

Modélisation statistique de la distribution spatiale de populations animales à partir de multiples jeux de données, appliquée aux papillons

Les observations naturalistes dites opportunistes (obtenues par iNaturalist, eBird, INPN espèces, GBIF,...) sont maintenant disponibles en masse dans diverses bases de données online. Par opposition aux données de présence / absence répondant à un protocole rigoureux, ces données présentent des biais spatiaux et taxonomiques souvent importants : les observations peuvent refléter plus l'effort de prospection dans des zones d'intérêt pour les naturalistes que l'abondance ou la diversité en espèces réelles. A l'inverse, les données protocolées de présence-absence ou d'abondance sont beaucoup plus faciles à analyser et bien moins biaisées, mais elles sont souvent en bien moins grande quantité également. Cela rend l'interpolation et l'extrapolation spatiale des indicateurs de biodiversité calculés avec les données protocolées très délicate, voire impossible à fine échelle spatiale, e.g., en dessous du km^2 .

Pour résoudre ce dilemme classique entre quantité et qualité des données, la littérature à l'interface entre statistique et écologie a mis en évidence la possibilité de combiner les données naturalistes opportunistes avec des données protocolées dans des modèles dits "intégrés" (Giraud et al. 2016, Pacifici et al. 2017, Isaac et al. 2020). Cela permet d'améliorer la prédiction spatiale, notamment en utilisant des processus de points (synthèse dans Isaac et al. 2020). Les données protocolées permettent alors de "corriger" les données opportunistes et d'estimer la probabilité de présence réelle des espèces, souvent à l'aide de variables latentes. Nous proposons d'émuler le travail de Fithian et al. (2015) à base de processus de points spatiaux, afin de prédire les tendances spatiales de diversité des papillons de jour sur la métropole bordelaise, pour lesquelles nous disposons à la fois de données opportunistes et de données protocolées.

Les techniques d'intégration de données sont le plus souvent utilisées à grande échelle spatiale (e.g., sur des cartes au maillage d'un 1 à 100 km^2), mais les processus de points sont idéaux pour considérer des échelles spatiales bien inférieures, ce qui sera nécessaire pour appliquer ces techniques au milieu urbain et péri-urbain qui sont très hétérogènes spatialement. Le stage a pour objectif de (1) tester la faisabilité de cette combinaison à fine échelle spatiale de données opportunistes et protocolées, sur l'exemple concret des papillons en milieu (péri)urbain, ainsi que plus généralement (2) quantifier l'amélioration de l'identifiabilité permise par la combinaison de données, qui est dans ce contexte inconnue. Cette seconde tâche fera appel autant à la théorie statistique qu'à la simulation de données ressemblant aux données réelles.

Supervision Frédéric Barraquand (CNRS), frederic.barraquand@u-bordeaux.fr.

Pratique Début à compter de janvier 2023, à l'Université de Bordeaux (Institut de Mathématiques de Bordeaux), gratification ~ 600 euros/mois.

Références

- Fithian, W., Elith, J., Hastie, T., & Keith, D. A. (2015). Bias correction in species distribution models: pooling survey and collection data for multiple species. *Methods in Ecology and Evolution*, 6(4), 424-438.
- Giraud, C., Calenge, C., Coron, C. & Julliard, R. (2016). Capitalizing on opportunistic data for monitoring relative abundances of species. *Biometrics*, 72, 649-658.
- Isaac, N. J., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., ... & Jarvis, S. (2020). Data integration for large-scale models of species distributions. *Trends in Ecology and Evolution*, 35(1), 56-67.
- Pacifici, K., Reich, B. J., Miller, D. A., Gardner, B., Stauffer, G., Singh, S., ... & Collazo, J. A. (2017). Integrating multiple data sources in species distribution modeling: a framework for data fusion. *Ecology*, 98(3), 840-850.