



STAGE DE RECHERCHE M2 ECOLOGIE EVOLUTION GENOMIQUE

Rentrée 2016

Estimation des paramètres de diffusion d'une population dans un paysage hétérogène à partir d'équations aux dérivées partielles et de données génétiques spatialisées

INRA Avignon

Encadrants :

Pierre Franck (pierre.franck@inra.fr), Olivier Bonnefon (olivier.bonnefon@inra.fr)

Mots clés et techniques à utiliser au cours du stage :

dynamique des populations ; génétique des populations, équations aux dérivées partielles, analyse numérique (méthode des éléments finis) ; estimation de paramètres par maximum de vraisemblance, algorithme numérique d'optimisation (méthode quasi-Newton), simulation de paysage

Projet de stage :

En écologie du paysage, les équations de réaction-diffusion permettent de modéliser la dynamique spatio-temporelle d'une population en représentant l'évolution temporelle de la densité d'individus $u(t, x)$ en chaque point de l'espace sous l'effet de la dispersion (terme de diffusion) et de la reproduction/mortalité (terme de réaction). Ces équations sont des EDP paraboliques de la forme : $\partial_t u(t, x) = \Delta(D(x)u) + f(x, u)$.

En particulier, dans un paysage hétérogène, un coefficient de diffusion $D(x)$ variable dans l'espace (déterminé par exemple par l'occupation du sol en chaque position) permet de modéliser des mobilités différentes des individus en fonction de l'environnement dans lequel ils se trouvent. Estimer les coefficients de diffusion d'un organisme en fonction de l'hétérogénéité de l'environnement est un enjeu important en écologie du paysage (notamment en termes de biologie de la conservation et de gestion des organismes nuisibles). Différents types d'observations sont habituellement utilisés pour suivre la dynamique spatio-temporelle d'un organisme : comptages, suivi de trajectoires, captures-marquages-recaptures. Dans un travail récemment publié (Roques et al .2016), nous avons développé un modèle statistique pour des dispositifs où les individus collectés dans des pièges spatialisés sont génotypés pour plusieurs marqueurs génétiques (*e.g.* locus microsatellites). Avec des simulations nous avons montré qu'il était possible d'estimer les coefficients de diffusion quand on connaît les fréquences alléliques des marqueurs des populations sources, localisées dans l'espace et qu'elles sont génétiquement différenciées. Cette méthode par maximum de vraisemblance d'estimations des coefficients de diffusion s'appuie sur les proportions d'individus issus de chaque source initiale capturés dans chaque piège après diffusion. Les proportions estimées génétiquement sont comparées aux proportions qui sont calculées à partir d'un système d'équations de réaction-diffusion, une pour chacune des sources initiales. L'objectif de ce stage est d'évaluer les performances de la méthode et de proposer des pistes d'amélioration en vue de son application à des données de génétiques du paysage. Les analyses réalisées à ce jour suggèrent au moins trois pistes d'amélioration possibles, lesquelles pourront être tout ou partie intégrées à la méthode dans le cadre du stage. D'une part, il est rarement possible de connaître exactement les fréquences alléliques dans les sources avant la diffusion, soit parce qu'on n'a échantillonné qu'un

petit nombre d'individus avant la diffusion, soit parce que plusieurs générations se sont succédées entre l'échantillon d'apprentissage des fréquences alléliques et l'échantillon d'individus piégés. D'autre part, quand les sources sont trop faiblement différenciées, la vraisemblance d'appartenance à une source peut ne pas être assez résolutive. Utiliser des vraisemblances d'apparement (plein- frères, demi-frères, non apparentés) entre paires d'individus pourrait être une option pour résoudre ce problème. Finalement le modèle de système d'EDP utilisé pourrait être amélioré pour prendre en compte un terme de convection. Ce stage consistera à écrire les équations de réaction-diffusion, à les mettre en œuvre numériquement sur une mosaïque paysagère stylisée, à adapter les fonctions de vraisemblance pour corriger les faiblesses identifiées ci-dessus, et à étudier les propriétés statistiques des estimateurs sur des jeux de données simulées, en considérant des écarts aux hypothèses du modèle d'observation.

Intérêts. Ce stage conduira l'étudiant(e) à acquérir des connaissances sur les principes de modélisation, la modélisation par EDP, la mise en œuvre numérique d'EDP, les méthodes statistiques associées aux marqueurs génétiques (écologie moléculaire), l'écologie et la génétique du paysage.

Modalités pratiques : Le stage se déroulera sur le site agroparc de l'INRA d'Avignon dans l'unité «Biostatistique & Processus Spatiaux » en étroite collaboration avec l'unité «Plantes & Systèmes de culture Horticoles ». Le stage est financé par le projet ANR « Peerless ». Il est indemnisé selon les conditions standards de gratification à l'INRA (554€ / mois).

Quelques références des équipes sur le sujet :

Franck P, Ricci B, Klein EK, Olivares J, Simon S, Cornuet JM, Lavigne (2011). Genetic inferences about the population dynamics of codling moth females at a local scale. *Genetica*, 139, 949-960. Roques L, Walker E, Franck P, Soubeyrand S, Klein EK (2016). Using genetic data to estimate diffusion rates in heterogeneous landscapes. *Journal of Mathematical Biology*, 73, 397–422. Roques L & Bonnefon O (2015). Modelling population dynamics in realistic landscapes with linear elements: A mechanistic-statistical reaction-diffusion approach. *PLoS One*, 11, e0151217.