

Métamodélisation pour la sélection de variables en grande dimension dans les modèles non linéaires à effets mixtes complexes.

Application en amélioration des plantes.

Proposition de stage niveau M2 (printemps 2024)

*Pour postuler envoyer CV et dernier relevé de notes à
maud.delattre@inrae.fr et marion.naveau@inrae.fr.*

Contexte applicatif

Les modèles à effets mixtes permettent d'analyser des observations collectées de façon répétée sur plusieurs individus. La variabilité intrinsèque aux données est alors attribuable à différentes sources (intra-individuelle, inter-individuelle, résiduelle) dont la prise en compte est essentielle pour caractériser sans biais les mécanismes biologiques à l'origine des observations. Dans un modèle à effets mixtes, la variabilité entre individus est décrite au moyen de covariables et d'effets aléatoires. Les covariables décrivent les différences entre individus dues à des caractéristiques observées tandis que les effets aléatoires représentent la part de la variabilité entre individus qui n'est pas attribuable aux covariables mesurées. En amélioration des plantes, les modèles non linéaires à effets mixtes sont utilisés pour décrire le développement des plantes en fonction de leurs génotypes et des conditions environnementales. Ils permettent de comprendre le rôle des interactions entre le génotype et l'environnement dans l'évolution de la plante et sont utilisés pour prédire les performances de différentes variétés dans des conditions environnementales spécifiques. Les covariables considérées sont généralement de grande dimension puisque les variétés sont caractérisées par des milliers de covariables génétiques (des marqueurs moléculaires par exemple). Par ailleurs, certains modèles non linéaires à effets mixtes utilisés en amélioration des plantes sont construits à partir de modèles écophysiologiques qui simulent de façon très précise les différentes étapes du développement de la plante et sont coûteux à évaluer. Dans ce cas spécifique, les temps de calculs requis par les algorithmes d'inférence classiques sont trop importants pour qu'ils puissent être utilisés sur données réelles.

Objectifs

Ce stage fait suite au travail récent de Marion Naveau dans lequel la sélection de covariables dans les modèles non linéaires à effets mixtes est réalisée en couplant l'algorithme SAEM et un prior bayésien de type spike and slab [2]. Cette procédure est itérative et nécessite plusieurs évaluations de la fonction de régression non linéaire à chacune de ses itérations. Lorsque la fonction de régression est coûteuse à évaluer, le temps d'exécution de la méthodologie complète devient prohibitif. Des approches de métamodélisation, déjà appliquées avec succès pour l'estimation de paramètres dans des modèles non linéaires à effets mixtes dans [1], pourraient permettre de réduire les temps de calcul.

Le stagiaire débutera par un travail bibliographique visant à comprendre le formalisme des modèles non linéaires à effets mixtes et les approches classiques de métamodélisation. Il réfléchira ensuite à l'utilisation de méta-modèles pour améliorer la méthodologie développée dans [2]. Puis il implémentera la méthode proposée et réalisera des simulations pour en valider le comportement numérique. Enfin, il l'appliquera à des données réelles. L'application sur données réelles se fera en collaboration avec Renaud Rincent (UMR GQE - Le Moulon - Paris Saclay).

Profil recherché

Le candidat doit être en formation de M2 (ou une formation équivalente) en statistique. Un intérêt pour la modélisation statistique, des notions d'apprentissage statistique (éventuellement en grande dimension) et de programmation en R ou Python sont attendus.

Il est à noter qu'aucune connaissance en sciences du vivant n'est exigée.

Conditions du stage

Laboratoires d'accueil

UR 1404 Mathématiques et Informatique Appliquées du Génome à l'Environnement (MaIAGE), INRAE, 78352 Jouy-en-josas

Encadrantes

Maud Delattre : maud.delattre@inrae.fr

Marion Naveau : marion.naveau@inrae.fr

Durée 4-6 mois

Gratification environ 550 euros nets par mois

Références

- [1] Barbillon, P., Barthélémy, C., & Samson, A. (2017) *Parameter estimation of complex mixed models based on meta-model approach*. *Statistics and Computing*, 27(4), 1111-1128.
- [2] Naveau, M., Kon Kam King, G., Rincent, R., Sansonnet, L. & Delattre, M. (2022) *Bayesian high-dimensional covariate selection in non-linear mixed-effects models using the SAEM algorithm*. arXiv preprint arXiv :2206.01012.
- [3] Lavielle, M. (2014) *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*. Chapman & Hall/CRC biostatistics series.