

STAGE DE RECHERCHE de MASTER 2^{ème} ANNEE

Master SOAC / WAPE

Année Universitaire 2024-2025

LABORATOIRE : Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA)

SUJET DU STAGE : Estimation of Pollutant Emissions from REMote sensing data and deep Learning (ESPEREL)

COORDONNEES DU RESPONSABLE :

Nom – Prénom : Dufour Gaëlle

Adresse: LISA – Université Paris Est Créteil – 61 avenue du Général de Gaulle – 94000 Créteil

Téléphone : 0145176546

E-mail : gaelle.dufour@lisa.ipsl.fr

NATURE DU SUJET :

Théorie	Pas du tout	Un peu	Beaucoup
Modélisation num.	Pas du tout	Un peu	Beaucoup
Expérimentation	Pas du tout	Un peu	Beaucoup
Analyse de données	Pas du tout	Un peu	Beaucoup
Instrumentation	Pas du tout	Un peu	Beaucoup

SUJET :

L'industrialisation et l'urbanisation rapides du XX^e siècle ont entraîné une hausse des émissions de polluants d'origine anthropique, faisant de la pollution de l'air un risque majeur pour la santé, avec environ 3 millions de décès annuels attribués à la pollution extérieure par l'OMS. Si les pays développés ont commencé à s'attaquer aux problèmes de qualité de l'air dans les années 1990, les réglementations efficaces dans les pays en développement, comme la Chine, sont plus récentes.

Des données précises, haute résolution et à long terme sur les émissions de polluants sont essentielles pour évaluer avec précision l'impact des variations d'émissions sur la chimie atmosphérique et le climat, mesurer l'efficacité des contrôles d'émission et améliorer la gestion de la qualité de l'air. Cependant, les inventaires actuels des émissions reposent largement sur les déclarations des émetteurs et des statistiques annuelles agrégées à l'échelle nationale, ce qui entraîne des incertitudes et ne permet pas de mettre à jour rapidement les inventaires, limitant ainsi l'efficacité des prévisions de qualité de l'air.

Les avancées en télédétection par satellite, notamment avec Sentinel 5P et les futurs Sentinel 4 et 5, fournissent des données fréquentes et haute résolution. Associées aux techniques d'assimilation de données, ces observations satellitaires devraient permettre des estimations d'émissions précises de manière journalière voire horaire, répondant à de nombreuses limitations des inventaires traditionnels. Cependant, la gestion et l'analyse des grandes

quantités de données satellitaires posent également des défis. Pour y répondre, notre projet vise à développer des systèmes basés sur le deep learning, capables de fournir des estimations d'émissions plus précises et en quasi-temps réel, avec une résolution spatiale et temporelle élevée.

En se concentrant sur les émissions d'oxydes d'azote (NO_x), polluants clés dans la formation de l'ozone et des particules fines, le projet propose d'utiliser les données de NO₂ de Sentinel 5P pour étudier la Chine, où les récentes réglementations ont réduit les émissions de plus de 20 %. En collaboration entre les laboratoires LISA et LIPADE et avec le soutien du Data Intelligence Institute de l'Université Paris Cité, nous avons commencé à développer des systèmes de deep learning basés sur l'émulation du modèle de chimie-transport CHIMERE et utilisant des données satellitaires synthétiques. Le ou la stagiaire de Master appliquera ces systèmes aux observations réelles de Sentinel 5P. Cela pourrait nécessiter de revoir les données d'apprentissage et d'utiliser des techniques à l'état de l'art des sciences des données pour prendre en compte les différences de distribution entre les observations réelles et synthétiques. Les estimations d'émissions issues des modèles de machine learning seront utilisées comme nouvelles entrées dans CHIMERE pour simuler les concentrations de polluants et les comparer aux mesures de concentration in situ pour évaluer les améliorations dans la modélisation de la qualité de l'air. Une publication est envisagée à partir de ce travail.

REMARQUES :

Le candidat doit posséder une solide formation en sciences de l'atmosphère, une connaissance des concepts de modélisation atmosphérique, de solides compétences en programmation—en particulier en Python—et soit des bases en machine learning, soit un intérêt de développer cette compétence.

Le stage sera encadré par une équipe interdisciplinaire incluant physico-chimistes de l'atmosphère (LISA) et informaticiens spécialistes des sciences des données (LIPADE). Le stage pourra durer jusqu'à 6 mois avec une gratification suivant la législation en vigueur et pourra se poursuivre en thèse en fonction des opportunités de financement.

Les candidatures sont à envoyer à gaelle.dufour@lisa.ipsl.fr et sylvain.lobry@u-paris.fr avec le sujet « [Internship ESPEREL 2025] Prénom Nom » et inclure un CV, les notes et classements de Master si possible, une lettre de motivation et les coordonnées d'un professeur pouvant recommander la candidature.