

# Proposition de stage de master : Identifiabilité des modèles démographiques ajustés à de multiples sources de données

Les paramètres des modèles démographiques pour les populations animales sont souvent difficiles à estimer à partir des données les plus communes, les comptages des nombres d'individus. Cela a motivé statisticiens et écologues à formuler des modèles qui intègrent plusieurs sources de données (comptages, fécondité, survie) dans la vraisemblance à maximiser [1], similaires aux modèles à "vraisemblance conjointe" de l'épidémiologie. En dépit de ces développements, on a surprenamment peu d'évaluations formelles de l'accroissement de l'identifiabilité (qui se manifeste par un jeu de paramètres estimés unique) et de la qualité de l'estimation permise par cette combinaison de données [3,4]. En parallèle, il a été montré que certains modèles démographiques pouvaient être ajustés sur des données de comptages seules, pourvu que l'on introduise des covariables pertinentes [4]. Il y a donc un manque de connaissances sur les conditions qui nécessitent la combinaison de plusieurs sources de données, sachant que ces données nécessitent beaucoup de temps d'acquisition.

En conséquence, le stage proposé vise à mieux comprendre, pour des modèles nonlinéaires simples à deux classes d'âge, dans quelles conditions la fusion de plusieurs sources de données est requise pour obtenir des paramètres de survie et de fécondité identifiables, ainsi que des estimateurs pour ces paramètres aux propriétés satisfaisantes. On considèrera des modèles dans lesquels la quantité d'aléa sur la fertilité et la mortalité sera variée, avec une structure générale

$$\mathbf{n}(t+1)|(\mathbf{n}(t), \xi_t) \sim \mathcal{D}(\mathbf{A}[\mathbf{n}(t), \xi_t]\mathbf{n}(t); \theta_{\mathcal{D}}) \quad (1)$$

où  $\mathbf{n}(t) = (n_J(t), n_A(t))^T$  est le vecteur des densités de population juvénile et adulte au temps  $t$ ,  $\mathcal{D}$  est une certaine distribution de probabilité (Poisson, log-normale, ...) et  $\xi_t$  un vecteur de paramètres (e.g., fertilité maximale, effets de la densité) qui fluctuent au cours du temps. La matrice  $A[\mathbf{n}, \xi]$  traduit quant à elle la démographie et le passage d'une classe d'âge à l'autre:

$$A[\mathbf{n}, \xi] = \begin{pmatrix} (1 - \gamma)s_J(\mathbf{n}, \xi) & f(\mathbf{n}, \xi) \\ \gamma s_J(\mathbf{n}, \xi) & s_A \end{pmatrix}. \quad (2)$$

La fertilité  $f$  et la survie juvénile  $s_J$  sont rendues aléatoires par le vecteur  $\xi$ , qui suit un processus stochastique  $(\xi_t)_{t \in \llbracket 1, T \rrbracket}$  dont on souhaite estimer les paramètres  $\theta_{\xi}$ . On considèrera entre autres le cas simple  $\xi_{t,i} = a_i + b_i x_{t,i}$  où  $x_t$  est un vecteur de covariables connu. Le stagiaire déterminera l'identifiabilité théorique et pratique des modèles [5], sous des scénarios réalistes d'acquisition de données, et ajustera ces modèles à des données de comptages, fécondité, et/ou capture-recapture provenant d'une population d'oiseaux.

**Profil de l'étudiant.e :** Master en statistique, mathématiques appliquées, ou modélisation pour la biologie / Goût pour l'écologie et la programmation (possible de travailler sur l'identifiabilité analytiquement néanmoins).

**Supervision :** Frédéric Barraquand (Institut de Mathématiques de Bordeaux, CNRS).

**Pratique :** Le stage pourra commencer dès début 2022 (gratification  $\sim 600$  euros par mois), et prendra place à l'Institut de Mathématiques de Bordeaux.

**Contact :** frederic.barraquand@u-bordeaux.fr

## Références

- [1] Besbeas, P., Freeman, S. N., Morgan, B. J., & Catchpole, E. A. (2002). Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58(3), 540-547.
- [2] Cole, D. J., & McCrea, R. S. (2016). Parameter redundancy in discrete state-space and integrated models. *Biometrical Journal*, 58(5), 1071-1090.
- [3] Barraquand, F., & Gimenez, O. (2019). Integrating multiple data sources to fit matrix population models for interacting species. *Ecological modelling*, 411, 108713.
- [4] Polansky, L., Newman, K. B., & Mitchell, L. (2020). Improving inference for nonlinear state-space models of animal population dynamics given biased sequential life stage data. *Biometrics*. doi:10.1111/biom.13267
- [5] Cole, D. (2020). *Parameter Redundancy and Identifiability*. CRC Press.