

## Offre de Stage IPSL 2021

(soutenu par le programme EUR IPSL-Climate Graduate School)

Titre du sujet de stage : Développement d'algorithmes d'apprentissage causaux et d'agrégation d'experts pour prendre en compte l'erreur multi-modèles dans les études de détection et d'attribution de records climatiques

### **Description du sujet** (1 page maximum) :

Les modèles climatiques globaux comme toutes les expériences numériques « in silico » sont entachés de différents types de biais. La quantification des incertitudes reste un défi dans toute analyse d'attribution climatique, en particulier lorsqu'il s'agit de quantifier l'influence des activités humaines sur le climat. Une question méthodologique fondamentale est de déterminer quelles mesures statistiques, tout en restant physiquement interprétables, peuvent être robustes vis-à-vis des erreurs présentes dans les modèles de climat. Dans ce stage, l'étudiant étudiera l'apport des méthodes causales pour améliorer la détection des signaux d'origine anthropique dans les études d'attribution du climat. L'objectif principal est d'étudier l'influence des activités humaines sur les extrêmes climatiques, ici des records de précipitations et de températures.

La théorie des valeurs extrêmes fournira un cadre mathématique pour modéliser la distribution de records et devra être couplée avec des méthodes d'agrégation d'experts, de correction de biais et de clustering pour synthétiser les informations parfois divergentes des différents modèles de climat tout en tenant compte de leurs incertitudes. Les extrêmes climatiques étant au centre de nombreux enjeux actuels, il est important que des méthodes soient développées pour évaluer et corriger le réalisme non seulement de la partie « centrale » des données générées par les modèles, mais également de la partie extrême, qui requiert des traitements spécifiques.

Des comparaisons avec des méthodes classiques d'apprentissage (de type GAN, Generative Adversarial Network) seront effectuées. En particulier, nous tenterons de voir si les méthodes d'apprentissage sont capables de générer des événements extrêmes et ce avec la bonne fréquence. Nous pourrions également évaluer si le discriminateur d'un GAN est capable de distinguer les événements extrêmes venant de simulations avec forçage anthropique des événements générés sans forçage anthropique. L'apport de la théorie des valeurs extrêmes sera également étudié dans ce cadre.

En termes d'applications, nous déduirons les temps d'émergence des précipitations à partir des cycles CMIP sur la période 1850-2100 selon le scénario RCP 8.5.

## Bibliographie :

1. A. Ribes, S. Thao, J. Cattiaux (2020), Describing the relationship between a weather event and climate change: a new statistical approach, *Journal of Climate* 33 (15), 6297-6314
2. J. Worms and P. Naveau (2020), Record events attribution in climate studies , *submitted* <https://hal.archives-ouvertes.fr/hal-02938596>
3. P. Naveau and S. Thao (2020), Integrating multi-model errors to infer emergence times in climate attribution studies, *submitted*.
4. A. Kiriliouk and P. Naveau (2019), Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory. *Annals of Applied Statistics*, <https://arxiv.org/abs/1908.03107>.
5. S. Bhatia, A. Jain, B. Hooi (2020), ExGAN: Adversarial Generation of Extreme Samples, *ArXiv* <https://arxiv.org/pdf/2009.08454v1.pdf>

## Résumé en anglais (5 lignes) :

Global climate models (GCM) like all “in silico” digital experiments are tainted by different types of bias. In this internship, the student will study the contribution of causal methods to improve signal detection in climate attribution studies. The main objective is to study extremes, here records of precipitation and temperatures, and this will imply coupling extreme value theory with bias correction methods to take into account inter-GCMs errors. Comparisons with traditional learning methods (such as GAN models) will be made, and precipitation emergence times will be deduced from the CMIP cycles over the period 1850-2100 according to the RCP 8.5 scenario.

## Responsable(s) du stage (Nom/prénom/statut) :

- Soulivanh Thao, IR, CEA contact : [soulivanh.thao@lsce.ipsl.fr](mailto:soulivanh.thao@lsce.ipsl.fr)
- Julien Worms, MC, UVSQ contact : [julien.worms@uvsq.fr](mailto:julien.worms@uvsq.fr)
- Philippe Naveau, DR, CNRS contact : [philippe.naveau@lsce.ipsl.fr](mailto:philippe.naveau@lsce.ipsl.fr)

## Laboratoire(s) concerné(s) :

- Laboratoire de mathématiques de Versailles UVSQ
- Laboratoire des Sciences du Climat et de l'Environnement

**Equipe de recherche concernée** (si pertinent) : EstimR + Laboratoire de mathématiques de Versailles

**Niveau du stage** (Licence, M1, M2, internship) : M2 avec des très bonnes connaissances en machine learning et/ou statistiques

**Licence ou Master(s) où sera proposé le sujet** : Masters de Data Sciences avec une bonne connaissance en modélisation statistique (e.g. LPSM, ENSTA, etc) + Masters associés à l'ED129 + TRIED

**Thème scientifique de l'IPSL concerné : SAMA ML**

**Durée du stage : 6 mois**

**Période : 03/02/2021 → 02/08/2021**

**Est-il prévu une thèse dans le prolongement du stage ? Cela dépendra des demandes financières**