

Modélisation bayésienne flexible pour l'analyse jointe de données longitudinales et de survie : application à l'étude de la propagation d'une maladie dans le contexte du palmier à huile

Que cela soit dans les domaines de la santé, de l'agronomie ou de la génétique, de nombreuses études permettent de collecter des données mesurées à plusieurs pas de temps sur un même individu (données longitudinales) jusqu'à l'apparition ou non d'un évènement d'intérêt (données de survie). En médecine, l'analyse jointe de ces deux types de données est devenue un outil incontournable pour étudier l'association entre l'évolution de biomarqueurs longitudinaux et la survenue d'un évènement (Gould et al., 2015). Récemment des développements ont été proposés afin d'obtenir une compréhension plus fine de relations complexes entre, par exemple, la survenue d'une maladie et l'évolution d'un caractère d'intérêt agronomique. Ces développements, proposés dans un contexte bayésien (Köhler et al., 2017), permettent une modélisation plus flexible des trajectoires longitudinales notamment au travers de l'utilisation de P-splines (Marx and Eilers, 1999).

Dans le contexte du palmier à huile, on souhaite comprendre les liens entre la propagation d'une maladie due au *Ganoderma* et les caractères liés à la production d'huile : un palmier produit-il moins quand il est infecté ? Comment la production d'huile évolue-t-elle au cours du temps entre un palmier infecté et un palmier non-infecté ? Dans ce contexte, l'objectif de ce travail de recherche appliquée consistera, dans un premier temps, à s'appropriier les concepts liés à l'analyse jointe et à prendre en main le package *bamlss* (Köhler et al., 2017) du logiciel libre R. Dans un second temps, un modèle joint permettant d'intégrer une composante génétique et une composante spatiale sera développé et mis en œuvre en utilisant les modèles du package *bamlss*.

Les modèles proposés seront appliqués dans un contexte d'amélioration génétique du palmier à huile sur des données issues d'essais génétiques mis en place par le CIRAD (Centre de coopération internationale en recherche agronomique pour le développement) et PalmElit, sa filiale pour la sélection et la production de semences.

Contacts encadrants : Marie Denis (marie.denis@cirad.fr)

Durée : 4 mois ou 6 mois

Indemnité mensuelle : 550 euros

Thématiques : analyse de survie, données longitudinales, modèles joints, modèles linéaires mixtes, méthodes Bayésiennes

Lieu du stage : Montpellier, CIRAD Lavalette, UMR AGAP

Lawrence Gould, A., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., & Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in medicine*, 34(14), 2181-2195.

Köhler, M. (2017). Flexible Bayesian joint models for longitudinal biomarkers and time-to-event outcomes with applications to type 1 diabetes research (Doctoral dissertation, Imu).

Marx, B. D., & Eilers, P. H. (1999). Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics*, 41(1), 1-13.