

## Offre de thèse

**Intitulé : Modéliser l'abondance de plusieurs espèces en même temps à différents niveaux hiérarchiques : améliorations probabilistes et prise en compte de l'incertitude sur les traits fonctionnels des espèces dans les modèles statistiques paramétriques multi-espèces**

**Unité d'accueil :** UR EFNO, Irstea, Domaine des Barres, F-45290 Nogent-sur-Vernisson

**Direction :** Frédéric Gosselin

**Organisme finançant la thèse :** Région Centre Val de Loire et Irstea

**Début :** 01/10/2016

**Durée :** 3 ans

### **Descriptif du projet de thèse :**

La modélisation statistique paramétrique ou semi-paramétrique de la biodiversité est en plein essor. En guise d'exemple, les modèles de distribution d'espèce ont permis de modéliser la réponse des espèces à différents aspects des changements globaux (changement climatique, changement d'occupation du sol...) et fournir ainsi de précieux renseignements aux gestionnaires ou décideurs.

Un des développements récents consiste à analyser simultanément toutes les espèces, en cherchant à expliquer la réponse de ces espèces à des variables environnementales ou à des pressions anthropiques – par exemple liés aux changements globaux –, par leurs attributs fonctionnels et écologiques. Ces modèles ont l'avantage de pouvoir porter dans le même outil une vision plurielle de la biodiversité – chaque espèce pouvant répondre aux variables explicatives de manière différente – et une vision plus intégrée et fonctionnelle – par les réponses estimées au niveau du groupe fonctionnel. Ces modèles ont ainsi le potentiel de nous indiquer si la biodiversité répond de manière plurielle et peu organisée aux changements globaux ou si au contraire la réponse est plus univoque. Toutefois, ces modèles font très souvent de nombreuses hypothèses – par exemple sur l'indépendance des observations ou sur la forme de relation entre les variables explicatives et la moyenne de l'abondance des espèces. Or, ces hypothèses faites sont très rarement vérifiées. De plus, l'incertitude sur les attributs fonctionnels des espèces n'est à ce stade pas prise en compte et pourrait avoir un impact important sur les conclusions obtenues en termes plus fonctionnels. Dans ce contexte, l'objectif de cette thèse est triple :

(i) développer une structure de modèle plus générale que les structures actuelles dans le cas de données de comptage. En effet, la plupart des modèles multi-espèces actuels ne modélisent que la présence des espèces. Or l'abondance de ces espèces est souvent une donnée disponible qu'il est dommage de ne pas prendre en compte. Il s'agit ici de proposer des modèles pour modéliser des données d'abondance ;

(ii) proposer des méthodes d'évaluation des hypothèses faites dans le modèle (dans différentes composantes du modèle) ;

(iii) de développer des modèles prenant en compte l'incertitude sur les attributs fonctionnels des espèces.

Pour les points (i) et (iii) nous étudierons l'impact de choix probabilistes sous-optimaux sur les conclusions tirées du modèle. Ces différents points seront appliqués sur le jeu de données existant, issu de la thèse d'Emmanuelle Dauffy-Richard (décédée en 2014) et permettront de continuer le travail entrepris avec elle sur les modèles multi-espèces durant sa thèse et après sa thèse. Les points (i) et (ii) font suite à des avancées acquises ou en cours dans le cadre de notre équipe de recherche, permettant de les valoriser davantage.

**Abstract.** The use of parametric or semi-parametric models to analyze biodiversity is developing. One of the recent advances has been to analyze in the same Bayesian model all the species, with the aim of explaining the response of species to environmental variables by their functional or ecological traits. Yet, the probabilistic quality of these models is rarely checked and could be important due to

the numerous assumptions made in such models. Furthermore, the uncertainty related to the functional traits of species is currently not taken into account and could have an important impact on the conclusions drawn from such models. The objectives of this PhD project are:

- (i) To develop a structure of models more general than the current models available in the case our abundance count data;
- (ii) To propose methods to diagnose the goodness of fit of different components of these parametric models;
- (iii) To develop models that take into account the uncertainty of functional traits of species. For points (i) and (iii), the impact of worse choices in terms of structure on the conclusions drawn from the model will be studied. A fourth more exploratory point will be to try to propose new methods of beta diversity quantification based on these models. These points will be applied on the PhD data of Emmanuelle Dauffy-Richard and shall allow us to continue the work in progress on multi-species models during her PhD and after her PhD.

**Profils recherchés :** Le candidat devra être de formation soit mathématique – avec une sensibilité pour l'écologie et les applications – soit biologique/écologique – avec un goût pour les outils statistiques et une solide formation associée. Connaissances en statistiques et dans un langage de programmation comme C, R ou Julia indispensables. L'approche proposée confèrera au candidat des compétences à l'interface entre statistiques appliquées (et plus généralement biométrie) et écologie.

Les candidatures devront être envoyées par courrier électronique avant le 10/06/2016 à Frédéric Gosselin ([frederic.gosselin@irstea.fr](mailto:frederic.gosselin@irstea.fr)). Elles comprendront un CV détaillé, une lettre de motivation, une copie du Master 2 (ou du Master 1 si le Master 2 est en cours) avec indication de mention et/ou notes obtenues et au minimum une lettre de recommandation.

### ***Publications de l'équipe sur des thématiques proches :***

Gosselin, F., 2011. Propositions pour améliorer l'équipement biométrique du détective écologique. Application à la modélisation de la relation entre gestion forestière et biodiversité. Mémoire de HDR, Université Pierre et Marie Curie, Paris.

Gosselin, F., 2011. A New Calibrated Bayesian Internal Goodness-of-Fit Method: Sampled Posterior p-values as Simple and General p-values that Allow Double Use of the Data. *Plos One*, 6(3), e14770.

Saas, Y. and F. Gosselin, 2014. Simulation-based comparative analysis of spatial count regression methods on regularly and irregularly-spaced locations. *Ecography*, 37(5), 476-489.

Gimenez, O., Buckland, Stephen T., Byron J.T. Morgan, Nicolas Bez, Sophie Bertrand et al., 2014. Statistical ecology comes of age. *Biology Letters*, 10(12), 20140698.

Gosselin, F. 2015. Reevaluating Europe's other debt with improved statistical tools. *Biodiversity and Conservation* 24(1):205-211. doi:10.1007/s10531-014-0799-y.

Herpigny, B., and Gosselin, F. 2015. Analyzing plant cover class data quantitatively: customized cumulative zero-inflated beta distributions show promising results. *Ecological Informatics* 26(3):18-26. doi:10.1016/j.ecoinf.2014.12.002.