

# Étude de distributions spatiales régionalisées de contaminants atmosphériques bio-accumulés par des mousses végétales

Thèse de doctorat en statistique appliquée à l'écologie

Encadrants :

- Isabelle Albert (INRAE, MIA Paris)
- Sébastien Leblond (UMS 2006 PatriNat OFB/CNRS/MNHN)
- Caroline Meyer (UMS 2006 PatriNat OFB/CNRS/MNHN)

Date de début : 01/12/2021

Laboratoire de rattachement : MIA Paris, Unité Mixte de recherche INRAE et AgroParisTech - Université Paris Saclay, <https://ww6.inrae.fr/mia-paris>

Contact : [isabelle.albert@inrae.fr](mailto:isabelle.albert@inrae.fr) ; [sebastien.leblond@mnhn.fr](mailto:sebastien.leblond@mnhn.fr) ; [caroline.meyer@mnhn.fr](mailto:caroline.meyer@mnhn.fr)

Compétences attendues : Solide formation en statistique ; Intérêt marqué pour l'écologie ; Compétence en programmation R ; Anglais scientifique.

**Résumé** Cette thèse s'intéresse à la production de cartes de distribution spatiale de retombées atmosphériques métalliques et azotées par l'analyse d'un organisme bio-accumulateur, la mousse végétale. De nombreuses études ont montré que les mousses étaient de bons capteurs des polluants environnementaux. Cependant l'étude des retombées atmosphériques à large échelle s'avère complexe, la France notamment étant divisée en zones biogéographiques qui peuvent influencer sur les concentrations accumulées. À partir des valeurs de concentrations en éléments métalliques et azotés dans des mousses, nous produisons des cartes regroupant les informations régionalisées apportées par les différents liens entre mousses et éléments contextuels : concentration du contaminant dans l'air (EMEP, RENECOFOR, MERA, etc.) ou le sol (réseau RMQS), altitude, type de mousse, etc.) avec des modèles spatiaux hiérarchiques, où les coefficients reliant concentration dans les mousses et éléments contextuels, varieront d'un territoire à l'autre. Différents champs spatiaux latents seront étudiés pour ajuster au mieux les données. La flexibilité du modèle permettra de rassembler sur une même carte les quantifications dans des mousses ayant des liens très divers avec leur environnement. La mise en œuvre du modèle sur la France entière (plus de 400 sites de prélèvements ruraux) et sa déclinaison à plus d'une vingtaine de contaminants nécessiteront des choix pertinents d'algorithmes pour l'estimation des paramètres et de leur incertitude associée.

**Enjeux scientifiques et socio-économiques** Enjeux scientifiques : développement de modèles spatiaux statistiques qui répondront à des problèmes d'effets régionalisés avec des données de grande dimension (en 2016 et 2021 : 445 sites de prélèvements, 26 éléments analysés sur chaque site).

Enjeux socio-économiques : produire des cartes de concentrations en différents éléments à l'échelle de la France métropolitaine, exploitables par les politiques publiques et le domaine de la santé.

**État de l'art scientifique** Les derniers travaux de modélisation faits sur des mousses urbaines et rurales par les équipes impliquées dans cette thèse ont montré les liens régionalisés entre les concentrations en cadmium dans les mousses et leurs concentrations atmosphériques (Van Honacker et al., 2018 ; Vieille et al., 2021). Il apparaît donc pertinent d'étudier cette relation plus en profondeur dans un travail de thèse pour permettre d'obtenir des cartes de concentrations sur l'ensemble du territoire métropolitain. La prise en compte de l'hétérogénéité spatiale du territoire dans la représentation des données et dans la modélisation des résultats est un enjeu crucial pour l'utilisation de ces données par les politiques publiques dans le cadre de suivis d'effets des réglementations, et par des chercheurs en écologie pour des suivis d'impacts sur les écosystèmes ou en épidémiologie pour des liens avec la mortalité ou différentes pathologies (Meyer et al., 2020 ; Lequy et al., 2019). Cette thèse de modélisation statistique s'intéressera donc à la modélisation spatiale (Cressie, 2015) des éléments métalliques et azotés mesurés dans les mousses à travers des modèles à coefficients de régression spatialisés, autrement dit à une modélisation hiérarchique du processus spatial avec un champ latent spatialisé qui gèrera la covariation spatiale de l'effet d'une ou plusieurs covariables explicatives de la contamination. Des premiers modèles de cette nature ont déjà été proposés par (Banerjee et al., 2014) mais devront être ici étendus pour prendre en compte la spécificité de la question biologique et l'importance du jeu de données, aspects non encore publiés.

**Question de recherche** Les premiers modèles développés précédemment devront être étendus pour prendre en compte des champs spatiaux latents non encore étudiés (Exponentiel, Matérn, etc.), avec ou sans pépite, et être adaptés à des champs spatiaux latents dépendant eux aussi de covariables. Le sujet nécessitera également la confrontation à des questions computationnelles car les matrices qui devront être inversées pour l'inférence des paramètres des modèles seront de grande dimension, des centaines de lignes et colonnes. Il faudra alors envisager des algorithmes d'inférence bayésiennes approchée ou bayésienne avec approximation, marginale (type bayésien variationnel) ou au niveau hiérarchique (type INLA), ou encore bayésienne via des algorithmes Hamiltonian MCMC.

**Matériel nécessaire et méthodes envisagées** Les données : le dispositif BRAMM (Biosurveillance des Retombées Atmosphériques Métalliques par les Mousses) est la participation française au programme européen de suivi des métaux dans les mousses (UNECE-LRATP). La France participe à ce programme européen depuis 1996 (<http://bramm.mnhn.fr/>). Les sites de prélèvement des mousses sont sélectionnés pour être représentatifs du niveau de fond, c'est-à-dire a priori non soumis aux sources locales de contamination. Il s'agit de sites ruraux, à l'exception de quelques sites périurbains en périphérie des grandes agglomérations (Paris) et en milieu forestier (sous couvert arboré) répartis de façon homogène sur l'ensemble du territoire métropolitain. Le nombre de sites échantillonnés, lors des différentes campagnes françaises, est de 511 sites en 1996, de 528 sites en 2000, de 536 en 2006, 449 en 2011 et de 445 en 2016 et 2021. Lors des deux dernières campagnes, vingt-six éléments (Al, As, C, Ca, Cd, Co, Cr, Cu, Fe, Hg, K, Mg, Mn, Mo, N, Na, Ni, P, Pb, Pd, Pt, Rh, S, Sb, V, Zn) ont été analysés. Les méthodes : nous envisageons d'abord tout un travail de modélisation qui devra aboutir à un modèle original prenant en compte des effets régionalisés pour l'analyse spatiale des polluants dans les mousses. Ensuite, nous devons résoudre des questions computationnelles pour l'inférence statistique comme évoquées précédemment.

**Programme de recherches** 1. Écriture de modèles pertinents ; 2. Inférence de ces modèles à partir des données (études de différentes techniques d'inférence possible) ; 3. Choix du modèle à partir

de critère statistique et biologique (quelles covariables sont pertinentes ? Quelles structures spatiales sur les champs latents spatialisés ?, etc.) ; 4. Production de cartes sur la France entière.

**Calendrier** 1ère année : étude bibliographique de la modélisation hiérarchique pour les données spatialisées et appropriation des bases de données « Mousses » et données modélisées existantes pour quelques éléments ; 2ème année : écriture et inférence des modèles pertinents. Sélection de modèle. Validation et prédiction ; 3ème année : production des cartes et communication des principaux résultats de la thèse.

**Publications envisageables** 1. Publication dans une revue de sciences environnementales des liens mis en évidence entre mesure sur un végétal (mousse) et pollution de l'air ; 2. Publication statistique sur la modélisation hiérarchique du processus spatial ; 3. Publication dans une revue d'écologie des cartes des polluants obtenues par les modèles mis en œuvre et discussion de leur intérêt.

**Compétences cognitives et techniques acquises par le doctorant en fin de thèse** Connaissance de la modélisation statistique avancée sur données spatiales. Capacité de dialogue avec des biologistes et physiciens. Programmation d'algorithmes d'inférence statistique.

**Partenariat scientifique et industriel** Ce travail de thèse s'inscrit dans une collaboration commencée en 2017 entre l'équipe de bryologie & biosurveillance de l'UMS 2006 Patrimoine Naturel (OFB, MNHN, CNRS) et l'équipe Morse de l'unité MIA-Paris (UMR 518 AgroParisTech, INRAE, Université Paris-Saclay).

## Références bibliographiques

Banerjee S., Carlin B.O., Gelfand A. E.. Hierarchical Modeling and Analysis for Spatial Data, Chapman and Hall, second edition, 2014, pp. 584.

Cressie N. Statistics for Spatial Data, Revised Edition, Wiley, 2015.

Lequy E., Siemiatyckid J., Leblond S., Meyer C., Zhiving S., Vienneau D., de Hoogh K., Goldberg M., Zins M., Jacquemain B.. Long-term exposure to atmospheric metals assessed by mosses and mortality in France. Environment International 2019, 129: 145-153

McElreath R. Statistical Rethinking: A Bayesian Course With Examples in R and Stan, Second Edition, Chapman and Hall, 2020.

Meyer C., Leblond S., Jacquemain B., Lequy E.. Métaux, pollution de l'air et santé: Les mousses, des alliées originales en épidémiologie. Medecine sciences: M/S 2020, 36 (4) : 376-381

Van Honacker J., Albert I., Parent E, Leblond S, Meyer C, Donnet S. Modélisation spatiale de contaminants métalliques. Application aux données de retombées atmosphériques métalliques enregistrées par les mousses, Rapport de M1 STDV. MIA 518 – 30/09/2018

Vieille B., Albert I., Leblond S., Couvidat F., Parent E. et al. Are Grimmia Mosses Good Biomonitors for Urban Atmospheric Metallic Pollution? Preliminary Evidence from a French Case Study on Cadmium, Atmosphere, MDPI 2021, 12 (4), pp.491.