

Proposition de stage INSA Rouen / Atmo Normandie

Méthodes statistiques pour la fusion de données

Lieu du stage : INSA Rouen Normandie et Atmo Normandie.

Niveau d'études : Master 2, Ecole d'ingénieur.

Sujet : Méthodes statistiques pour la fusion de données.

Date de début : Mars 2020.

Durée du contrat : 6 mois.

Gratification : 577,50 euros par mois, conformément à la grille officielle en vigueur.

Contexte. Atmo Normandie, l'Association Agréée de Surveillance de la Qualité de l'Air en Normandie (<http://www.atmonormandie.fr>), et le Laboratoire de Mathématiques de l'INSA Rouen Normandie (<http://lmi.insa-rouen.fr>) recherchent un stagiaire pour l'étude statistique des mesures de la qualité de l'air effectuées par divers types appareils en des positions géographiques différentes. A l'heure des micro-capteurs, des objets connectés et autres API, une récente étude menée par l'université Paris-Sud et l'INSA Rouen Normandie avec Atmo Normandie, a permis de recenser les méthodes statistiques de fusion/combinaison des mesures provenant de plusieurs réseaux, tant fixes que mobiles. Ces réseaux de mesures peuvent être classés en trois catégories pour simplifier :

- le réseau 1 est de haute qualité mais réduit à quelques dizaines de stations fixes à l'échelle régionale, concentrées autour des zones de population denses pour l'essentiel ;
- le réseau 2 est de qualité moyenne et constitué d'un réseau fixe, et bien plus dense que le réseau 1 mais réduit aux agglomérations typiquement ;
- le réseau 3 constitué des micro-capteurs qui équipent des objets connectés est de basse qualité mais potentiellement très dense, en particulier dans les zones dans lesquelles les deux autres réseaux sont muets.

L'enjeu à terme est de tirer profit des nouvelles mesures provenant des réseaux 2 et 3, pour améliorer les cartographies basées sur les mesures du réseau 1 et des modèles physico-chimiques.

Sujet : Le stage portera sur les méthodes statistiques de correction mutuelle des mesures de chacun de ces réseaux. Parmi les possibilités, le sujet pourrait intégrer :

- La lecture d'une synthèse bibliographique.
- La lecture d'une première étude de données fournies par les réseaux 1 et 2 mettant en évidence les relations entre les deux types de mesures.

- L'extension aux aspects spatiaux de ces relations. - L'utilisation de cette modélisation pour la simulation de pseudo-données.
- Plus généralement, l'exploration de méthodes de simulation de pseudo-données réalistes.
- Un travail combinant des données réelles ou simulées du réseau 2 par une stratégie de correction itérative des micro-capteurs par le réseau fixe. L'idée consiste en deux étapes exécutées itérativement :
 - o La première vise à partir du réseau 2, à construire par krigeage une carte. Le krigeage fournit en particulier une estimation en tous les points du réseau 1.
 - o On peut alors, lors d'une deuxième étape, adapter statistiquement la carte obtenue en utilisant les mesures.

Une première étape du travail sur ce point est en cours et l'un des enjeux du stage sera de le prolonger et de le consolider tout en l'exploitant sur des données réelles. Le sujet du stage sera dimensionné en fonction des candidat.e.s intéressé.e.s.

Profil du candidat : Ecole d'ingénieur ou Master en mathématiques appliquées, avec de bonnes compétences en statistiques et en informatique, et la pratique du langage de programmation R.

Candidature : Envoyer un message électronique accompagné d'un CV et d'une lettre de motivation à Bruno Portier : bruno.portier@insa-rouen.fr, ainsi qu'à Michel Bobbia : michel.bobbia@atmonormandie.fr.