

Implémentation et développement de modèles spatio-temporels pour l'étude de la propagation de maladies infectieuses

Que ce soit dans les domaines de la santé, de l'agronomie ou bien de l'écologie, de nombreuses études ont été mises en œuvre pour étudier la propagation de maladies infectieuses dans l'espace et dans le temps. Le cadre hiérarchique bayésien est couramment utilisé pour traiter des données spatio/temporelles (Berliner, 1996; Arab et al., 2008). Ces approches permettent une modélisation flexible de processus complexes en considérant simultanément la variabilité de l'échantillonnage, les incertitudes sur les paramètres et les structures de dépendances entre les observations.

Dans le contexte de l'étude de maladies infectieuses chez le palmier à huile deux modèles hiérarchiques bayésiens ont été développés (Denis et al., 2017). L'inférence des paramètres a été obtenue avec l'approche « Integrated Nested Laplace Approximation » et le package « INLA » du logiciel libre R (Rue et al., 2009). Ce package est un outil flexible qui a permis de mettre en œuvre les différentes approches. Cependant son utilisation a mis en évidence deux limitations. La première réside dans la considération d'une structure de voisinage constante dans le temps. La seconde porte sur la modélisation des paramètres associés aux processus autorégressifs avec l'estimation d'un effet commun à l'ensemble des positions spatiales. Cette modélisation ne permet pas de traduire les différents mécanismes d'infection possibles à différentes localisations.

Le stage a pour objectif, dans un premier temps, d'implémenter et d'étendre les modèles développés avec le package INLA dans le langage de programmation STAN (Carpenter et al., 2017). Ce développement permettra l'estimation de paramètres autorégressifs différents en fonction des positions géographiques. De plus, afin d'estimer une structure de voisinage au cours du temps, des méthodes d'estimation de matrices de précision (Graph Lasso, Friedman et al., 2008) seront explorées.

Les différentes approches développées seront appliquées 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 : 4-6 mois

Indemnité mensuelle : 550 euros

Thématiques : Modèle spatio-temporel, méthodes Bayésiennes, épidémiologie, INLA, STAN

Lieu du stage : Montpellier, CIRAD Lavalette, UMR AGAP

Arab, A., Hooten, M. B., & Wikle, C. K. (2008). Hierarchical spatial models. In Encyclopedia of GIS (pp. 425-431). Springer US.

Berliner, L. M. (1996). Hierarchical Bayesian time series models. In Maximum entropy and Bayesian methods (pp. 15-22). Springer Netherlands.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2016). Stan: A probabilistic programming language. *Journal of Statistical Software*, 20, 1-37.

Denis, M., Cochard, B., Syahputra, I., de Franqueville, H., Tisné, S. (2017). Evaluation of spatio-temporal Bayesian models for the spread of infectious diseases in oil palm. *Spatial and Spatio-temporal epidemiology journal*, in revision.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3), 432-441.