développement de l'enfant

Contexte

La santé humaine peut être affectée par de nombreuses expositions environnementales, parmi lesquelles un grand nombre de preuves ont démontré l'impact de la pollution de l'air ambiant, des températures ambiantes élevées et des vagues de chaleur sur la morbidité et la mortalité (Schwartz, 1994). Les expositions au début de la vie sont particulièrement préoccupantes : la vie intra-utérine est une période critique de plasticité au cours de laquelle les agressions environnementales peuvent modifier les programmes de développement avec des effets immédiats visibles à la naissance ou des effets différés qui apparaissent pendant l'enfance, à la puberté ou à l'âge adulte. Ces travaux sont liés au domaine de recherche DOHaD (Developmental Origins of Health and Diseases), qui affirme que plusieurs maladies fréquentes peuvent avoir une origine pendant la période de développement (Sinclair et al. 2007). Chez les animaux, un large corpus de recherche soutient cette hypothèse (Hanson and Gluckman 2011) y compris pour la pollution de l'air (Veras et al. 2009) et l'exposition à la chaleur pendant la gestation (van Wettere et al. 2021).

Chez l'homme, des études épidémiologiques ont démontré les effets néfastes de la pollution atmosphérique et de la chaleur sur l'issue de multiples naissances (par exemple, diminution du poids à la naissance, augmentation des naissances prématurées, mortes et néonatales) (Jakpor et al. 2020; Chersich et al. 2020; Song et al. 2023). Plus tard dans la vie, le faible poids du bébé à la naissance contribue de manière significative à la morbidité et à la mortalité périnatales (Pallotto and Kilbride 2006). Au-delà de la période néonatale, le retard de croissance intra-utérin peut avoir des ramifications tout au long de la vie en augmentant le risque de plusieurs maladies cardiovasculaires, respiratoires et métaboliques (Fowden et al. 2006).

Du point de vue de l'exposition, les études examinent de plus en plus l'ensemble de la trajectoire d'exposition d'un individu, en tenant compte de chaque semaine, de la conception à la naissance (Gasparrini et al. 2010; Jakpor et al. 2020), en tirant parti de modèles d'exposition de pointe dotés d'une résolution spatiale et temporelle fine (Hough et al. 2020; Hough et al. 2021). Du point de vue du développement, le poids de naissance est une mesure a posteriori qui ne reflète pas nécessairement les différentes trajectoires

de croissance au cours de la vie fœtale. Certaines études ont examiné la contribution de l'échographie à la détection d'éventuels effets précoces et de phénomènes de croissance de rattrapage (Iñiguez et al. 2016). En outre, l'identification des fenêtres critiques d'exposition à l'environnement pendant la grossesse présente un intérêt considérable pour mieux prévenir les effets sur la santé du fœtus. Cependant, aucune étude n'a évalué les effets de l'ensemble de la trajectoire d'exposition sur la trajectoire de croissance du fœtus, en tenant compte des aspects longitudinaux de l'exposition et du résultat.

Objectifs

L'objectif du stage de Master 2 est de développer des modèles statistiques et des méthodes d'estimation prenant en compte la nature longitudinale des données d'exposition et de résultats. La première étape consistera à revoir les packages existants `dlnm`, `nlme` et à identifier leurs limites. La seconde étape consistera à développer un nouveau package incluant l'extension aux effets aléatoires et aux interactions entre variables.

Ce projet de Master 2 est la première étape d'un projet de doctorat plus général visant à produire de nouvelles connaissances sur les effets de la pollution de l'air et de l'exposition à la chaleur pendant la grossesse sur la croissance du fœtus (dans un contexte de données répétées à la fois pour l'exposition et le résultat).

Methodes

Les méthodes statistiques incluront des modèles de lag distribué (DL). Ces modèles de séries temporelles incluent des covariables qui dépendent des effets passés et différés des covariables. Par exemple, l'espérance μ_t de la variable de croissance Y_t au temps t est modélisée comme une fonction (non linéaire) des covariables $x_t, x_{t-1}, \dots, x_{t-L}$:

$$g(\mu_t) = \alpha + \sum_{k=0}^L \beta_k s(x_{t-k})$$

où s est une fonction de spline.

Ces modèles DL permettent de quantifier l'effet d'une exposition différée sur les résultats. Dans la littérature, les modèles DL ont été appliqués pour décrire l'effet de la pollution de l'air et de l'exposition à la chaleur sur le poids de naissance.

Au lieu de se concentrer sur la mesure a posteriori du poids de naissance, nous envisageons d'étendre cette approche à la modélisation de l'ensemble de la trajectoire de croissance du fœtus en utilisant les données d'une cohorte. Le modèle correspondant inclura donc des effets aléatoires pour décrire la variabilité interindividuelle entre les sujets de la cohorte. Une première étape consistera à proposer une méthode d'estimation des paramètres de ce modèle longitudinal mixte. Une approche de type maximum de vraisemblance sera envisagée, avec l'adaptation d'un algorithme de type EM (Dempster, 1977) à cette classe spécifique de modèles mixtes, par exemple un algorithme SAEM adapté aux modèles à effets mixtes non linéaires (Kuhn, 2005).

La deuxième étape consistera à proposer de nouvelles façons d'inclure les effets de l'exposition dans le modèle, afin de décrire simultanément l'exposition aiguë (jours) et l'exposition chronique (semaines, mois) à la chaleur ou à la pollution de l'air. Ces différentes échelles pourraient être modélisées par différentes fonctions splines, en sélectionnant les covariables significatives par la sélection de modèles ou en les intégrant dans le temps.

Enfin, les covariables telles que la pollution atmosphérique sont des covariables spatiales et ne sont pas faciles à introduire dans le modèle de régression. Des outils spécifiques tels que la fonction d'intensité pourraient être utilisés pour résumer l'exposition et servir d'indicateur dans le modèle DL.

Compétences requises

- Quelques notions dans les méthodes statistiques avancées
- Maîtrise de R, en particulier dans le développement de packages.
- Solides compétences organisationnelles et souci du détail
- Intérêt pour l'épidémiologie et la santé environnementale

Candidature : Le CV et la lettre de motivation (y compris les perspectives de carrière) doivent être envoyés aux adresses suivantes adeline.leclercq-samson@univ-grenoble-alpes.fr, vincent.brault@univ-grenoble-alpes.fr, johanna.lepeule@univ-grenoble-alpes.fr

Ces documents peuvent être rédigés en français ou en anglais.

English version

Title: The longitudinal association between exposure to environmental factors and development of the child

Background

Human health can be affected by many environmental exposures as, among them, a large body of evidence has demonstrated the impact of ambient air pollution and high ambient temperatures and heatwaves on morbidity and mortality (Schwartz, 1994). Early life exposures are of particular concern: intrauterine life is a critical period of plasticity during which environmental insults can alter developmental programs with immediate effects visible at birth or delayed effects that appear in childhood, at puberty or in adulthood. Such work relates to the DOHaD (Developmental Origins of Health and Diseases) research field, which states that several frequent diseases may have an origin during the developmental period (Sinclair et al. 2007). In animals, a wide research corpus supports this hypothesis (Hanson and Gluckman 2011) including for air pollution (Veras et al. 2009) and heat exposure during gestation (van Wettere et al. 2021).

In humans, epidemiological studies demonstrated adverse effects of air pollution and heat on multiple birth outcomes (e.g. decreased birth weight, increase in premature, still- and neonatal births) (Jakpor et al. 2020; Chersich et al. 2020; Song et al. 2023). Later on in life, lower birth weight of the baby contributes significantly toward perinatal morbidity and mortality (Pallotto and Kilbride 2006). Even beyond the neonatal period, intra uterine growth retardation can have ramifications throughout the entire life course by increasing the risk of several cardiovascular, respiratory and metabolic diseases (Fowden et al. 2006).

From an exposure point of view, studies increasingly examine the whole trajectory of exposure of an individual, looking at each week from conception to birth (Gasparrini et al. 2010; Jakpor et al. 2020), taking advantage of state-of-the-art exposure models with fine spatial and temporal resolution (Hough et al. 2020; Hough et al. 2021). From a developmental point of view, birth weight is an a posteriori measurement that does not necessarily reflect the different growth trajectories during fetal life. Some studies have looked at the contribution of ultrasound to detect possible early effects and catch-up growth phenomena (Iñiguez et al. 2016). Besides, there is a huge interest in identifying the critical exposure windows to environmental

exposure during pregnancy to better prevent health effects on the fetus. However, no study has evaluated the effects of the whole trajectory of exposure on the fetal growth trajectory, accounting for the longitudinal aspects of both the exposure and the outcome.

Objectives

The objectives of the Master 2 internship is to develop statistical models and estimation methods accounting for both the longitudinal nature of exposure and outcome data. The first step will consist in reviewing the existing packages `dlnm`, `nlme` and identify their limits. The second step will then consist in developing a new package including the extension to random effects, interaction between variables.

This Master 2 project is the first step of a more general PhD project aiming at producing new knowledge on the effects of air pollution and heat exposure during pregnancy on the growth of the fetus (in a context of repeated data for both the exposure and the outcome).

Methods

Statistical methods will include Distributed Lag (DL) models. These models of time series include covariates that depend on the past and delayed effects of covariates. For example, the expectation μ_t of the growth observation Y_t at time t is modeled as a (non-linear) function of delayed covariates of exposure $x_t, x_{t-1}, \dots, x_{t-L}$:

$$g(\mu_t) = \alpha + \sum_{k=0}^L \beta_k s(x_{t-k})$$

where s is a spline function.

These DL models allow to quantify the effect of delayed exposure on the output. In the literature, DL have been applied to describe the effect of air pollution and heat exposure on birth weight.

Instead of focusing on the a posteriori measurement of birth weight, we plan to extend this approach to modeling the whole fetal growth trajectory using cohort data. The corresponding model will thus include random effects to describe the inter-individual variability between the subjects of the cohort. A first step will be to propose an estimation method of the parameters of this mixed longitudinal model. A maximum likelihood approach will be considered, with the adaptation of an EM-type algorithm (Dempster et al 1977) to this specific class of mixed models, for example a SAEM algorithm adapted to non-linear mixed effects models (Kuhn Lavielle 2005).

The second step will be to propose new ways of including exposure effects in the model, to describe simultaneously acute (days) exposure as well as chronic (weeks, months) exposure to heat or air pollution. These different scales could be modeled either by different spline functions, by selecting the significant covariates by model selection or by integrating them over time.

Finally, covariates such as air pollution are spatial covariates and are not easy to introduce in the regression model. Specific tools such as intensity function could be used to summarize the exposure and used as a proxy in the DL model.

Key competences needed

- Familiarity with advanced statistical methods
- Proficient in R, specifically in the development of packages
- Strong organizational skills and attention to detail

Inserm

La science pour la santé
From science to health



LABORATOIRE
JEAN KUNTZMANN
MATHÉMATIQUES APPLIQUÉES - INFORMATIQUE

- Interest in epidemiology and environmental health

Application: CV and cover letter (including career perspectives) should be sent to adeline.leclercq-samson@univ-grenoble-alpes.fr, vincent.brault@univ-grenoble-alpes.fr, johanna.lepeule@univ-grenoble-alpes.fr

Those documents can be written either in French or English.