

Construction de nouveaux critères multi-objectifs pour la sélection variétale intégrant la variabilité environnementale

Stage niveau Master 2, début 2022

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 pour appréhender ces questions est celle de l'interaction entre la plante et son environnement au sens large (conditions météorologiques, conditions de sol, présence de ravageurs, conduite de culture de l'agriculteur). Comment les variétés cultivées actuellement vont-elles s'adapter aux effets du changement climatique, en particulier aux situations de stress ? Quelles sont les variétés les plus robustes et les plus résilientes les plus à même de s'adapter aux nouvelles conditions ?

L'objectif de ce stage est de construire de nouveaux critères pour la sélection variétale, en tenant compte de plusieurs objectifs simultanément. On pourra par exemple chercher à maximiser le rendement et la surface foliaire de la plante. Une autre cible serait de tenir compte de la variabilité environnementale, par exemple en considérant le rendement dans différentes conditions environnementales. La prise en compte de plusieurs objectifs simultanément permettra de sélectionner des variétés robustes, par exemple dans plusieurs environnements ou dans des environnements extrêmes dans un contexte de changement climatique.

Les données de rendement disponibles 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

- identifier les méthodologies de construction de critères dans un contexte multi-objectifs adaptées au contexte d'étude fondées sur une élicitation de préférence dans un cadre où les objectifs sont parfaitement connus,
- après avoir identifié le front de Pareto sur les données réelles de rendement des génotypes observés, mettre-en-œuvre une ou plusieurs de ces méthodes sur ces données,
- formaliser la généralisation du front de Pareto dans un contexte où les objectifs sont entachés d'erreurs,
- proposer une méthodologie de construction de critères dans un contexte multi-objectifs lorsque les objectifs sont entachés d'erreurs, la mettre en œuvre sur un jeu de données réelles,
- apporter des garanties théoriques sur la méthodologie proposée,
- étudier la robustesse de la méthodologie proposée en cas d'erreur dans le processus d'élicitation et de construction de l'objectif, en simulation et en théorie,
- 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 l'élicitation de préférence pour toutes les parties concernant les données supposées parfaitement connues. Dans un cadre entaché d'erreur, les erreurs pourront être modélisées par des lois aléatoires (en première approche par des lois normales ou supposées bornées). L'obtention de règles de décision pour la construction de critères combinés dans un contexte multicritère pourra faire appel à des méthodes statistiques et à la théorie de la décision. Les garanties théoriques pourront être obtenues à partir de raisonnements en probabilité ou à partir d'un formalisme ensembliste. Du point de vue informatique, la détermination du front de Pareto dans un contexte connu et inconnu fera appel à de l'optimisation combinatoire et devra faire l'objet de développements informatiques spécifiques. 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 principalement au centre INRAE de Jouy-en-Josas dans l'unité MaIAGE et de façon ponctuelle dans l'unité Heudiasysc de l'Université de Technologie de Compiègne (UTC). La durée du stage sera de cinq ou six mois, entre février et septembre 2022. 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). 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] Emmerich, M. T., Deutz, A. H. (2018). A tutorial on multiobjective optimization : fundamentals and evolutionary methods. *Natural computing*, 17(3), 585-609.
- [2] 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.