

**Offre de stage :**

Clustering of rainfall extremes by coupling Kullback-Leibler divergence and machine learning algorithms

**Niveau :** M2

**Cadre :**

- encadrant principal : Naveau Philippe, DR CNRS
- co-encadrant(s) : Prof. Olivier Wintenberger
- Laboratoire(s) : LSCE, LPSM

**Durée et Période :** 5 ou 6 mois, idéalement du 03/02/2020 au 02/08/2020

**Lieu :** 4 Place Jussieu, Paris, France

**Sujet:**

En théorie de l'apprentissage automatique et en théorie l'information, la divergence de **Kullback-Leibler** (KL) est souvent utilisée comme fonction de perte, par ex. en renforcement et / ou apprentissage supervisé. Pour comparer les comportements extrêmes, Naveau et al. (2014) ont proposé une version de la KL adaptée aux cas extrêmes. Au cours du stage, l'étudiant étudiera comment une telle divergence peut être utilisée pour le regroupement spatial de données de précipitations, c'est-à-dire comment améliorer l'Analyse de Fréquence Régionale (**RFA**) utilisée en hydrologie. Il est bien connu que de fortes précipitations peuvent déclencher des crues soudaines, principal risque naturel dans certaines régions et provoquer des dégâts importants, e.g. les pluies Cévenols en région Méditerranéenne. Les facteurs qui expliquent la distribution spatiale de ces occurrences de pluies intenses sont complexes et variés. Cette complexité implique qu'une modélisation fine des incertitudes est nécessaire et, dans ce contexte, plusieurs approches probabilistes ont été proposées. En particulier, la RFA s'appuie sur le concept de régions homogènes pouvant être définies de manière contiguë dans l'espace. Ce découpage régional a le double avantage de simplifier la modélisation à l'intérieur d'une région et pouvoir interpréter climatologiquement les contrastes entre les différentes régions. Toutefois, la création de telles régions homogènes n'est pas simple car elle doit être basée sur l'intensité d'événements extrêmes de pluie, leurs proximités spatiales. La théorie des valeurs extrêmes [**EVT**, Coles, 2001] fournit un cadre mathématique couramment utilisé pour modéliser la distribution d'extrêmes. L'EVT univarié a été appliquée dans le cadre de la RFA, voir Hosking et Wallis (2005) et Carreau, Naveau and Neppel (2017). Dans le cadre multi-dimensionnel qui intégrerait l'EVT multivarié pour prendre en compte la structure spatiale dans les champs de précipitations extrêmes, il n'existe pas, à notre connaissance, de méthode de clustering spatiale MEVT applicable à la RFA. Pour développer de

telles méthodes, il est fondamental d'avoir un critère simple de comparaison de densités. Une version multivariée de la KL divergence et adaptée aux extrêmes multivariés nous permettra de proposer un critère simple pour la RFA. Ce critère devra être intégré aux algorithmes de clustering spatial utilisés en machine learning. Dans ce cadre, l'étudiant(e) étudiera, à partir de simulations et d'analyse de données de pluies, la faisabilité d'une telle proposition à l'intersection de trois domaines: machine learning, EVT and rainfall analysis.

### **Références bibliographiques :**

P. Naveau, A. Guillou, and T. Rietsch. A non-parametric entropy based approach to detect changes in climate extremes. Journal of the Royal Statistical Society : Series B (Statistical Methodology), 2014.

J.Carreau, P.Naveau ,an dL.Neppel .Partitioningintohazardsubregionsfor regional peaks-over-threshold modeling of heavy precipitation. Water Resour. Res., 53 :4407–4426, 2017.

**Autres :** une thèse dans le prolongement du stage est prévu