



**Titre : Modélisation statistique des concentrations des polluants dans l'atmosphère**

**Direction de stage :**

- Yousri SLAOUI (LMA de Poitiers, email : [yousri.slaoui@math.univ-poitiers.fr](mailto:yousri.slaoui@math.univ-poitiers.fr))

- Zaher Khraibani (LEESU, Paris Est Créteil, email : [zaher.khraibani@gmail.com](mailto:zaher.khraibani@gmail.com))

**Lieu de stage : LMA, Université de Poitiers**

**Date de début souhaitée : Avril 2021.**

**Durée et Rémunération/Gratification : gratification de stage (taux légal), 3 mois**

**Résumé :** L'environnement est le milieu de vie des êtres vivants. Il est ainsi le centre des actions, réactions et coactions entre les êtres vivants. Il exerce sur les êtres vivants des actions diverses : par exemple, l'ensemble des facteurs écologiques de l'environnement détermine la nature et la diversité des espèces présentes dans une région, leur abondance relative et leur répartition. La pollution atmosphérique est l'une des plus grandes menaces pour l'environnement qui atteint les êtres vivants, les villes, les forêts, les écosystèmes aquatiques, etc. L'objectif de ce stage est

d'étudier la pollution atmosphérique en utilisant la modélisation statistique pour prédire les concentrations des particules dans l'air. La plupart des travaux dans la littérature se basent sur les séries temporelles, des modèles non-linéaire par exemple les réseaux de neurones. Nous envisageons dans ce stage d'étudier d'une part des modèles dynamiques basés sur la théorie du chaos, d'autre part sur des modèles probabilistes, basés sur les chaînes de Markov. Ces modèles sont peu appliqués dans le domaine de la pollution de l'air. Citons par exemple le travail du Dong et al 2009 qui prédit la concentration du PM2.5 dans l'atmosphère avec un modèle de chaîne de Markov caché (Hidden Markov Model, HMM). Ces modèles nécessitent une estimation supplémentaire (diffère de modèle classique qui nécessite que les états sont récurrents).

## **OBJECTIFS**

- 1) Estimation des paramètres des modèles proposés par des techniques de type d'apprentissage statistiques comme les réseaux de neurones récurrent de type (LSTM (Long Short Term Memory)).
- 2) Application des modèles proposées sur des données réelles PM2.5, PM10, NOx, O3, dont l'objectif est de prédire les concentrations des polluants sur deux niveaux (faible/élevée). Des données qui sont déjà disponibles.
- 3) Implémenter sous R la méthode proposée avec une comparaison avec les autres approches proposées dans la littérature.

## **Références Bibliographiques**

Dong, M., Dong, Y., Kuang, Y., He, D., Erdal, S. and Kenski, D. (2009) PM2.5 concentration prediction using hidden semi-Markov model-based times series data mining. *Expert Systems with Applications*, 36: 9046-9055

Jovanovic, R. Z., Sretenovic, A. A. and Branislav D. Zivkovic, B. D. (2015) Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings* 94 (2015) 189–199.

Lee, A. X., Nagabandi, A., Abbeel, P. and Levine, S. (2020) Stochastic Latent Actor-Critic: Deep Reinforcement Learning with a Latent Variable Model. *Neural Information Processing Systems*.

Xiao Feng, Qi Li, Yajie Zhu, Junxiong Hou, Lingyan Jin, Jingjie Wang, Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation, Atmospheric Environment, Volume 107, 2015, Pages 118-128, ISSN 1352-2310, <https://doi.org/10.1016/j.atmosenv.2015.02.030>.