

Méthode bayésienne de réduction de dimension dans le cadre de modèles à effets aléatoires : Application à l'amélioration génétique

Contexte et objectifs :

Que ce soit dans les domaines de la médecine, de l'agronomie ou de l'écologie, une compréhension des processus physiologiques (croissance, floraison, sénescence, ...) est cruciale surtout dans le contexte du changement climatique. Ces processus sont gouvernés par des facteurs environnementaux et des mécanismes biologiques complexes. Une approche communément utilisée consiste à identifier les régions du génomes impliquées dans la variabilité du processus en utilisant des données génomiques (marqueurs moléculaires, expression de gènes, ...) et des caractéristiques environnementales (pluviométrie, température, types de sol, ...). Depuis quelques années, ces données sont toujours plus abondantes allant par exemple jusqu'à des dizaines de milliers de marqueurs génétiques. Dans ce contexte de grande dimension, il est primordial de développer des méthodes statistiques efficaces qui permettent de réduire ces dimensions et de régulariser les modèles utilisés.

Récemment, une approche innovante ([Heuclin et al., 2023](#)) reposant sur un modèle linéaire mixte couplé à une méthode bayésienne de sélection d'effets aléatoires a été proposée afin de sélectionner les régions importantes. Cette approche considère chaque région génomique comme un effet aléatoire où la matrice de variance-covariance est calculée à partir du sous-ensemble de marqueurs disponibles sur une région donnée. Cette méthode soulève trois difficultés : (i) les marqueurs sont redondants et réduire la dimension de la matrice s'avère important d'un point de vue biologique mais aussi computationnel, (ii) le choix de la métrique ne prend pas en compte la qualité d'ajustement du modèle aux données et enfin (iii) sélectionner les régions pertinentes est crucial.

Dans ce contexte, les objectifs du stage seront i) de proposer, en utilisant une approche bayésienne de réduction de dimension, une métrique d'apparement génétique qui tienne compte de la qualité d'ajustement du modèle aux données et ii) de discuter d'une ou plusieurs extensions au cas multivarié afin d'étudier les effets pléiotropiques des gènes sur un ensemble de processus physiologiques.

Les approches proposées seront appliquées dans un contexte d'amélioration génétique du palmier à huile sur des données issues d'essais génétiques mis en place par le [Cirad](#) (Centre de coopération internationale en recherche agronomique pour le développement) et PalmElit, sa filiale pour la sélection et la production de semences.

Profil recherché :

- École d'ingénieur ou Master 1 ou 2 en (bio)statistique ou mathématique appliquée ou agronomie avec dominante en analyse et modélisation statistique des données
- Maîtrise théorique des modèles linéaires mixtes et des concepts bayésiens
- Bonnes compétences en programmation R
- Bonne capacité rédactionnelle (français/anglais)

Déroulement du stage :

Ce stage se déroulera à Montpellier, au sein de l'UMR AGAP du Cirad, pour une durée de 4 à 6 mois entre février et août 2024. Une poursuite en thèse est fortement envisagée. Le stage sera donc assez exigeant sur le plan mathématique.

Encadrants principaux : Xavier Bry, Marie Denis, Frédéric Mortier.

Indemnité mensuelle : 611 euros

Modalités de candidature : CV et lettre de motivation à envoyer par email conjointement à xavier.bry@umontpellier.fr, marie.denis@cirad.fr, frederic.mortier@cirad.fr