

Optimisation multi-critère et estimation en présence d'aléa. Application à la sélection de variétés multi-performantes en végétal en présence de variabilité environnementale.

Stage niveau Master 2, début 2023

Unité Mathématiques et informatique appliquées
du génome à l'environnement (MaIAGE)
INRAE, Jouy-en-Josas

Pour postuler, merci d'envoyer un **dossier complet contenant votre CV, votre lettre de motivation et vos derniers relevés de notes de Master 1 ou équivalent** à Estelle Kuhn (estelle.kuhn@inrae.fr) et à Jean-Benoist Leger (jbleger@hds.utc.fr).

Contexte

Dans le contexte actuel de changement climatique, l'agriculture est au cœur des préoccupations, à la fois comme l'une des causes de ce processus, mais aussi du fait des bouleversements majeurs qu'elle subira et auxquels elle devra s'adapter. L'une des notions clés qu'il est nécessaire de mieux comprendre et maîtriser pour appréhender ces questions est celle de l'interaction entre la plante et son environnement au sens large (conditions météorologiques, conditions de sol, conduite de culture par l'agriculteur). Les interactions entre plantes sont également importantes, en particulier en agro-écologie. Par ailleurs, la plante peut être vue comme un système pouvant rendre des services multiples, par exemple comme plante de service dans des cultures en association ou pour réduire l'empreinte carbone d'une culture. Il est ainsi fondamental de pouvoir optimiser la sélection de variétés performantes sur plusieurs critères simultanément, en intégrant la variabilité environnementale qui constitue une source d'incertitude.

L'objectif de ce stage est de développer une méthodologie pour réaliser une optimisation multi-critères intégrant des composantes aléatoires décrivant l'environnement de la plante (climat, sol...), d'établir ses propriétés et de quantifier ses performances. Des critères d'intérêt classiques sont par exemple le rendement et la valeur économique. D'autres critères complémentaires peuvent être par exemple l'empreinte carbone et des aspects liés à la morphologie de la plante dans un contexte de culture en mélange ou en association. Cette optimisation simultanée de plusieurs objectifs permettra de sélectionner des variétés à la fois rentables, éco-responsables et robustes dans un contexte de changement climatique.

Les données de rendement disponibles pour le stage proviennent d'une expérience de 220 variétés (appelées également génotypes) de blé observées dans 42 environnements avec des conditions de cultures et climatiques contrastées. Leur génome a été caractérisé (par génotypage dense) ce qui permet de calibrer des modèles pour prédire de nouvelles variétés. Le génotypage consiste à identifier des variations le long de l'ADN.

Objectifs du stage

- formaliser la définition de l'ensemble des génotypes pertinents en multi-critères dans un environnement aléatoire, en généralisant la définition classique du front de Pareto issu du contexte déterministe.
- proposer un estimateur de cet ensemble à partir d'observations des valeurs des critères sur un échantillon d'environnements et établir ses propriétés théoriques
- proposer une méthode numérique de calcul de cet estimateur et la mettre en oeuvre sur les données du projet
- proposer un estimateur de cet ensemble à partir de prédictions des valeurs des critères obtenues via un modèle prédictif et établir ses propriétés théoriques

Des perspectives possibles à plus long terme consisteraient à étendre la méthodologie développée en la combinant à des modèles de prédiction génomique de la littérature pour pouvoir considérer des génotypes supplémentaires non observés.

Aspects mathématiques et informatiques L'approche envisagée reposera sur l'optimisation multicritère et sur la formalisation probabiliste de l'aléa de l'environnement, en intégrant la question biologique visée. Les aspects théoriques feront intervenir des concepts ensemblistes et des probabilités. Les aspects appliqués feront intervenir des méthodes statistiques et des outils algorithmiques. L'essentiel de ces développements devrait idéalement être effectué en Python.

Profil recherché

Formation niveau BAC+5 (Master 2 ou école d'ingénieurs), connaissance en mathématiques appliquées et en informatique, ayant un intérêt pour le travail à l'interface et les applications en sciences du vivant ; maîtrise d'un langage de programmation indispensable ; rigueur scientifique, curiosité intellectuelle, facilité de communication.

Modalités pratiques

Le stage s'inscrit dans le cadre du projet ANR Stat4Plant. Il se déroulera au centre INRAE de Jouy-en-Josas dans l'unité MaIAGE. La durée du stage sera de cinq ou six mois, entre février et septembre 2023. La gratification mensuelle est d'environ 550 euro (taux légal). L'encadrement sera réalisé par Estelle Kuhn (MaIAGE, INRAE) et Jean-Benoist Leger (Heudiasysc, UTC ; MIA Paris-Saclay, INRAE). La partie appliquée du stage se fera en forte interaction avec Renaud Rincent, généticien de l'unité INRAE Génétique Quantitative et Évolution Le Moulon à Gif-sur-Yvette, impliqué dans le projet Stat4Plant sur des données de blé.

Références bibliographiques

- [1] Ehrgott, M. (2005). Multicriteria optimization (Vol. 491). Springer Science & Business Media.
- [2] Feller W., (1968) An Introduction to Probability Theory and Its Applications, Wiley, 3rd edition.
- [3] Bickel P.J., Doksum K.A. (2001) Mathematical statistics, Prentice Hall.
- [4] Rincent R., Malosetti M., Ababaei B., Touzy G., Mini A., Bogard M., Martre P., Le Gouis J., van Eeuwijk F.. (2019) Using crop growth model stress covariates and AMMI decomposition to better predict genotype-by-environment interactions. *Theor Appl Genet*, 12 (132) 3399-3411.
- [5] Akdemir, D., Beavis, W., Fritsche-Neto, R., Singh, A. K., Isidro-Sánchez, J. (2019). Multi-objective optimized genomic breeding strategies for sustainable food improvement. *Heredity*, 122(5), 672-683.