

## OFFRE DE STAGE

### Apport des approches de machine learning pour la cartographie haute résolution de la qualité de l'air en France métropolitaine

Nos réf. : Ineris - [217830] - ID 2726173

**Date de publication :** 20/11/2023

**Lieu :** Verneuil-en-Halatte (60) - accessible en transports en commun, à 40 mn au Nord de Paris

**Type de contrat :** stage

**Contact :** [antoine.guion@ineris.fr](mailto:antoine.guion@ineris.fr) - Tél. : 06 62 76 62 97 pour plus d'information

Le stage proposé s'effectuera au sein de l'unité MOCA (Modélisation Atmosphérique et Cartographie Environnementale). L'unité compte plus de 10 permanents travaillant dans des projets nationaux et internationaux portant sur la modélisation et la cartographie de la qualité de l'air.

Le stage a pour l'objectif de développer des nouveaux outils de cartographies de la qualité de l'air, afin de renforcer la qualité des produits disponibles sur la plateforme nationale de prévision de la qualité de l'air ([www.prevoir.org](http://www.prevoir.org)) et la cartothèque de l'Ineris (<https://www.ineris.fr/fr/ineris/actualites/20-ans-evolution-qualite-air-cartographies-ineris>). Ces informations servent notamment d'aide à la décision pour définir des politiques publiques d'amélioration de la qualité de l'air. Le développement de méthodes de cartographie de la qualité de l'air à haute résolution est un sujet en pleine expansion pour répondre à la fois aux besoins d'informations de proximité (pollution près des zones d'émissions) et régionales (pollution de fond).

La cartographie des différents polluants réglementés à l'échelle de la France métropolitaine est une activité phare de l'unité MOCA qui alimente la cartothèque publique de l'INERIS retraçant l'évolution de la qualité de l'air depuis les années 2000. Celle-ci repose principalement sur une approche de descente d'échelle par interpolation (méthode de krigeage) à partir des simulations numériques du modèle CHIMERE (codéveloppé à l'INERIS), combinées à des données de mesures collectées sur le terrain.

Une méthode alternative de cartographie, parfois complémentaire, repose sur l'utilisation de modèles de régressions statistiques depuis les caractéristiques d'utilisation du sol (modèles LUR, Land Use Regression). Bénéficiant des avancées récentes dans le domaine du machine learning, les modèles LUR offrent des opportunités considérables en termes de résolution spatiale et de temps de calcul pour la représentation des sources de pollutions locales (p. ex. Gouldsbrough et al., 2023<sup>1</sup> ; Horálek et al., 2020<sup>2</sup>).

Le stage proposé consistera à contribuer au développement de modèles LUR par des approches supervisées de régression en machine learning (p. ex. algorithmes 'random forest'

---

<sup>1</sup> [EGUsphere - A machine learning approach to downscale EMEP4UK: analysis of UK ozone variability and trends \(copernicus.org\)](https://www.copernicus.org/)

<sup>2</sup> [ETC/ATNI Report 11/2020: European air quality interim mapping under ETC/ATNI. Evaluation of AQ mapping using UTD measurement and CAMS forecast modelling data: an approach for more timely European AQ annual maps? — Eionet Portal \(europa.eu\)](https://eionet.europa.eu/)

ou 'regularized regression'). De manière plus précise, le travail pourrait suivre le programme de tâches suggérées ci-après :

- Appropriation du sujet et exploration de la littérature scientifique
- Récupération et homogénéisation des déterminants spatiaux comme données d'entrée dans le modèle
- Mise en place du/des modèle(s) LUR pour une cartographie annuelle
- Validation de la méthode et ajustement des hyperparamètres
- Exploration pour extension de la méthode au domaine Europe
- Exploration pour l'intégration de la composante dynamique temporelle

## PROFIL

Bac+4 ou Bac+5 (Ingénieur / Master 1 ou 2). Ce stage s'adresse à un.e étudiant.e ayant suivi des enseignements en (géo) statistiques et/ou machine learning. Des compétences en programmation R ou python sont nécessaires et en Bash sont recommandées. L'environnement de travail est sous LINUX.

## Expérience / Compétence

Des compétences en science des données et modélisation sont requises pour ce stage, et une connaissance des enjeux environnementaux tel que la qualité de l'air serait un plus. Le stage s'insère dans un cadre de développement d'outils et de produits ayant vocation d'apporter une aide à la décision pour la gestion de problématiques environnementales, un fort intérêt pour la recherche et le développement d'approches scientifiques appliquées est également apprécié.

## DIVERS

Durée : 4 - 6 mois

Indemnité mensuelle : environ 530 € (stage à temps plein). Remboursement abonnement SNCF (à hauteur de 50%).

**Ce poste est ouvert aux personnes en situation de handicap.**