



Sujet de stage de Master 2.

Prise en compte de contraintes physiques pour l'apprentissage statistique de codes de simulation.

Centre d'accueil : CEA/DAM/DIF, F-91297, Arpajon, France

Contact : Guillaume Perrin (guillaume.PERRIN2@cea.fr).

Mots clés : machine learning, krigeage, réseaux de neurones, small data.

Contexte. La simulation est de plus en plus utilisée pour la garantie et la conception de systèmes complexes. De manière classique, ces analyses passent par la mise au point de codes numériques, puis par l'appel de ces codes en un grand nombre de configurations, que l'on suppose pouvoir être décrites par un vecteur \mathbf{x} (en faisant varier la pression, la température, les dimensions, ou les matériaux du système par exemple).

Lorsque le coût numérique d'un appel au code est élevé (l'appel à certains codes multi physiques peut nécessiter plusieurs jours de calculs sur les supercalculateurs du CEA), ces analyses demandent souvent l'introduction d'approximations (aussi appelées métamodèles) pertinentes et rapides à calculer d'une ou plusieurs quantités d'intérêt $y(\mathbf{x})$ fournies par le code, à partir d'un nombre limité de simulations.

Descriptif de l'étude. Les méthodes d'apprentissage statistique telles que la régression par processus gaussien (krigeage) ou les réseaux de neurones permettent de construire des prédicteurs de la fonction $\mathbf{x} \rightarrow y(\mathbf{x})$ à partir d'un nombre fini de couples entrées-sorties. Dans un contexte « small data », c'est-à-dire pour lequel le nombre maximal d'appels au code pour l'apprentissage est relativement faible par rapport à la dimension de \mathbf{x} , le caractère prédictif de tels approximations peut être fortement amélioré par l'intégration d'une ou plusieurs contraintes physiques sur le comportement du système, comme des contraintes de monotonie ou de convexité, ou par l'existence de relations entre les différentes quantités d'intérêt (équations de conservation, lois d'état...).

L'objectif du stage est alors de proposer des adaptations des formalismes classiques de la régression par processus gaussien et des réseaux de neurones pour leur permettre de tirer profit de ces informations physiques supplémentaires. L'efficacité des approches développées sera finalement testée sur des cas tests numériques proches des problèmes traités au CEA.

Connaissances requises. Ce stage demande des connaissances en probabilités/statistique et en programmation de type python, R, matlab ou Julia.

Détails pratiques. Le poste est basé sur le centre CEA DAM Île de France situé à Bruyères-le-Châtel. Des lignes de bus CEA desservent le centre depuis Paris et la banlieue parisienne.