

Proposition de stage de Master 2 pour l'année 2019

Analyse de sensibilité pour modèles stochastiques

Encadrement

Gildas Mazo, Elisabeta Vergu (INRA, UR1404 MaIAGE)

Contexte scientifique

La simulation numérique occupe aujourd'hui une place prépondérante en science et en ingénierie. Les méthodes utilisées, souvent empiriques, ouvrent de nombreuses perspectives de recherche en mathématique et en informatique. Par exemple, les modèles construits pour simuler un phénomène d'intérêt dépendent en général de paramètres inconnus. Pour palier ce problème, il est naturel de spécifier une plage de valeurs à l'intérieur de laquelle les paramètres sont tirés au hasard, et d'observer les sorties qui en résultent. La sortie est donc vue comme un variable aléatoire dont la variance représente l'incertitude due à notre méconnaissance des entrées. On parle de propagation d'incertitude : l'incertitude sur les paramètres est propagée jusqu'à la sortie du modèle.

Quels sont, dès lors, les paramètres importants à prendre en compte ? L'analyse de sensibilité peut apporter des éléments de réponse, en ce qu'elle permet de quantifier l'influence de chaque entrée. Plus précisément, elle permet de calculer, par échantillonnage Monte-Carlo, la proportion de variance qui serait éliminée si tel ou tel paramètre était fixé. Les quantités calculées sont appelées des indices de Sobol. Cela permet au scientifique ou à l'ingénieur de mettre des oeuvre des stratégies de réduction de la complexité du modèle ou de son incertitude.

Il existe des modèles intrinsèquement stochastiques, c'est à dire que, même si les paramètres étaient fixés, la sortie du modèle serait aléatoire. Dans ce cas, réaliser une analyse de sensibilité est plus complexe car sa performance dépendra d'un compromis entre le nombre d'explorations de l'espace des paramètres et du nombre de répétitions du modèle. Dans un travail en cours, nous avons proposé une approche théorique qui exploite le rapport signal-sur-bruit et avons démontré son optimalité asymptotique. Néanmoins, sa performance doit être évaluée à taille d'échantillon finie et sur des modèles réalistes.

Objectifs

L'objectif de ce stage est d'évaluer la performance de la méthode que nous avons proposée sur des modèles réalistes de propagation d'épidémies. Il s'agit de savoir si cette méthode permet effectivement d'obtenir des estimations des indices de Sobol plus précises que les méthodes empiriques utilisées aujourd'hui. L'étudiant mènera une étude de cas qui, idéalement, donnera lieu à la publication d'un article de recherche.

Sous réserve de l'obtention de financement, il est envisagé que ce stage se prolonge par une thèse dans laquelle une extension de notre méthode à des modèles dynamiques devra être proposée. Ainsi, en plus de l'aspect pratique, la thèse comportera un important volet théorique.

Compétences recherchées

Master 2 en mathématiques appliquées avec une forte composante en statistique. Compétences en programmation (R ou Python ou C++), inférence statistique, bon équilibre entre intérêts pour les applications et théorie.

Déroulement du stage

Le stage se déroulera dans l'unité MaIAGE (Mathématiques et Informatique Appliquées du Génome à l'Environnement) du centre INRA de Jouy-en-Josas (78350). La durée prévue est de 5 à 6 mois. Gratification : 550 euros/mois environ (taux légal).

Candidature

Envoyer une lettre de motivation, votre CV et la liste des cours suivis en M1/M2 avec toutes les notes disponibles à gildas.mazo@inra.fr, elisabeta.vergu@inra.fr.