

Modélisation et analyse de spectres proche infra-rouge via des modèles déformables pour la prédiction de phénotypes chez les plantes

Stage niveau M2
début 2022 durée de 5 à 6 mois

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

Pour postuler, merci d'envoyer un **dossier complet** contenant **vos CV, votre lettre de motivation et vos derniers relevés de notes de Master 1 ou équivalent** à estelle.kuhn@inrae.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, ...). Comment adapter les variétés cultivées aux effets du changement climatique, en particulier aux situations de stress? Peut-on identifier des leviers d'action biologiques afin de sélectionner les variétés adaptées aux nouvelles conditions?

En amélioration des plantes, on réalise des croisements entre lignées élites d'une espèce de manière à générer de nouvelles variétés qui peuvent potentiellement combiner des caractéristiques intéressantes. Ces variétés candidates sont observées aux champs de manière à évaluer leurs performances (productivité, tolérance aux stress ...). Pour des raisons de budget et de contraintes de temps, il n'est pas possible d'observer en détail toutes les variétés générées à chaque nouveau cycle de croisement. Des approches prédictives ont été développées pour prédire les performances sur la base de caractérisation du génome des plantes (sélection génomique, Meuwissen et al. 2001), ou sur la base de spectroscopie proche infrarouge de tissu prélevé sur les plantes (SPIR, sélection phénotypique, Rincent et al. 2018). La SPIR consiste à mesurer à différentes longueurs d'ondes la proportion de lumière absorbée par des échantillons de tissu prélevés sur chacune des variétés. Les spectres obtenus rendent compte de la composition chimique du tissu et sont influencés à la fois par la génétique de la plante et par les conditions environnementales dans lesquelles elle pousse. Ils permettent en particulier d'estimer des similarités entre variétés. Cette approche s'est révélée efficace pour prédire des caractères complexes chez de nombreuses espèces. Une des limites de la sélection phénotypique vient du fait que chaque année les spectres sont collectés dans des conditions environnementales différentes et qu'il est difficile de combiner ces différents spectres dans un unique modèle. De nouvelles approches statistiques sont nécessaires pour permettre de prédire les spectres de toutes les variétés dans une vaste gamme d'environnements. Ces prédictions de spectres seraient utilisées dans un deuxième temps pour prédire les caractères d'intérêts comme par exemple le rendement.

Objectifs du stage

- * proposer un modèle statistique pour analyser les spectres,
- * proposer un estimateur des paramètres du modèle

- * développer et implémenter un algorithme d'estimation et le valider sur données simulées,
- * ajuster le modèle proposé aux données réelles,
- * valider les prédictions obtenues sur les données réelles.

Aspects mathématiques L'approche envisagée pour la modélisation statistique reposera sur une extension des modèles déformables d'analyse d'image qui pourront rendre compte des variabilités génotypiques et environnementales des spectres observés. L'estimation des paramètres du modèle pourra se faire via des algorithmes stochastiques. Les prédicteurs seront construits à partir du modèle ajusté. Une étude comparative sur la performance des prédictions avec d'autres méthodes de la littérature sera réalisée.

Profil recherché

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

Modalités pratiques

Le stage se déroulera principalement au centre INRAE de Jouy-en-Josas dans l'unité MaIAGE et de façon ponctuelle dans l'unité INRAE GQE Le Moulon située à Gif-sur-Yvette. 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 (INRAE, MaIAGE), Renaud Rincet et Tristan Mary-Huard (INRAE, GQE Le Moulon).

Références bibliographiques

- [1] Allasonnière S., Kuhn E., Trouvé A.. Construction of Bayesian deformable models via a stochastic approximation algorithm: A convergence study. *Bernoulli*, Vol. 16, No. 3, pp 641-678, 2010.
- [2] Allasonnière S. and Kuhn E., Convergent Stochastic Expectation Maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation. *Computational Statistics and Data Analysis*, Vol. 91, pp 4-19, 2015.
- [3] Kuhn E., Matias C., Rebafka T., Properties of the Stochastic Approximation EM Algorithm with Mini-batch Sampling. *Statistics and Computing*, Vol. 30, pp 1725–1739, 2020.
- [4] Meuwissen T., Hayes B., Goddard M.E., Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, Vol. 157, pp 1819–1829, 2001 (<https://doi.org/10.1093/genetics/157.4.1819>).
- [5] Rincet R., Charpentier J-P, Faivre-Rampant P., et al, Phenomic Selection Is a Low-Cost and High-Throughput Method Based on Indirect Predictions: Proof of Concept on Wheat and Poplar. *G3 Genes/Genomes/Genetics*, 2018 (<https://doi.org/10.1534/g3.118.200760>)