

Proposition de sujet de stage

Méthodes statistiques pour l'inférence du réseau d'interaction d'une zoonose

Laboratoires :

Unité Biométrie et Intelligence Artificielle (INRA Toulouse) et Unité Biostatistique et Processus Spatiaux (INRA Avignon). Le stage se déroulera à Toulouse.

Encadrants :

Nathalie Peyrard	nathalie.peyrard@toulouse.inra.fr	05.61.28.54.39
Rachid Senoussi	rachid.senoussi@avignon.inra.fr	04.32.72.21.84

Description

Le risque d'émergence au sein d'un territoire d'une zoonose, dépendant de contacts interspécifiques concernant l'Animal et l'Homme, est fonction de nombreux facteurs démo-géographiques (zones d'habitats, densités et flux de populations humaines et animales, ...), socio-écologiques (habitudes de vie, modes d'élevages, ...), environnementaux (climat, présence de sites à risque tels que les abattoirs, zones protégées, ...) et épidémiologiques (sujets susceptibles, politique sanitaire,...). Dans le cadre du projet "Cartographie et modélisation spatio-écologique des risques zoonotiques en France à des fins de prévision des épidémies", initié par le consortium ZooRisq, nous nous intéressons à la modélisation spatio-temporelle du risque d'occurrence et de la propagation des zoonoses en fonction des variables considérées comme cofacteurs. La première question à laquelle nous souhaitons répondre est : comment identifier le réseau d'interactions spatiales caractérisant une zoonose donnée? C'est à dire, comment identifier quel site influence quel site, dans le temps et dans l'espace, et quels sont les cofacteurs qui expliquent cette dépendance? Comment avoir une pondération ou pseudo-distance entre sites qui "mesure" cette influence? Ce problème peut être vu comme un problème d'inférence d'un graphe valué à partir des cofacteurs du risque et des cas de maladie. Les nœuds du réseau sont les sites/entités administratives, les arêtes sont pondérées par des indices de corrélation (ou pseudo-distances).

Dans ce cadre, l'objectif du stage proposé est l'étude de méthodes basées sur la régression pour l'inférence du réseau dans le cas où l'on connaît toutes les observations des cas et les cofacteurs. Les cofacteurs peuvent être quantitatifs ou catégoriels, et peuvent être attachés à un site ou à une paire de sites (déplacements). De plus, les observations peuvent être corrélées dans le temps.

Il s'agira donc dans un premier temps de construire un modèle de régression, a priori de type GLM. Puis l'étudiant devra étudier des méthodes d'estimation des coefficients de la régression dans un but de sélection des plus significatives (par exemple Lasso, qui a l'avantage de pénaliser les petits coefficients). Enfin, il s'agira de construire le graphe d'interaction à partir de ces coefficients. La validité de la ou des méthodes proposées pourront être testées sur des données simulées.