

Modélisation et estimation statistique d'un modèle Resource Balance Analysis (RBA) chez *Arabidopsis Thaliana*

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 Anne Goelzer (anne.goelzer@inrae.fr).

Contexte

Le développement de modèles mathématiques capables de prédictions quantitatives du comportement des plantes en conditions de stress est un enjeu majeur en biologie. Un premier modèle, basé sur l'allocation parcimonieuse des ressources de la plante entre les différentes fonctions cellulaires a récemment été développé par l'équipe BioSys du laboratoire MaIAGE (INRAE) pour décrire la plante *Arabidopsis thaliana* (appelé RBA, voir [1]). Ce modèle semble très prometteur pour prédire la réponse de cette plante en condition de stress. Il correspond à un problème d'optimisation convexe de taille importante. En collaboration avec plusieurs équipes de biologistes, de bioinformaticiens et biostatisticiens de l'INRAE, il s'agit maintenant de développer les méthodes statistiques adéquates pour estimer les paramètres du modèle, tout en tenant compte des contraintes expérimentales (données manquantes, données censurées, données agrégées, etc). Pour cela, deux jeux de données omiques et phénotypiques sont disponibles (Jeu 1) ou en cours d'acquisition (Jeu 2) pour différentes conditions de croissance d'*Arabidopsis thaliana* :

- Jeu 1 : protéomique quantitative, traits phénotypiques pour 6 réplicats biologiques, 3 types de tissus, 2 stades de développement de la plante
- Jeu 2 : transcriptomique, métabolomique, protéomique quantitative, phénotypage pour 3 réplicats biologiques, un stade de développement, 9 conditions environnementales combinant contraintes hydriques et azotées, et 5 génotypes.

Objectifs du stage

- proposer une modélisation stochastique pour le problème de calibration des paramètres du modèle RBA chez la plante
- développer une méthode numérique adaptée pour l'estimation des paramètres et pour la construction d'intervalle de confiance

- mettre en oeuvre la méthodologie développée sur les données réelles de Arabidopsis thaliana
- valider la qualité des estimations de paramètres via celles des prédictions obtenus par le modèle

L'approche envisagée pourra dans un premier temps faire intervenir des modèles statistiques à effets mixtes pour prendre en compte dans le modèle les variabilités présentes dans les données. L'estimation des paramètres sera réalisée via des algorithmes numériques adaptés et efficaces. Les intervalles de confiance seront construits soit par bootstrap soit via des propriétés asymptotiques.

Profil recherché

Formation niveau BAC+5 (Master 2 ou école d'ingénieurs), connaissance nécessaire en mathématiques appliquées et statistiques, connaissance souhaitée en génétique, 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 se déroulera au centre INRAE de Jouy-en-Josas dans l'unité MaIAGE. La durée du stage sera de cinq ou six mois, à partir de janvier 2023. La gratification mensuelle est d'environ 550 euro (taux légal). L'encadrement sera réalisé par Estelle Kuhn et Anne Goelzer (MaIAGE, INRAE).

Références bibliographiques

- [1] Goelzer, A., Muntel, J., Chubukov, V., Jules, M., Prestel, E., Nölker, R., ... , Fromion, V. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic engineering*, 32, 232-243.
- [2] Davidian, M., Giltinan, D. M. (2003). Nonlinear models for repeated measurement data : an overview and update. *Journal of agricultural, biological, and environmental statistics*, 8(4), 387-419.
- [3] Kuhn E., Matias C., Rebafka T. (2020) Properties of the Stochastic Approximation EM Algorithm with Mini-batch Sampling, *Statistics and Computing*, Vol. 30, pp 1725–1739.