

Fiche sujet de stage (à remplir par le proposant)

Machine Learning contraint par la physique pour l'émulation de paramétrisations sous-maille

Description du sujet :

Les modèles de climat reposent sur une division du travail entre d'une part un solveur de mécanique des fluides qui calcule l'évolution de grandeurs dites « résolues » (vitesse, température, pression), c'est-à-dire connues avec un degré de détail spatial dicté par la taille des mailles du modèle, et d'autre part des sous-modèles qui représentent statistiquement l'effet de processus d'échelle plus petite que la taille de la maille (turbulence, convection, nuages, interaction rayonnement-atmosphère) sur les grandeurs résolues. Ces sous-modèles sont appelés paramétrisations physiques et représentent une grande partie d'un modèle de climat. Grossièrement parlant, une paramétrisation associe à un profil vertical de grandeurs résolues (entrées) leurs dérivées temporelles sous l'effet des processus sous-maille (sorties). Une tendance actuelle est d'obtenir cette association entrées->sorties par des techniques de machine learning (ML), par exemple des réseaux de neurones profonds (Gentine et al. 2018 ; Geophysical Research Letters ; 45 ; 5742-5751).

Parmi les avantages potentiels d'une telle approche, on trouve la possibilité de créer ces paramétrisations à partir d'observations ou de simulations à plus haute résolution résolvant les processus en question, ou la possibilité de les exécuter plus efficacement sur des architectures de calcul dédiées au ML. Le stage proposé se place dans la deuxième optique, où l'on envisage d'émuler par ML une paramétrisation considérée comme satisfaisante mais coûteuse en calcul.

L'idée de remplacer une paramétrisation physique par un émulateur obtenu par ML soulève des questions fondamentales comme la capacité de telles paramétrisations à affronter des conditions météorologiques inconnues, ou à maintenir la stabilité du modèle. Ce stage aborde la question du respect de structures physiques fondamentales comme la conservation de l'énergie. En effet il est indispensable pour que les simulations issues d'un modèle de climat soient exploitables scientifiquement que, par exemple, l'énergie gagnée ou perdue par l'atmosphère soit égale à la somme des échanges d'énergie (flux radiatifs, flux de chaleur sensible, de chaleur latente) à sa base et à son sommet.

Le but du stage est de formuler des problèmes d'apprentissage qui respectent des structures physiques fondamentales, de leur appliquer des techniques de ML de type réseaux de neurones profonds, et d'évaluer la performance du processus d'apprentissage d'une part, de l'émulateur obtenu d'autre part. Le stage s'appuiera sur un jeu de paramétrisations simplifiées, dans une atmosphère sèche, évitant la manipulation d'un modèle de climat complet. En premier lieu, un jeu de données d'entrées sera créé à partir de simulations idéalisées existantes. Le problème de l'émulation « structurée » de paramétrisations physiques sera ensuite abordé, en commençant par une paramétrisation très simple de transfert radiatif. Dans ce cas précis, plutôt que de définir comme sorties des taux de chauffage de l'atmosphère, on pourra définir les sorties comme les flux radiatifs (dont on peut déduire les taux de chauffage). Se posera aussi la question de la fonction de coût à minimiser lors de la phase d'apprentissage. Selon l'avancement du travail, on pourra ensuite envisager :

- d'évaluer le climat produit par l'introduction de l'émulateur dans un modèle de circulation générale
- et/ou d'appliquer d'autres types d'émulateurs ou d'autres représentations des entrées/sorties (p.ex. représentation en bases d'ondelettes) afin d'accélérer/améliorer l'apprentissage
- et/ou d'aborder d'autres processus physiques (turbulence, ajustement sec)

Résumé en anglais : Deep learning techniques can be useful to emulate parts of numerical models of the atmosphere, especially subgrid-scale models (parameterizations). The internship will explore formulations of the machine learning problem that incorporate physical constraints that enforce conservation laws. The internship will focus initially on emulating a simplified radiative transfer computation. It may then take several directions, such as evaluating the climate obtained when using the emulator, improving the structure of the emulator and the learning procedure, or addressing other physical processes.

Responsable du stage (Nom/prénom/statut) : DUBOS Thomas, PR Polytechnique/LMD
KAGEYAMA Masa, CEA/LSCE
VRAC Mathieu, DR CNRS/LSCE

Laboratoire concerné : LMD / LSCE

Equipe de recherche concernée (si pertinent) :

Niveau du stage (Licence, M1, M2, internship) : M2

Licence ou Master(s) où sera proposé le sujet : MOCIS/WAPE

Thème scientifique de l'IPSL concerné : Statistics for Analysis, modeling and assimilation (SAMA)

Durée du stage : 4-6 mois

Période : mars – août 2020

Est-il prévu une thèse dans le prolongement du stage ?

Oui, un financement CEA a été demandé, sinon financement à trouver (ED)