Latent variable models for ecology and evolution

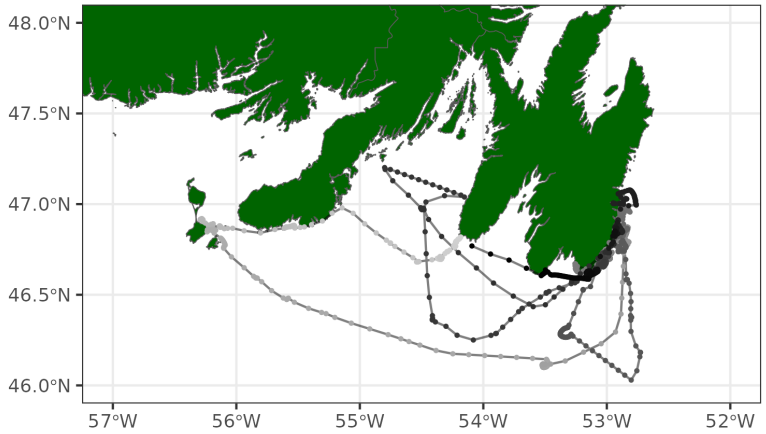
Presentation

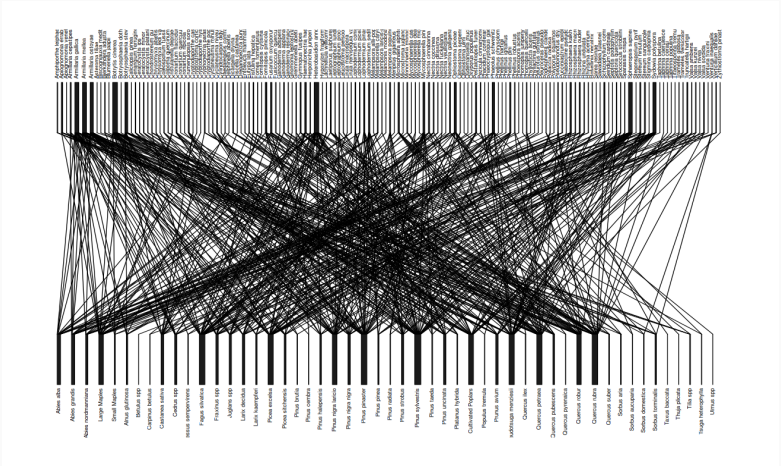
Sophie Donnet. INRAE

Master 2 MathSV. December 16, 2025

Ecological context

Ecology requires to handle complex phenomena and dependencies





Objective Propose models that take into account these dependencies.

Objectives

- **Purpose of this lecture**: present a series of statistical models involving hidden (also called latent) variables used in ecology and evolution:

$$\begin{aligned} Y|Z &\sim p_{\theta^{obs}}(Y|Z) \qquad \text{Observed} \\ Z &\sim p_{\theta^{lat}}(Z) \qquad \text{Hidden} \end{aligned}$$

- Estimation of the parameters

EM algorithm and extensions

$$\widehat{\theta} = \arg\max_{\theta} \log \ell(Y;\theta)$$

$$\log \ell(Y;\theta) = \log \int_Z p_{\theta^{obs}}(Y|Z) p_{\theta^{lat}}(Z) dZ$$

- Likelihood Expectation can be difficult to compute because of the integral form
- Resort to Expectation-Maximisation algorithm [Dempster et al., 1977] and extensions (Variational EM, Monte-Carlo EM, etc...)
- Model selection criterion

Examples

- **Mixture models** and EM
  - Gaussian mixture models: sizes of animals
  - Zero inflated Poisson : abundances for biodiversity
- **Hidden Markov models** and EM
  - Trajectories of animals (movement ecology)
  - Gaussian models for traits evolution
- **Stochastic Block models** and VEM
  - Parasitism network , pollination networks
  - Gaussian models for traits evolution
- **Poisson log-normal** and VEM
  - Dependent couting data : abundances of sereral interdependent species
- **Variational auto-encoder** and VEM
  - Deep learning

Mercredi après midi

14:00 - 17:00. Campus Agro Palaiseau,

- 07/01/2026: Room C1.0.21
- 14/01/2026: Room C1.0.21
- 21/01/2026: Room C1.0.21
- 04/02/2026: Room C1.0.21
- 11/02/2026: Room C1.0.21
- 18/02/2026: Room C1.1.27
- 25/02/2026: Room C1.1.27

Examen écrit de 3h le 25/03/2026 (sans documents)

Ressources

https://sophiedonnet.github.io/LatentVariableModelsCourse/

Le manuel sera en ligne

References

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977).
Maximum likelihood from incomplete data via the EM algorithm.
Jr. R. Stat. Soc. B, 39:1–38.