

Offre de CDD – Développement et optimisation d'un package R pour la sélection de variables en génétique

CV et lettre de motivation à envoyer à : maud.delattre@inrae.fr

Contexte

Ce contrat s'inscrit dans le cadre du projet ANR Stat4Plant¹, qui vise à développer des outils statistiques pour mieux comprendre les interactions entre les plantes et leur environnement, dans un contexte de changement climatique marqué. L'un des enjeux du projet est de rendre ces outils accessibles à la communauté des biologistes et généticiens via des implémentations logicielles robustes et faciles à utiliser.

Un des développements du projet concerne une méthode de sélection de variables en grande dimension destinée à exploiter des données issues d'expérimentations végétales². Cette méthode a déjà été testée avec succès sur des jeux de données de taille moyenne, mais son implémentation actuelle n'est pas conçue pour des jeux de données de grande dimension, comme ceux générés par le génotypage à l'échelle du génome. Une adaptation et une optimisation du code sont nécessaires pour en permettre une utilisation plus large.

Objectifs du poste

Le travail portera principalement sur l'optimisation et l'extension de cette méthodologie, avec pour finalité sa mise à disposition sous la forme d'un package R à destination de la communauté scientifique.

Les objectifs concrets incluent :

- Optimiser les performances des codes existants, en particulier pour le traitement de données de grande dimension ;
- Utiliser des bibliothèques efficaces comme Rcpp et Armadillo pour accélérer les calculs, notamment les opérations matricielles lourdes ;

¹<https://stat4plant.mathnum.inrae.fr/>

²Détails méthodologiques: Naveau, Marion, et al. "Bayesian high-dimensional covariate selection in non-linear mixed-effects models using the SAEM algorithm." *Statistics and Computing* 34.1 (2024): 53

- Structurer l'ensemble des fonctions dans un package R ;
- Rédiger la documentation technique du package, ainsi qu'éventuellement une vignette illustrant son utilisation.

Profil recherché

Le ou la candidate devra avoir un niveau M1 ou M2 et une formation en informatique avec de solides bases en méthodes numériques, ou une formation en mathématiques appliquées avec une expérience confirmée en programmation scientifique.

Le poste requiert de solides compétences en programmation, notamment en R (y compris le développement de packages), ainsi qu'une bonne maîtrise de C++ et de l'interfaçage avec R via Rcpp.

Une connaissance des modèles à effets mixtes ou des approches bayésiennes serait un plus, mais n'est pas indispensable : ces aspects seront pris en charge par l'équipe d'accueil, qui porte l'expertise statistique du projet.

Informations complémentaires

- Durée : 2 à 3 mois (à définir selon disponibilité)
- Début souhaité : dès que possible
- Lieu : INRAE, Jouy-en-Josas, France
- Encadrement : Maud Delattre, unité MaIAGE.
- Rémunération indicative : 2150 € bruts/mois, ajustée selon l'expérience ou l'ancienneté.