

Etudiant : Arnaud Becheler
Directeur : Stéphane Dupas (EGCE)
Co-encadrante : Camille Coron (LMO)

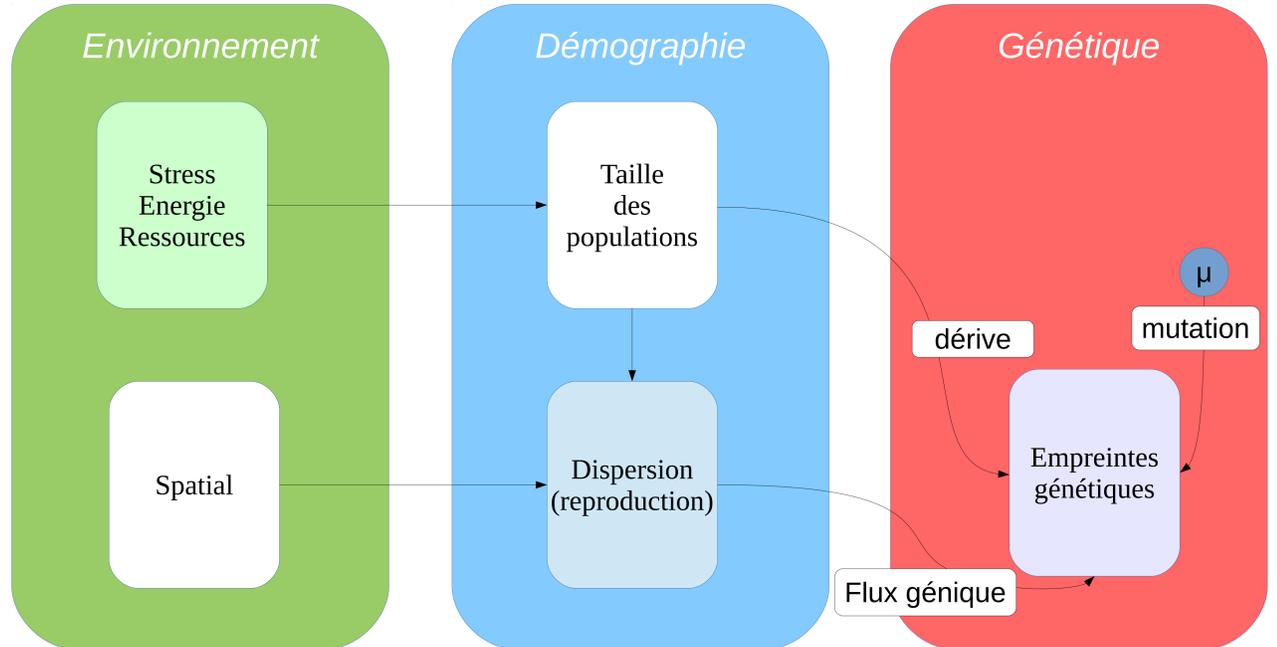
INTRODUCTION

La réponse des espèces aux changements environnementaux est au cœur de nombreuses préoccupations sociétales.

Les facteurs environnementaux affectent les caractéristiques démographiques des populations, ce qui laisse des empreintes génétiques plus ou moins aisément identifiables.

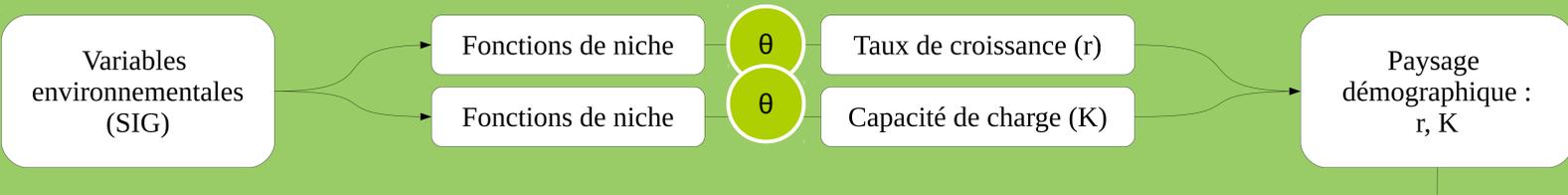
Les marqueurs génétiques se font donc mémoire d'une histoire complexe, et les patrons génétiques populationnels au temps présent permettent d'inférer l'histoire passée.

La modélisation stochastique de ces processus permettrait non seulement de comprendre l'histoire des populations à partir de données génétiques et environnementales, mais aussi de mieux projeter leurs réponses face aux changements paysagers actuels.

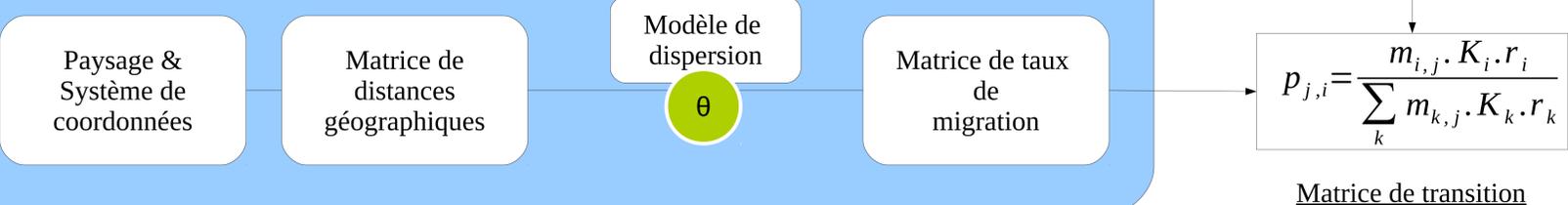


MODÈLE GÉNÉRATIF

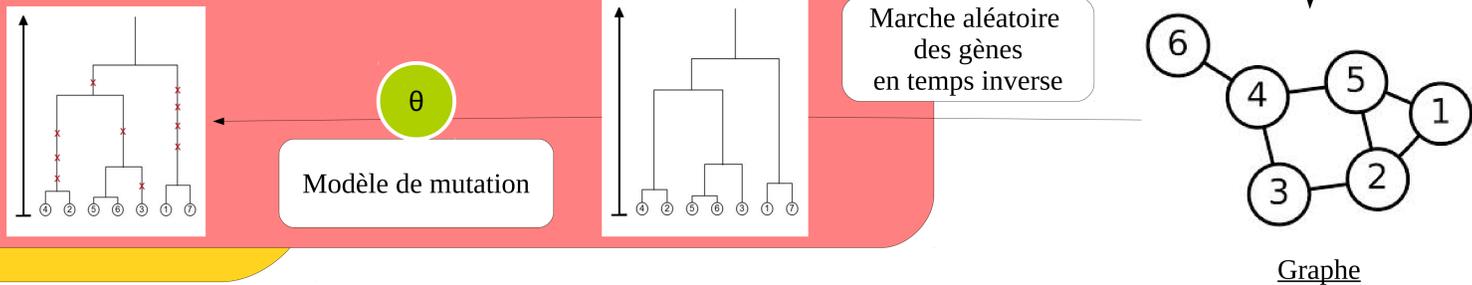
Modèles de niche : relier l'environnement à la démographie



Modèle de dispersion : rendre compte du mouvement des populations



Modèle de mutation pour générer de la différenciation génétique



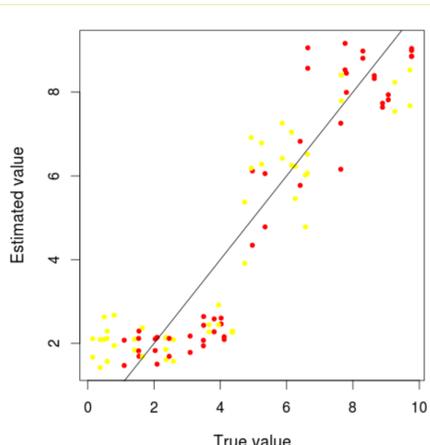
ESTIMATION DES PARAMÈTRES PAR CALCUL BAYÉSIEN APPROCHÉ (ABC)

On possède des données génétiques géoréférencées de *Sesamia nonagrioides* (Kader Neino & Bruno Le Rü) en Camargue, ainsi que des cartes environnementales (SIG).

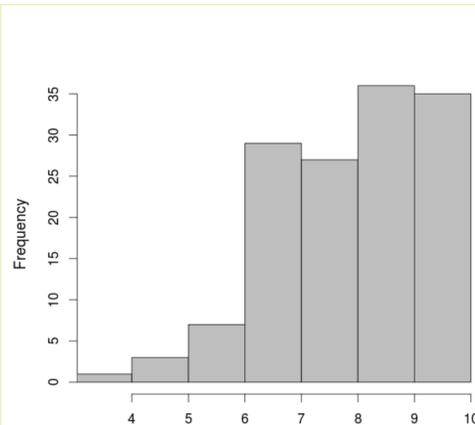
Le modèle génératif permet de simuler en masse des pseudo-données en faisant varier les paramètres d'entrée du modèle génératif (exploration).

On peut intégrer des connaissances a priori, en insistant sur l'exploration de certaines zones de l'espace des paramètres (définition de fonctions de priors).

Les valeurs de paramètres des simulations conduisant à des données "proches" du réel sont donc les plus probables. On obtient des *posterior* sur les paramètres, qui représentent notre degré de croyance sur les valeurs.



Test de fiabilité du modèle : validation croisée du modèle pour le deuxième paramètre β de la loi de dispersion d'un modèle "fat-tail". Chaque point représente une simulation pour laquelle on ré-estime ses paramètres déjà connus. L'ajustement du nuage de points à la droite $y=x$ résume la fiabilité.



Histogramme du *posterior* pour le même paramètre β , obtenu en comparant les données génétiques réelles aux données génétiques simulées. On a ici peu de confiance dans des valeurs de paramètres β inférieures à 6.

MAIS PAS SEULEMENT !

- Bibliothèque R en cours de développement.
- Accès à de nombreux modèles
- Relâcher les hypothèses d'environnement constant
- Utiliser la théorie des graphes pour prédire directement la différenciation génétique
- ...