

**INRAE**

➤ **Modéliser la distribution des espèces**

« Des modèles simples, des modèles complexes, des modèles utiles »

**Frédéric Archaux**

19 octobre 2020 – 10h15-12h15

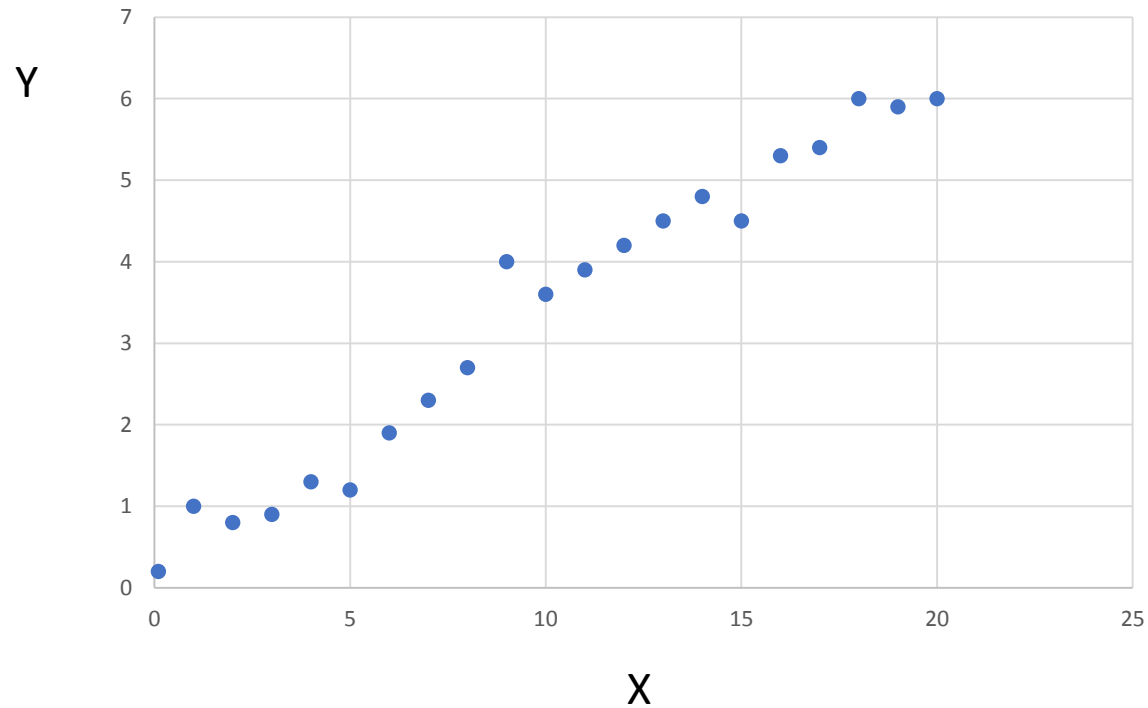
## ➤ Qu'est-ce qu'un modèle?

- Un modèle est une « **représentation humaine** » de ce que pourrait être le Réel
- **Un modèle est toujours faux mais il peut être utile** pour résoudre/faire avancer des problèmes appliqués ou théoriques
- Un modèle est constitué **d'hypothèses de travail**
- (Souvent) le modèle est une **formulation** mathématique approximative du mécanisme qui a généré des observations
- **Modèle de biodiversité = représentation de comment l'environnement façonne la distribution et la dynamique de la biodiversité**



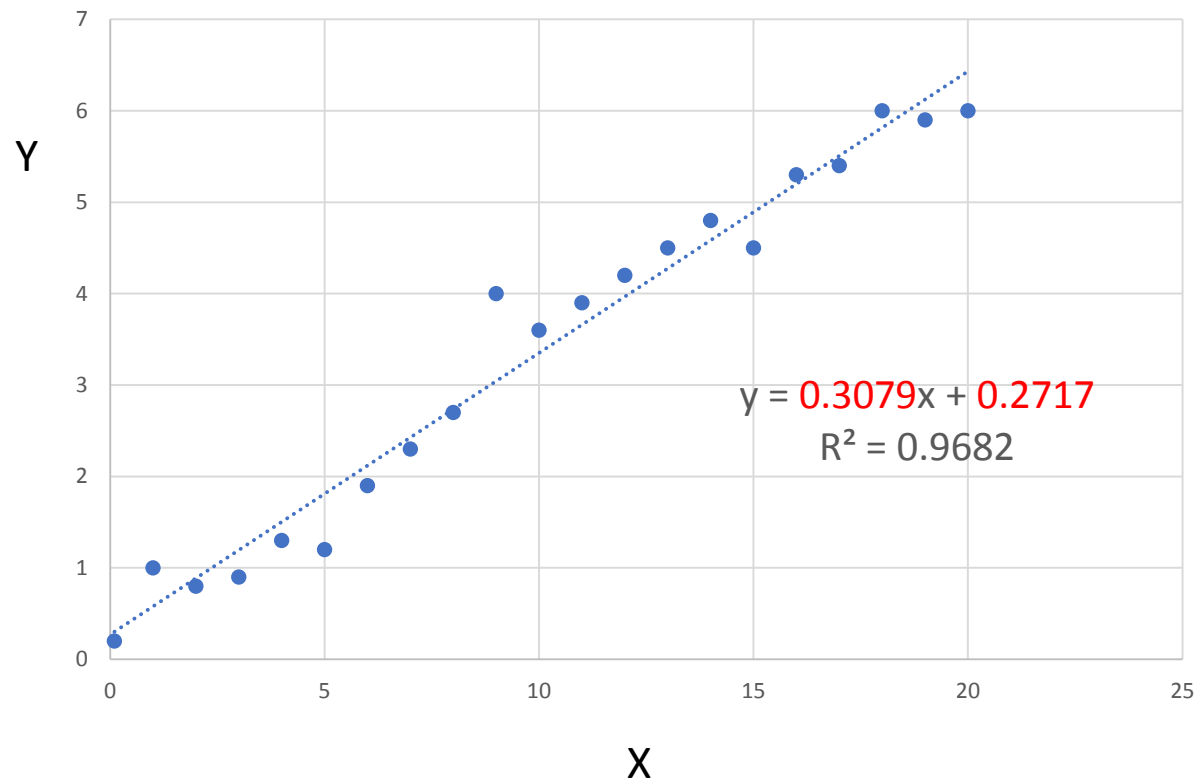
## ➤ Qu'est-ce qu'un modèle?

**Hypothèse** : la variable Y **croît de manière linéaire** avec la variable X



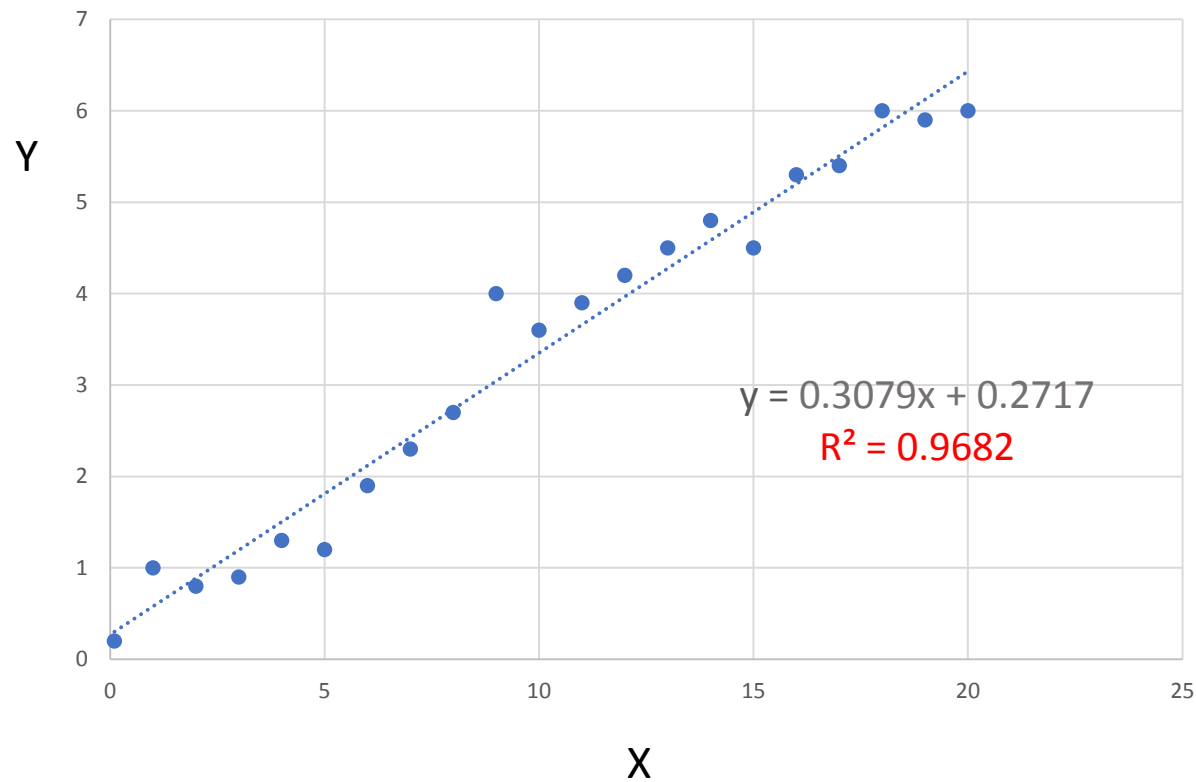
# ➤ Qu'est-ce qu'un modèle?

Estimation des **paramètres** du modèle (calibration)



# ➤ Qu'est-ce qu'un modèle?

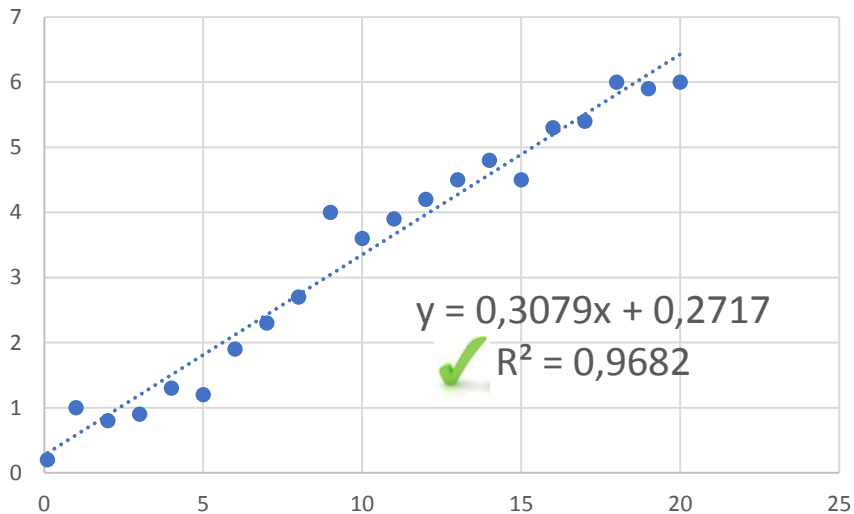
## Evaluation du modèle



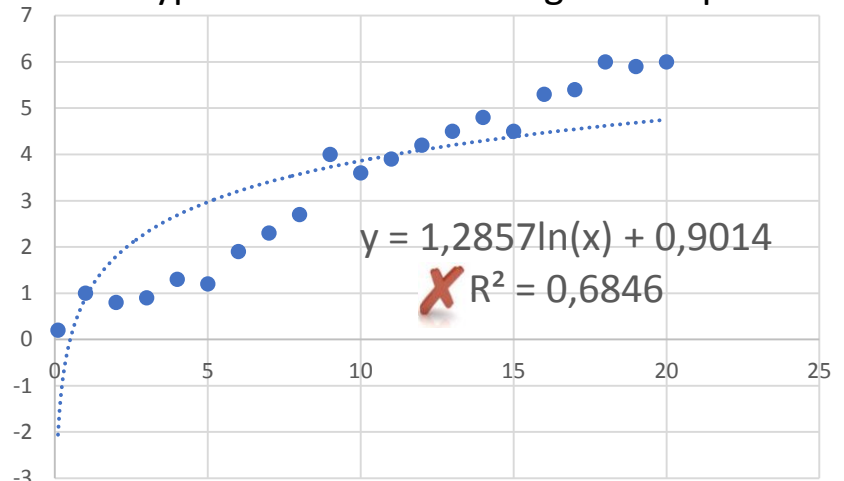
# ➤ Qu'est-ce qu'un modèle?

Comparaison de modèles  
(=test d'hypothèses alternatives)

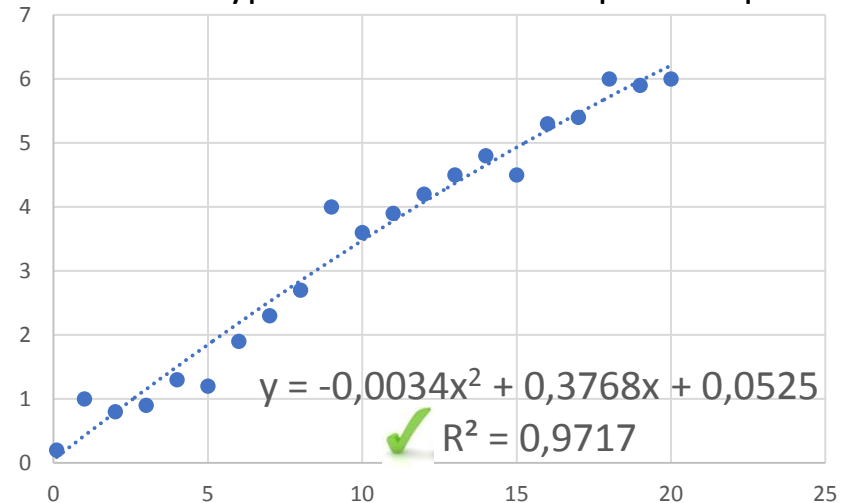
Hypothèse A: relation linéaire



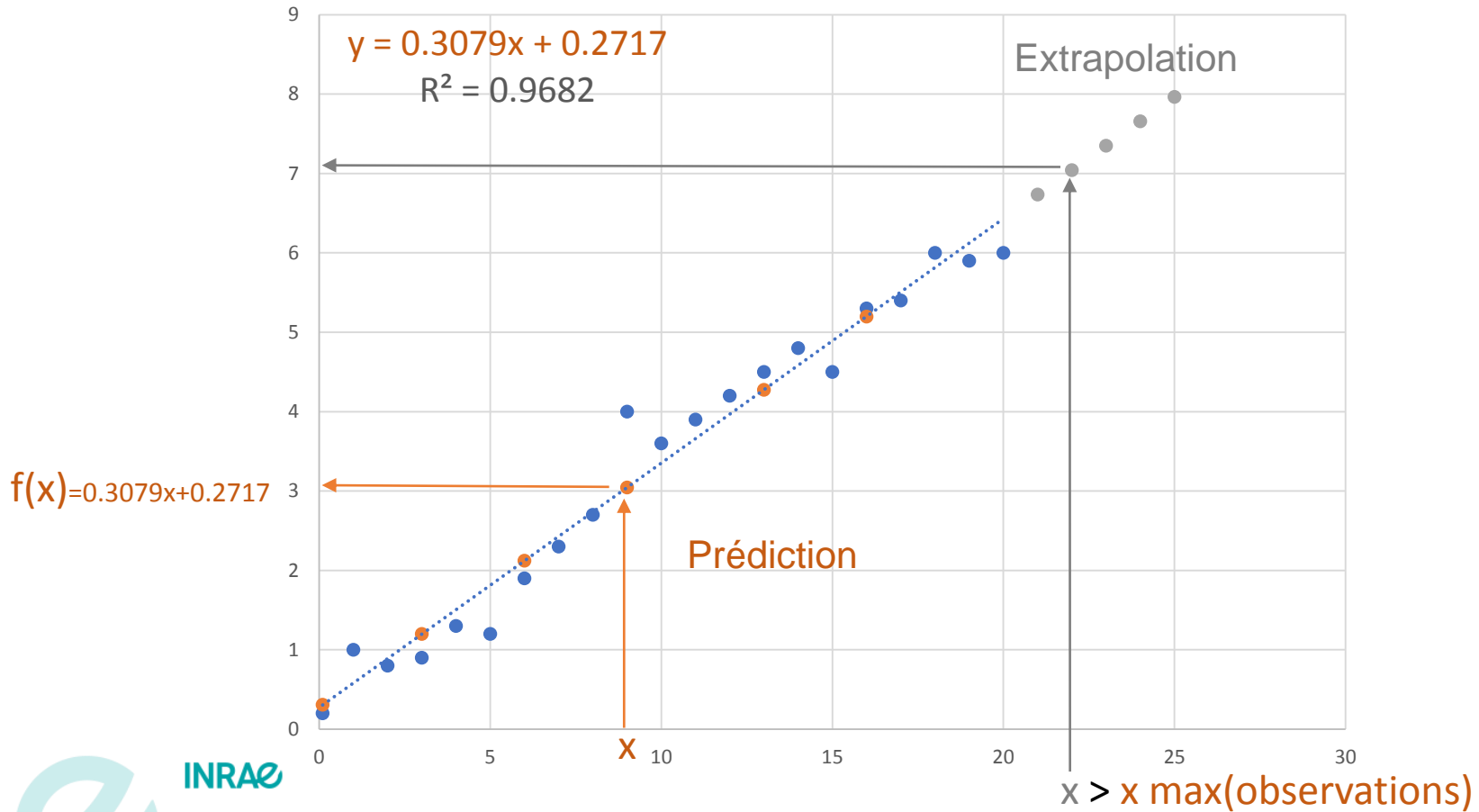
Hypothèse B: relation logarithmique



Hypothèse C: relation quadratique



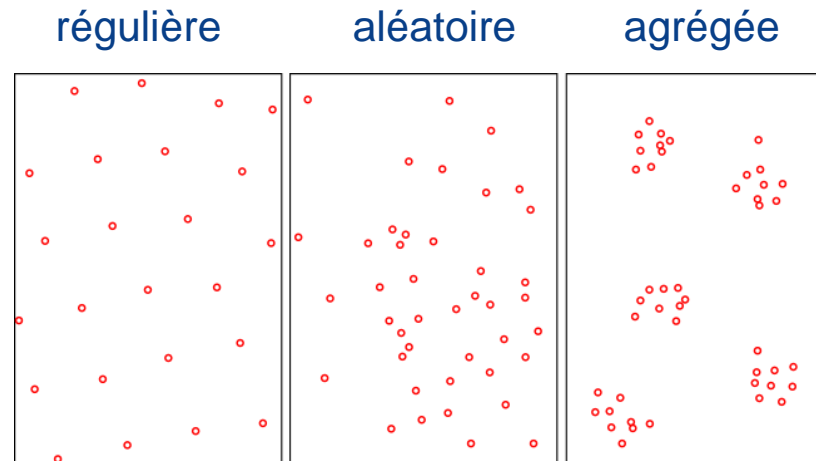
# Prédiction/extrapolation



## ➤ Qu'est-ce que la distribution d'une espèce?

La **distribution spatiale**, en écologie, désigne l'arrangement **spatial** des organismes vivant dans leur milieu naturel. Elle peut aussi se définir comme la fluctuation **spatiale** de l'abondance des organismes dans leur aire de répartition.

Trois archétypes de dispersion sont généralement reconnues :

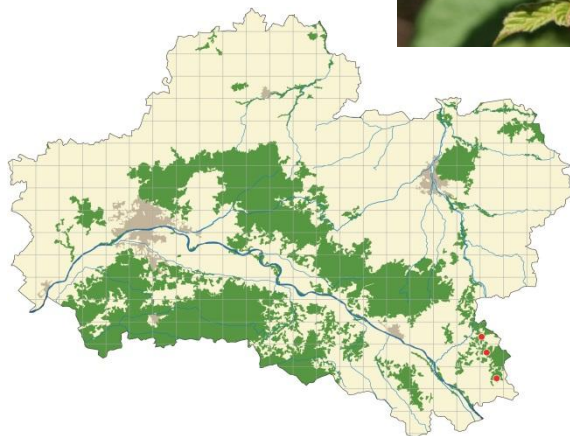




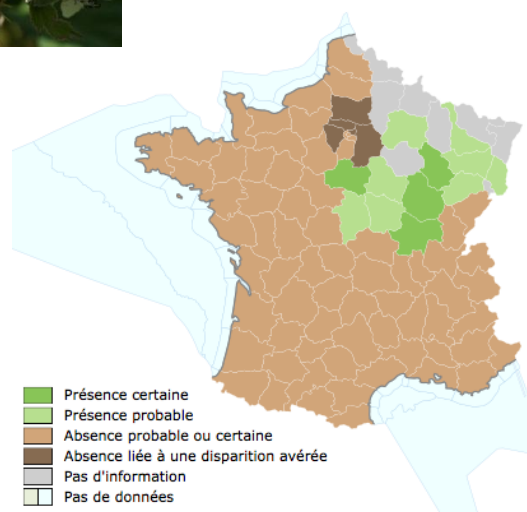
# ➤ Pourquoi s'intéresser à la distribution des espèces?

- Evaluer le statut de conservation (listes rouges, PNA), d'invasion

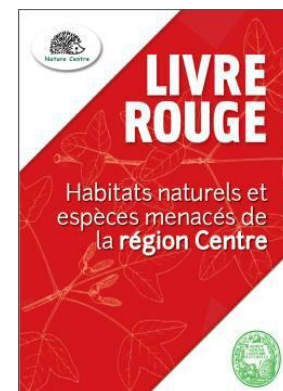
Damier du frêne  
(*Euphydryas maturna*)



Liste rouge régionale: CR - critique



Liste rouge nationale : EN - en danger

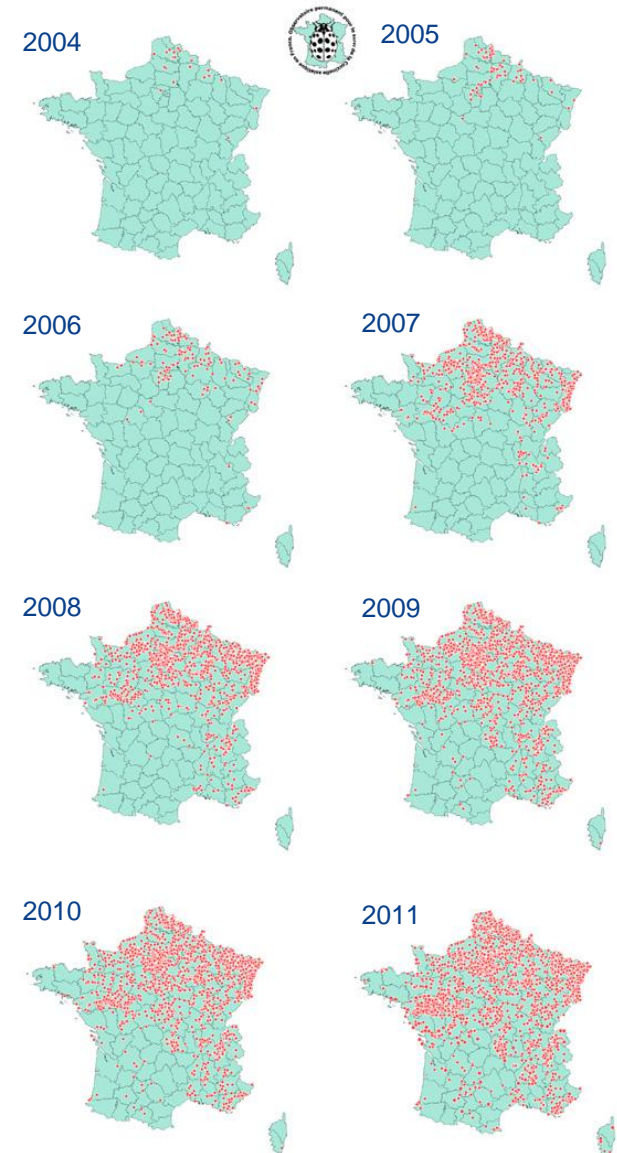


# ➤ Pourquoi s'intéresser à la distribution des espèces?

- Suivre la dynamique de la biodiversité dans le temps et l'espace



[http://vinc.ternois.pagesperso-orange.fr/cote\\_nature/Harmonia\\_axyridis/index.htm](http://vinc.ternois.pagesperso-orange.fr/cote_nature/Harmonia_axyridis/index.htm)



**INRAE**

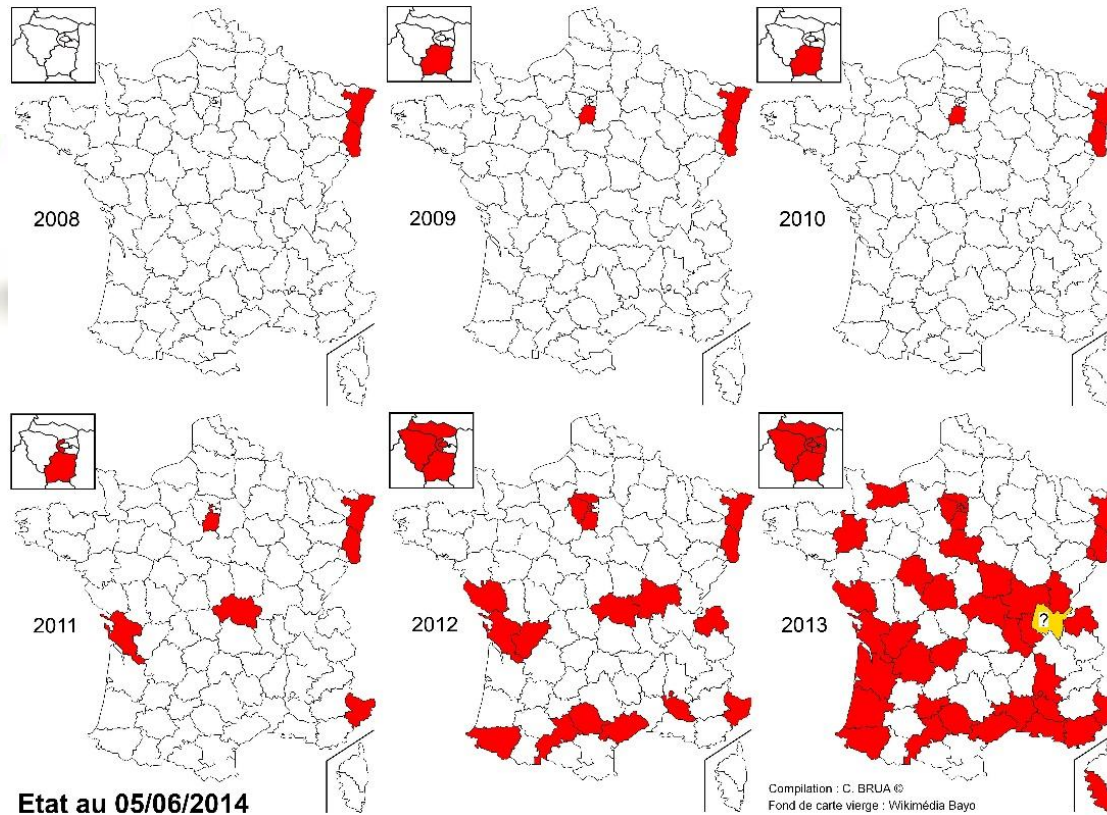
Master 1 BICG - FMB

Frédéric Archaux

# ➤ Pourquoi s'intéresser à la distribution des espèces?

- Adapter des mesures de conservation/d'éradication

Frelon asiatique



# ➤ Modéliser la distribution de la biodiversité?

## Important : attentions aux termes

- Modèle de distribution des espèces (Species Distribution Models)
- Modèle de la niche (écologique/environnementale/bioclimatique) des espèces
- Modèle de l'habitat des espèces
- ...

## On ne modélise pas :

- la distribution des espèces
- la niche écologique des espèces

## On modélise:

- La **probabilité** de présence ou l'abondance **potentielle** de l'espèce en un point de l'espace, estimée d'après les conditions environnementales

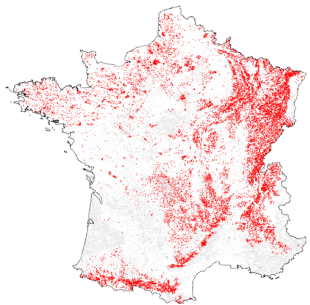




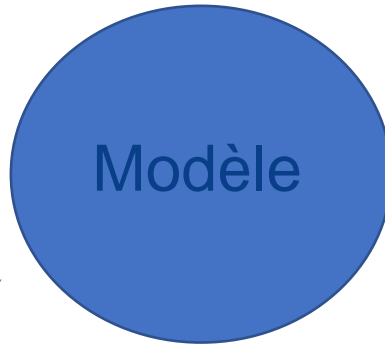
# ➤ Pourquoi modéliser la distribution des espèces?

Observations ponctuelles (Y)

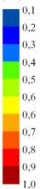
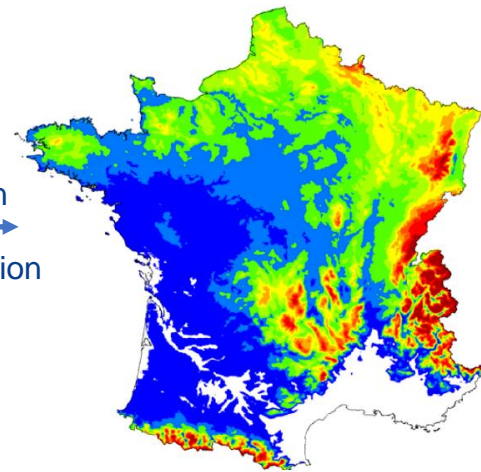
Probabilité de présence/abondance potentielle



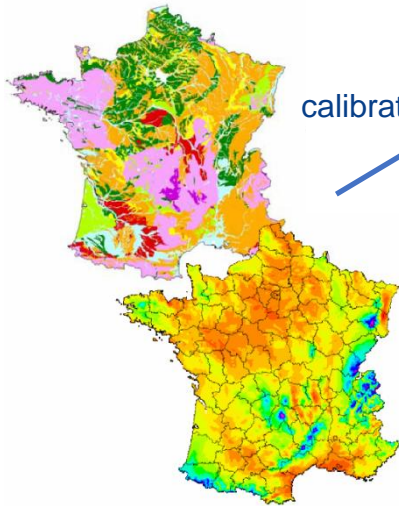
évaluation  
calibration



prédiction  
extrapolation



Informations environnementales (X)



calibration



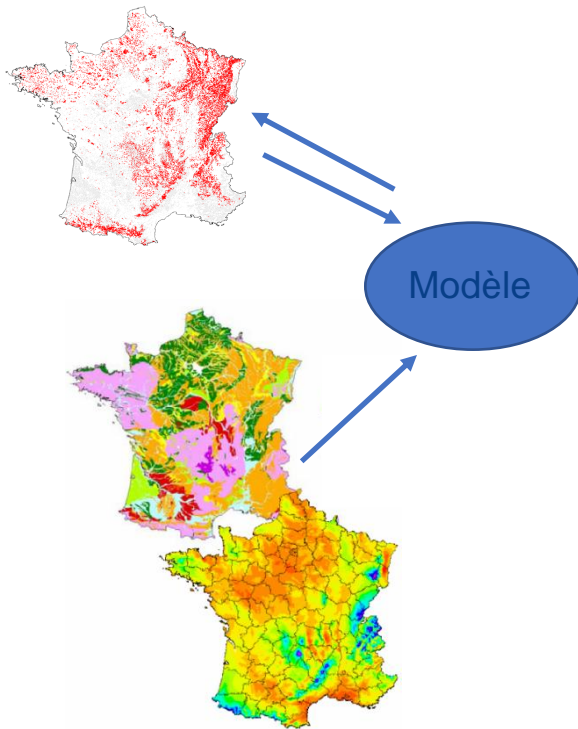
INRAE

Master 1 BICG - FMB

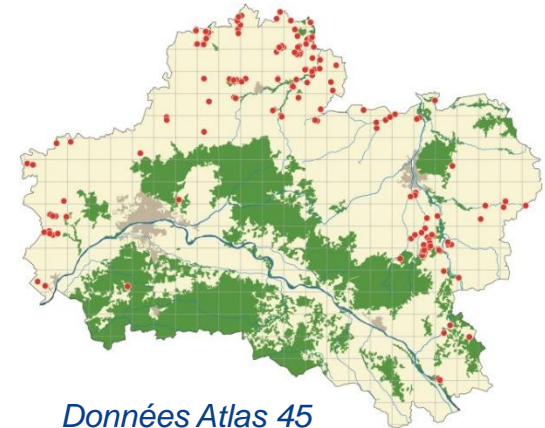
Frédéric Archaux

# ➤ Pourquoi modéliser la distribution des espèces?

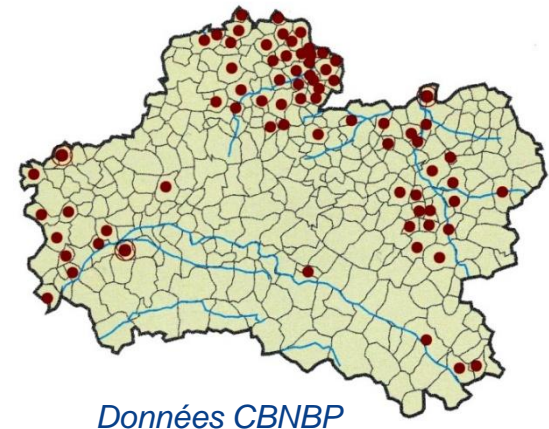
- Comprendre les facteurs déterminants pour la persistance de l'espèce



Azuré bleu-céleste  
(*Lysandra bellargus*)

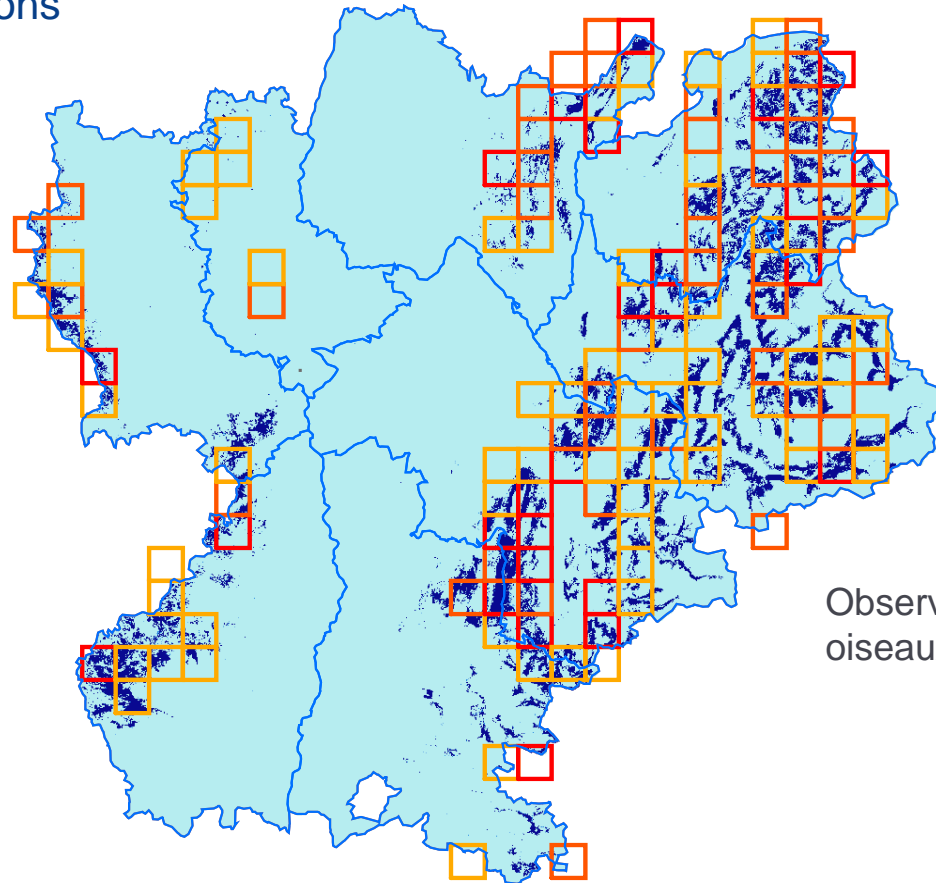
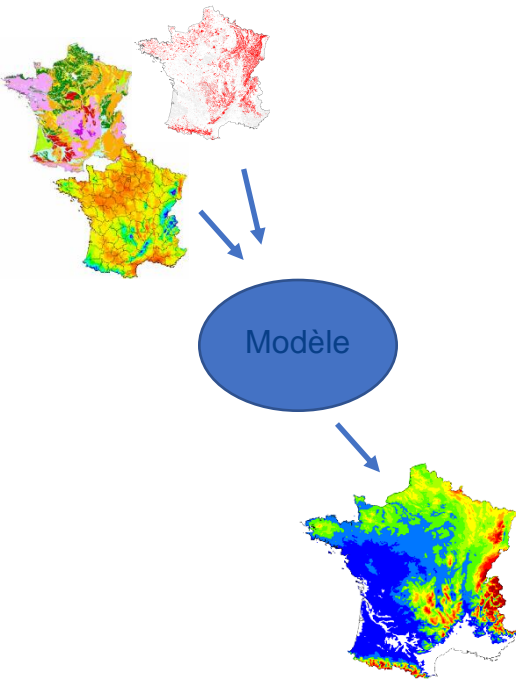


Hippocrépide à touppet  
(*Hippocrepis comosa*)



# ➤ Pourquoi modéliser la distribution des espèces?

- Orienter des prospections



Nyctale de Tengmalm



Photo : J. Séon ©

Observations (atlas oiseaux nicheurs)

Nidification certaine

Nidification probable

Nidification possible

Modélisation (ONF-LPO)

● forte probabilité de présence



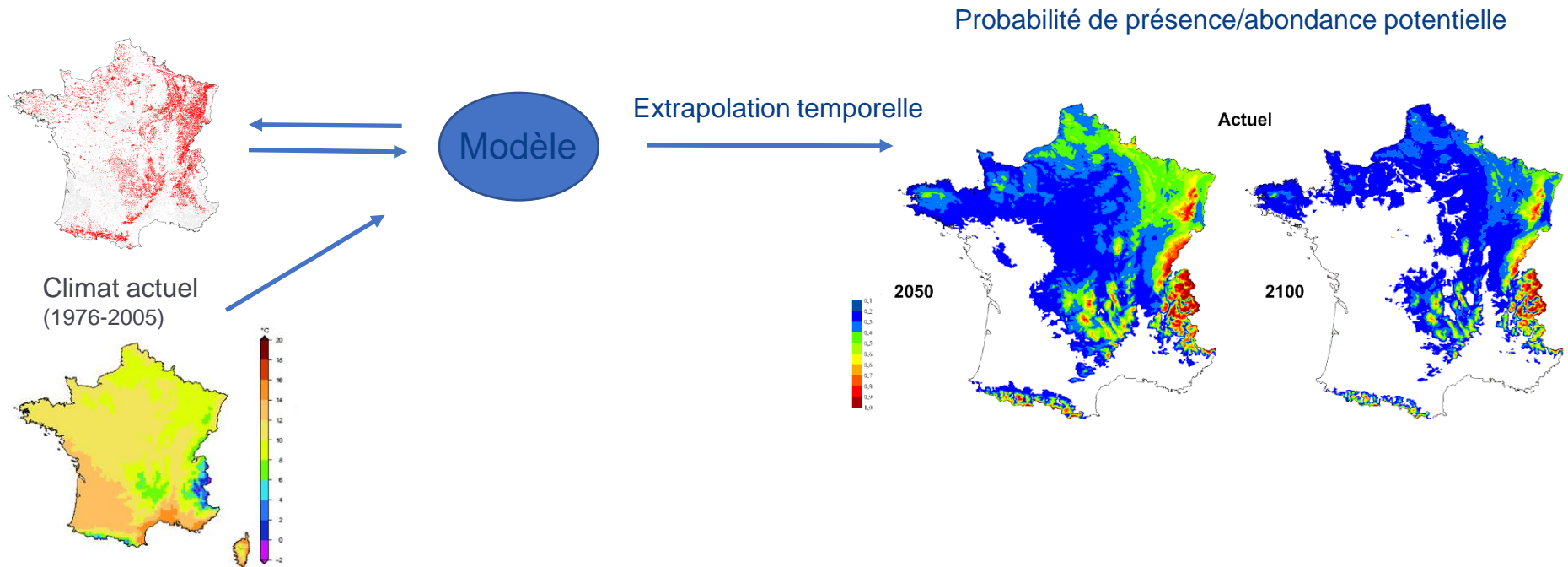
INRAE

Master 1 BICG - FMB

Frédéric Archaux

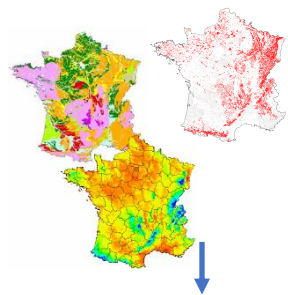
# ➤ Pourquoi modéliser la distribution des espèces?

- Anticiper les changements de distribution

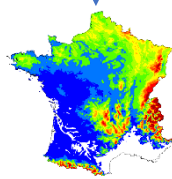




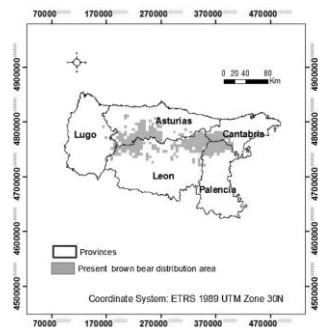
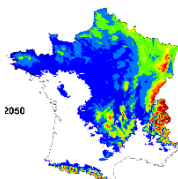
# ➤ Pourquoi modéliser la distribution des espèces?



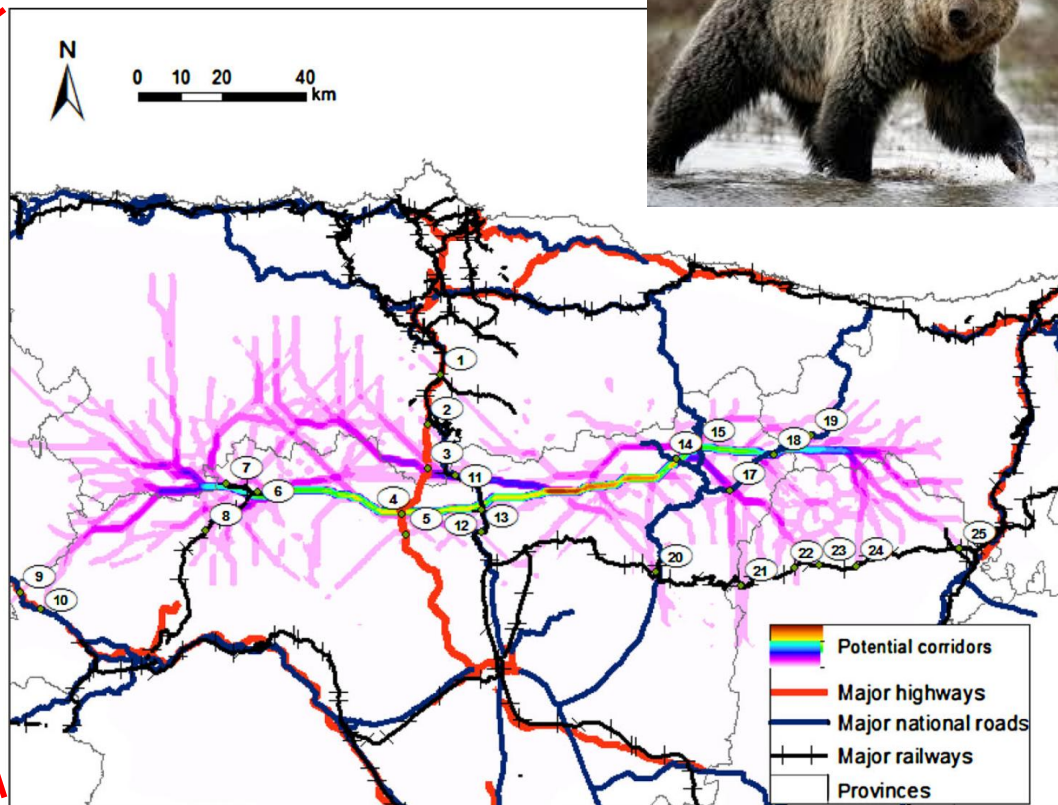
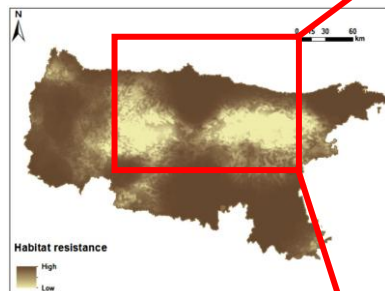
Modèle « d'habitat »



Modèle « de dispersion »



- Localiser les habitats à protéger en priorité – les ouvrages pour permettre le passage de la faune



# > Comment modéliser la distribution?

Pléthore d'approches !

Modèle corrélatif

Modèle mécaniste

Modèle métapopulationnel

Automate cellulaire

Modèle individu-centré

Simple



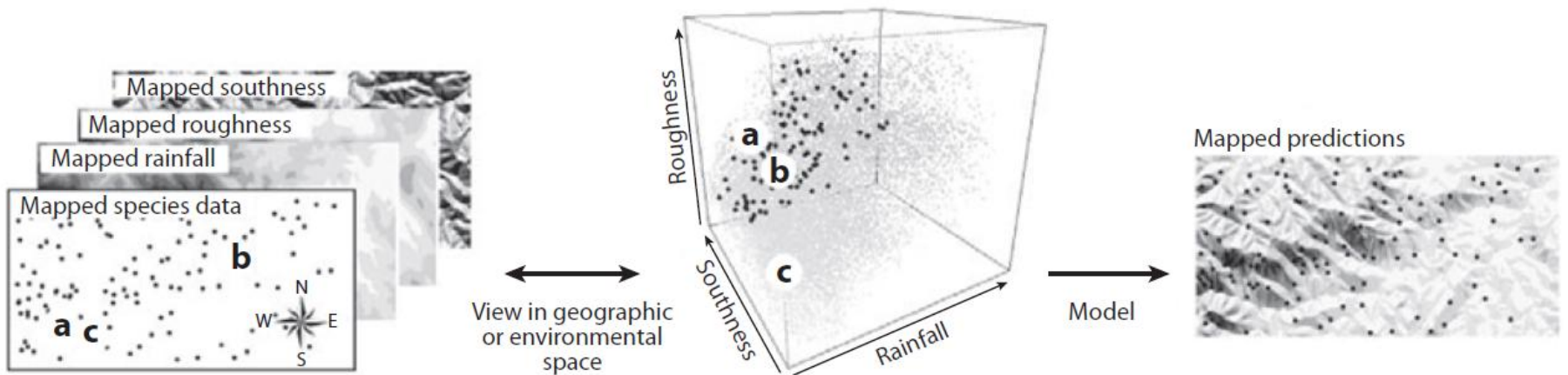
Complexe

Grande surface



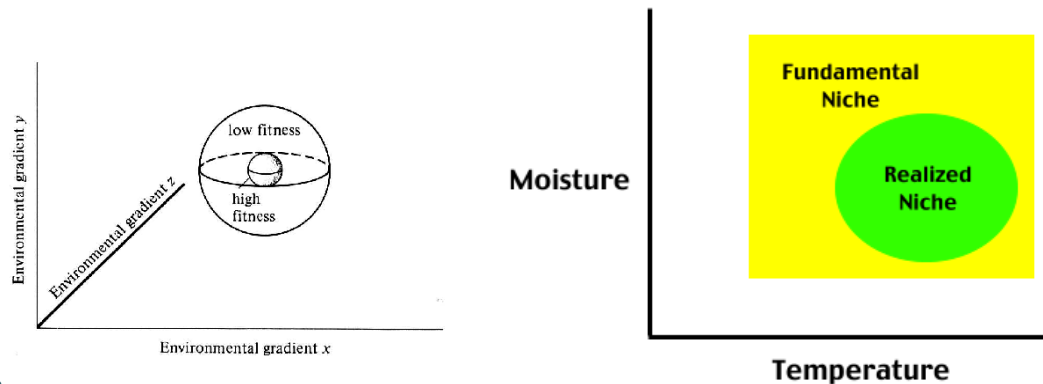
Petite surface

# ➤ Principe de l'approche corrélative



Probabilité de présence  
Abondance prédite

## Niche d'Hutchinson

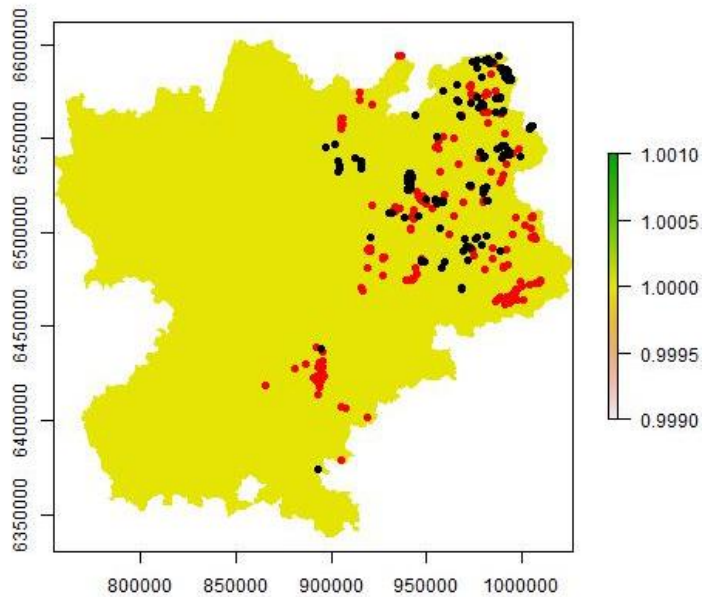
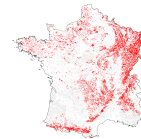


INRAE

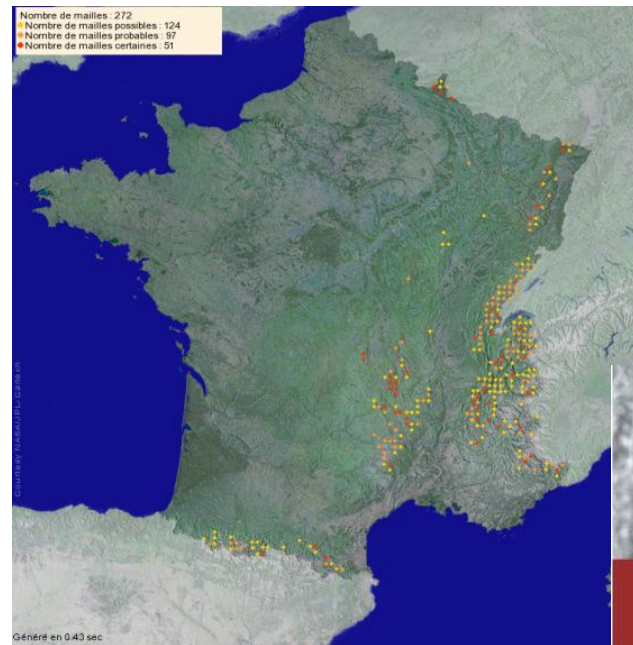
Master 1 BICG - FMB  
Frédéric Archaux

# ➤ Données en entrée : les observations

- Récolter des « observations » pertinentes



- présence avérée
- présence recherchée mais non avérée



Nyctale de Tengmalm  
(*Aegolius funereus*)

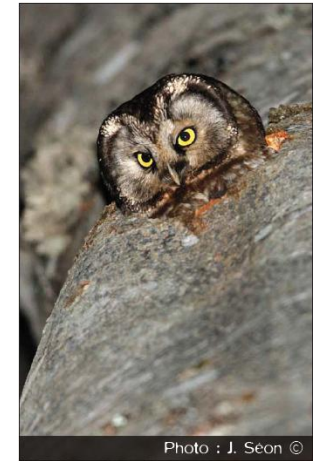


Photo : J. Séon ©



Bilger et al, données ONF-LPO



INRAE

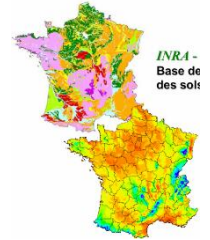
Master 1 BICG - FMB

Frédéric Archaux

# ➤ Données d'entrée: environnement

## - Récolter/calculer des données/variables environnementales pertinentes

- altitude (mnt)
- la topographie (dérivée du mnt, selon 4 classes : 1-vallée, 2-plaine, 3-versant, 4-crête)
- distance à la plus proche zone urbaine
- distance à la plus proche prairie
- distance à la plus proche route
- le type d'occupation du sol/type de peuplements forestiers



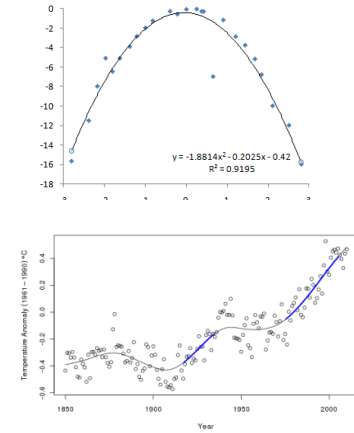
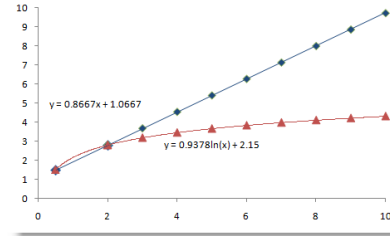
- 1) Cours et voies d'eau, Plans d'eau, Marais intérieurs, Terres arables hors périmètres d'irrigation, Vignobles, Vergers et petits fruits, Oliveraies, Systèmes cultureux et parcellaires complexes
- 2) Lande et broussailles, Forêt et végétation arbustive en mutation, Formation herbacée, Végétation sclérophylle, Forêt fermée ou ouverte sans couvert arboré
- 3) Forêt mélangée, Forêt fermée à mélange de feuillus prépondérants et conifères
- 4) Forêt fermée de sapin ou épicéa
- 5) Forêt fermée de conifères purs, de mélange de conifères ou mélange conifères prépondérants et feuillus
- 6) Forêt fermée de feuillus purs ou de mélange de feuillus
- 7) Prairies, pelouses, pâturages naturels et végétation clairsemée
- 8) Forêt ouverte de feuillus, conifères ou mélange feuillus conifères



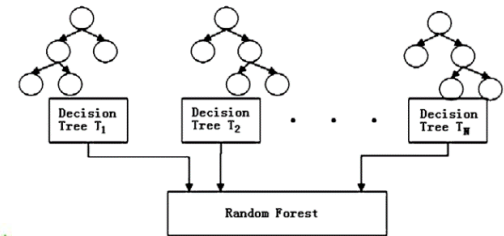
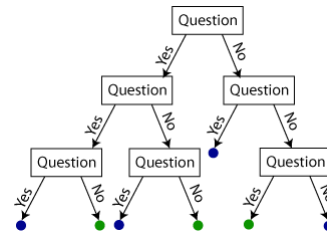


# ➤ Choix du modèle statistique

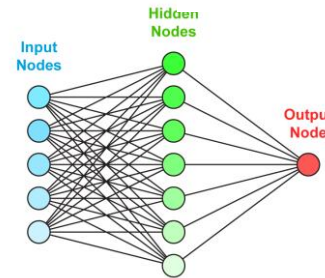
- Régressions linéaires/non linéaires



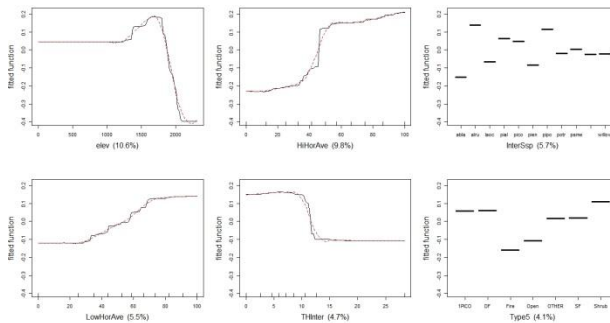
- Arbres de décisions



- Réseaux de neurones



- MaxEnt



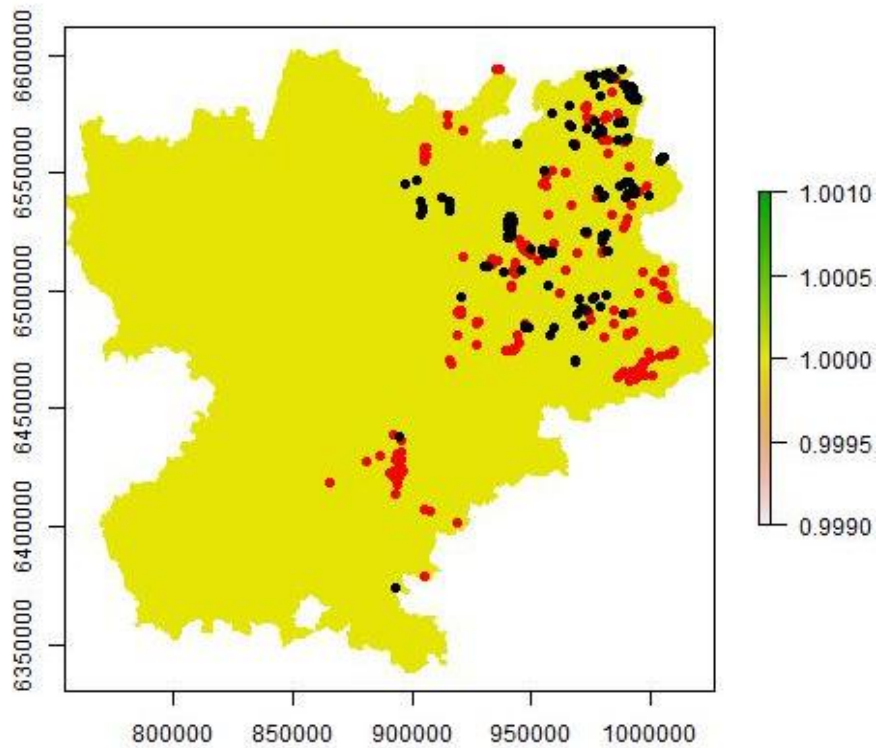
INRAE

Master 1 BICG - FMB

Frédéric Archaux

# ➤ Choix de la métrique analysée (1)

- Analyser en abondance, en présence/absence, en présence seulement?

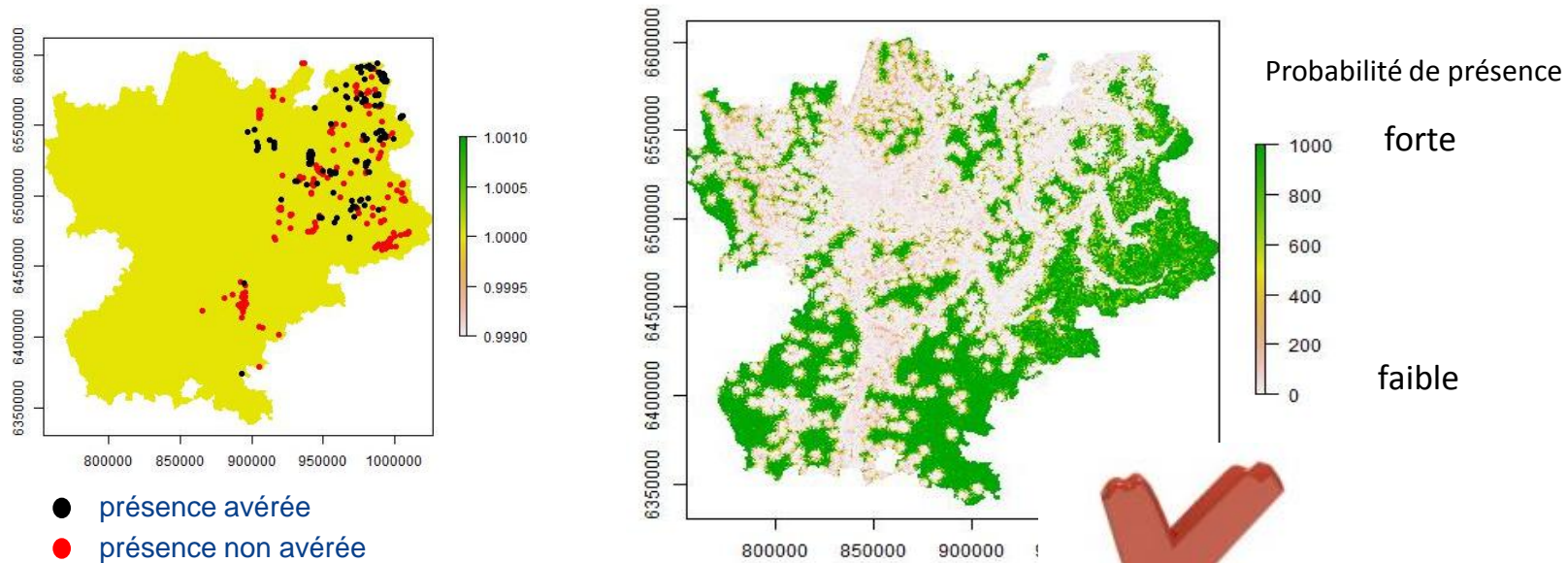


● présence avérée

● présence non avérée

## ➤ Choix de la métrique analysée (2)

- Analyser en abondance, en présence/absence, en présence seulement?



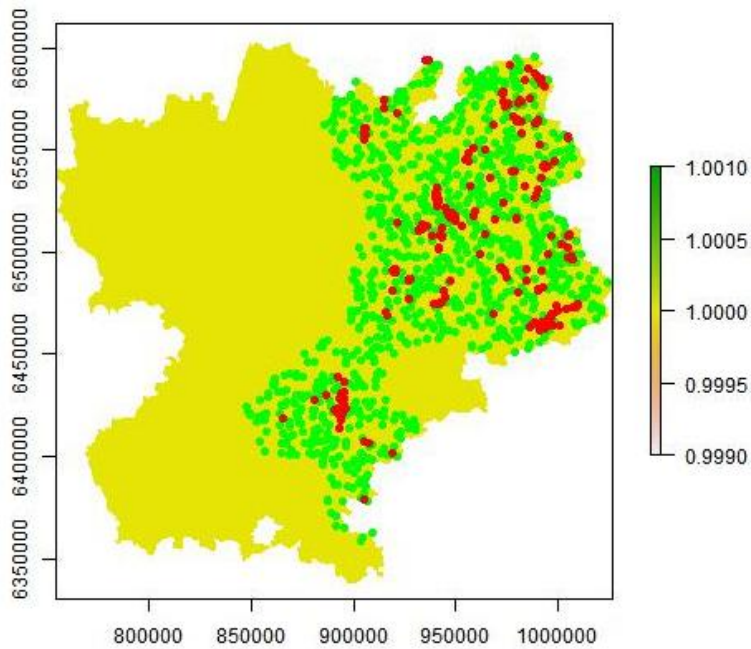
Trop optimiste !



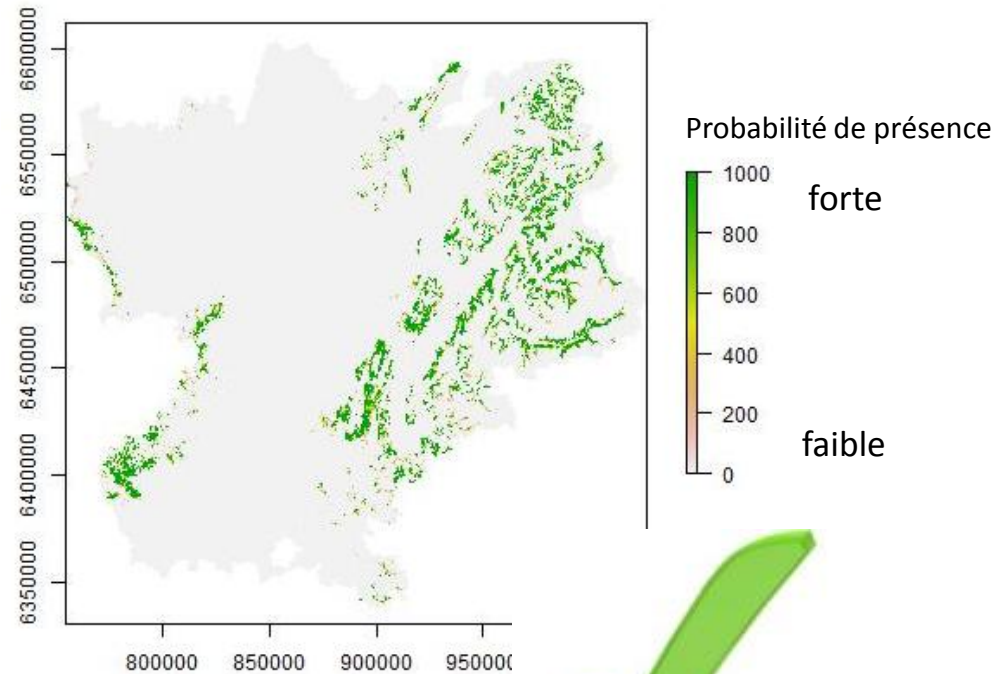


## ➤ Choix de la métrique analysée (3)

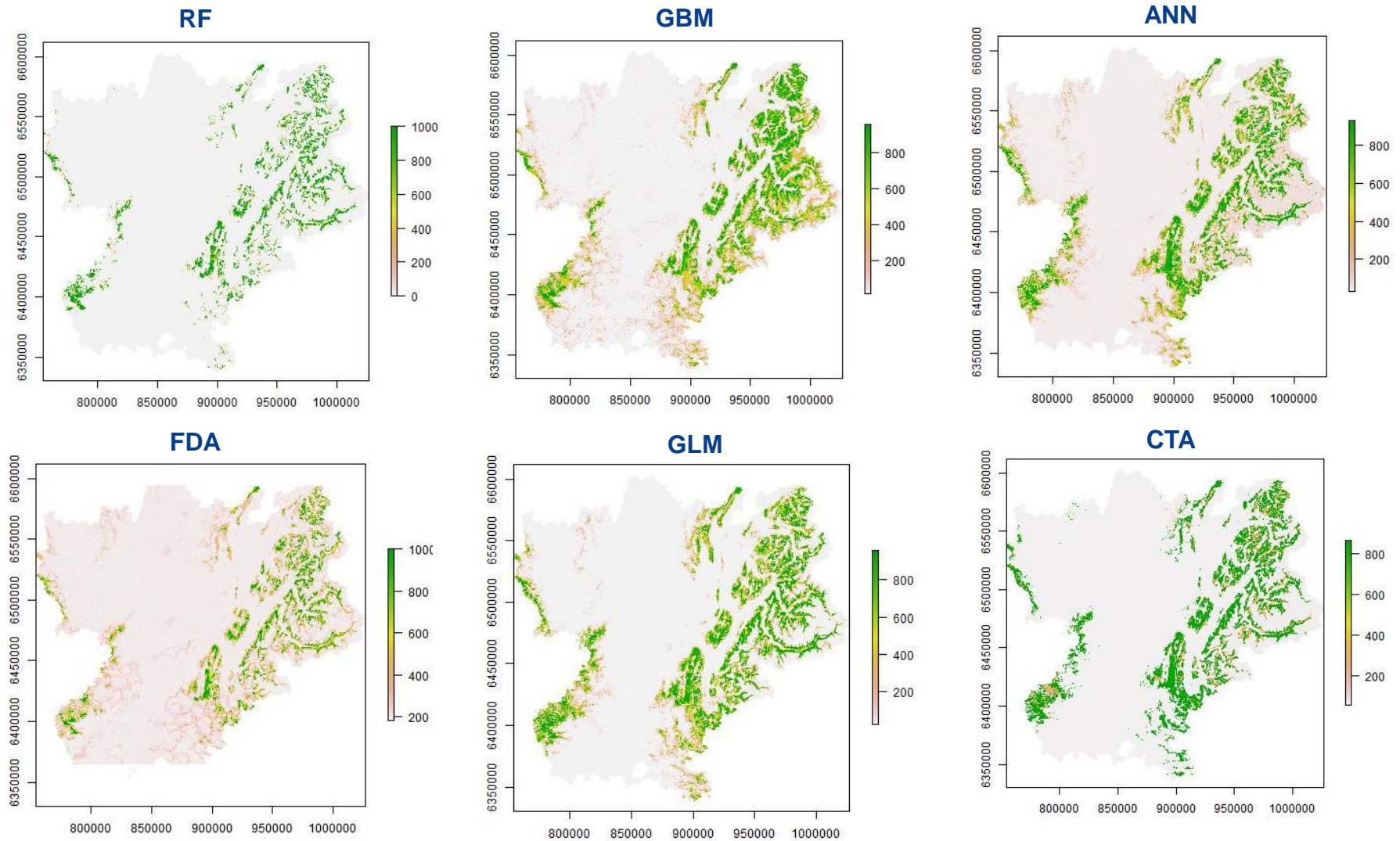
- Analyser en abondance, en présence/absence, en présence seulement?



● présence avérée  
● pseudo-absence



# ➤ Validation du (des) modèle(s) (1)



**INRAE**





Master 1 BICG - FMB  
Frédéric Archaux

Random Forest, Generalized Boosted Model, Artificial Neuronal Network, Flexible Discriminant Analysis, Generalized Linear Models, Classification Tree Analysis

## ➤ Validation du (des) modèle(s) (2)

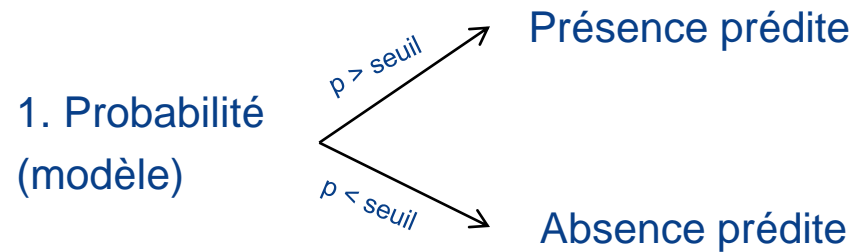
Validation interne au jeu de données :

- Réserver une partie de l'échantillon analysé (10 à 30%)
- Calibrer le modèle sur le reste de l'échantillon (estimation des paramètres)
- Comparer prédictions et observations sur l'échantillon réservé

	Présence observée	Absence	
Présence prédite			Sensibilité=% vraies présences
Absence prédite			Spécificité=% vraies absences

# ➤ Validation du (des) modèle(s) (3)

## Processus itératif



	Présence observée	Absence	
Présence prédite			Sensibilité=% vraies présences
Absence prédite			Spécificité=% vraies absences

2. Déterminer le seuil qui maximise la sensibilité et la spécificité
3. Evaluer la qualité prédictive du modèle pour ce seuil

# ➤ Validation du (des) modèle(s) (4)

## Validation interne

Qualité prédictive du modèle	Valeur des tests statistiques pour l'évaluation de modèles		
	TSS	KAPPA	ROC
Faible	0.2 à 0.4	0.2 à 0.4	0.6 à 0.7
Correcte	0.4 à 0.6	0.4 à 0.6	0.7 à 0.8
Bonne	0.6 à 0.8	0.6 à 0.8	0.8 à 0.9
Excellente	0.8 à 1	0.8 à 1	0.9 à 1



Statistiques alternatives pour juger la qualité prédictive du modèle basées sur la sensibilité et la spécificité

**RF** (kappa=0.979, Roc=1)

**GBM** (kappa=0.704, Roc=0.954)

**FDA** (kappa=0.661, Roc=0.917)

**ANN** (kappa=0.654, Roc=0.917)

**GLM** (kappa=0.62, Roc=0.916)

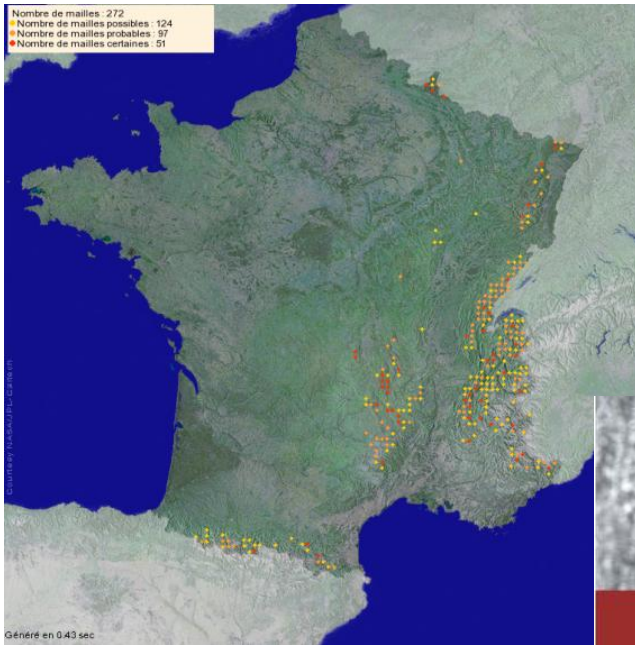
**CTA** (kappa=0.619, Roc=0.888)



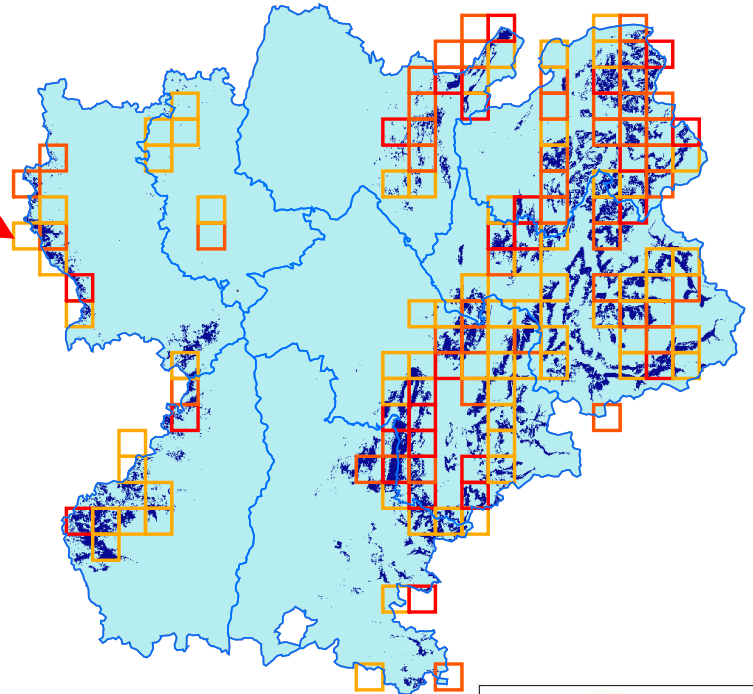


# ➤ Validation du (des) modèle(s) (5)

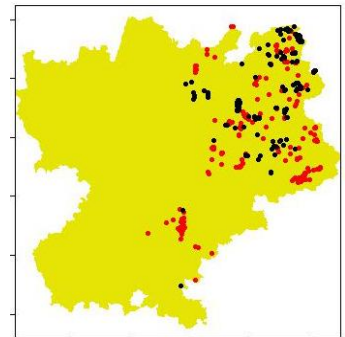
## Validation externe



Données  
indépendantes  
(=atlas des oiseaux  
nicheurs)



- Nidification certaine
- Nidification probable
- Nidification possible



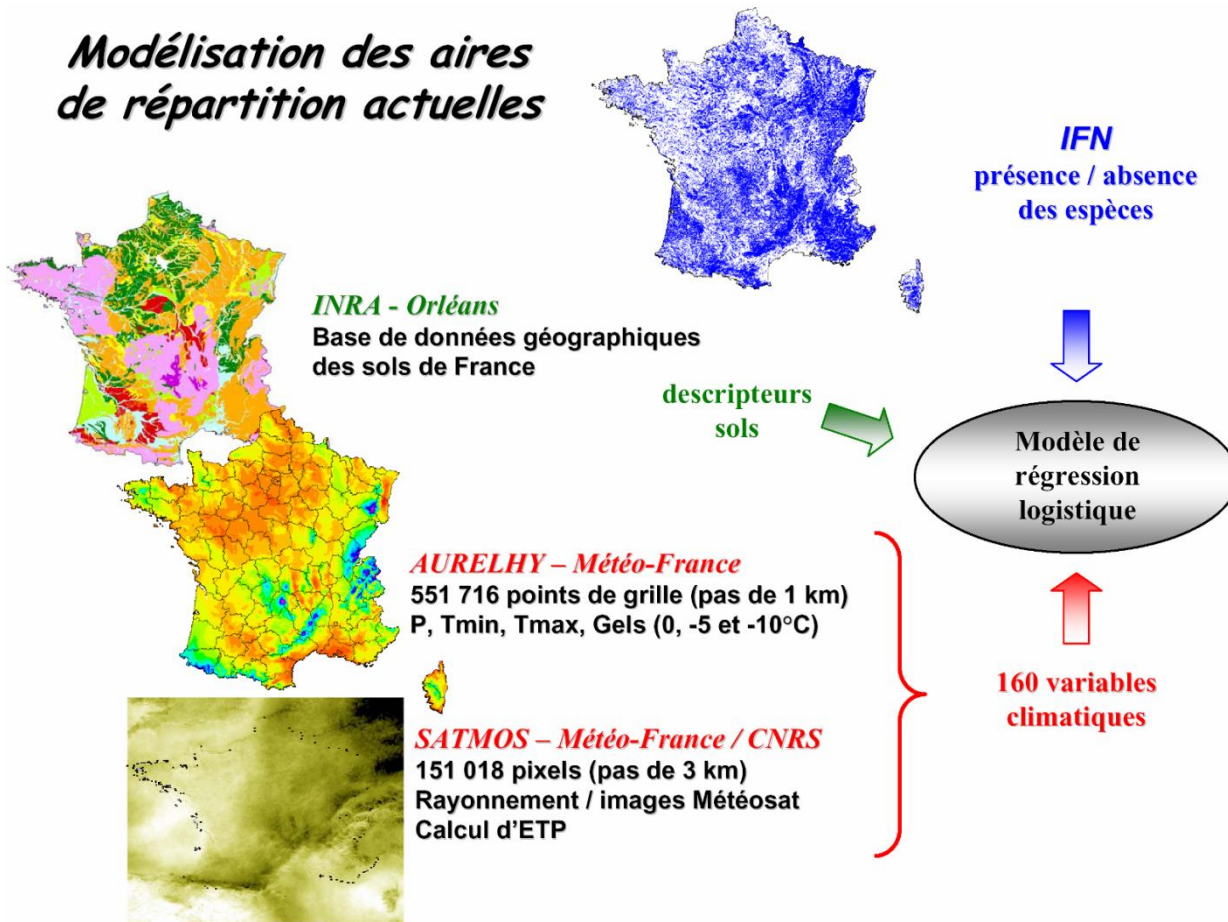
INRAE

Master 1 BICG - FMB

Frédéric Archaux

# ➤ Exemple d'utilisation de modèle corrélatif

*Modélisation des aires de répartition actuelles*



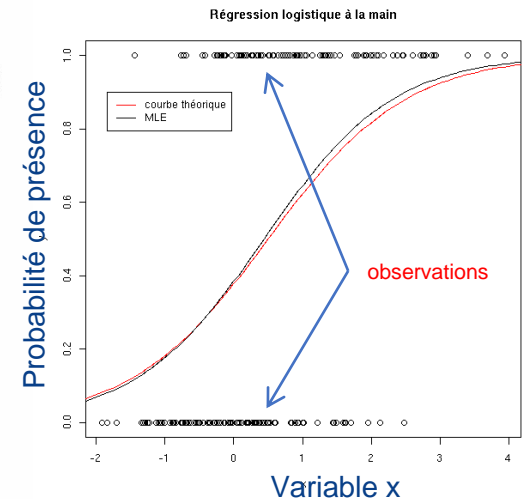
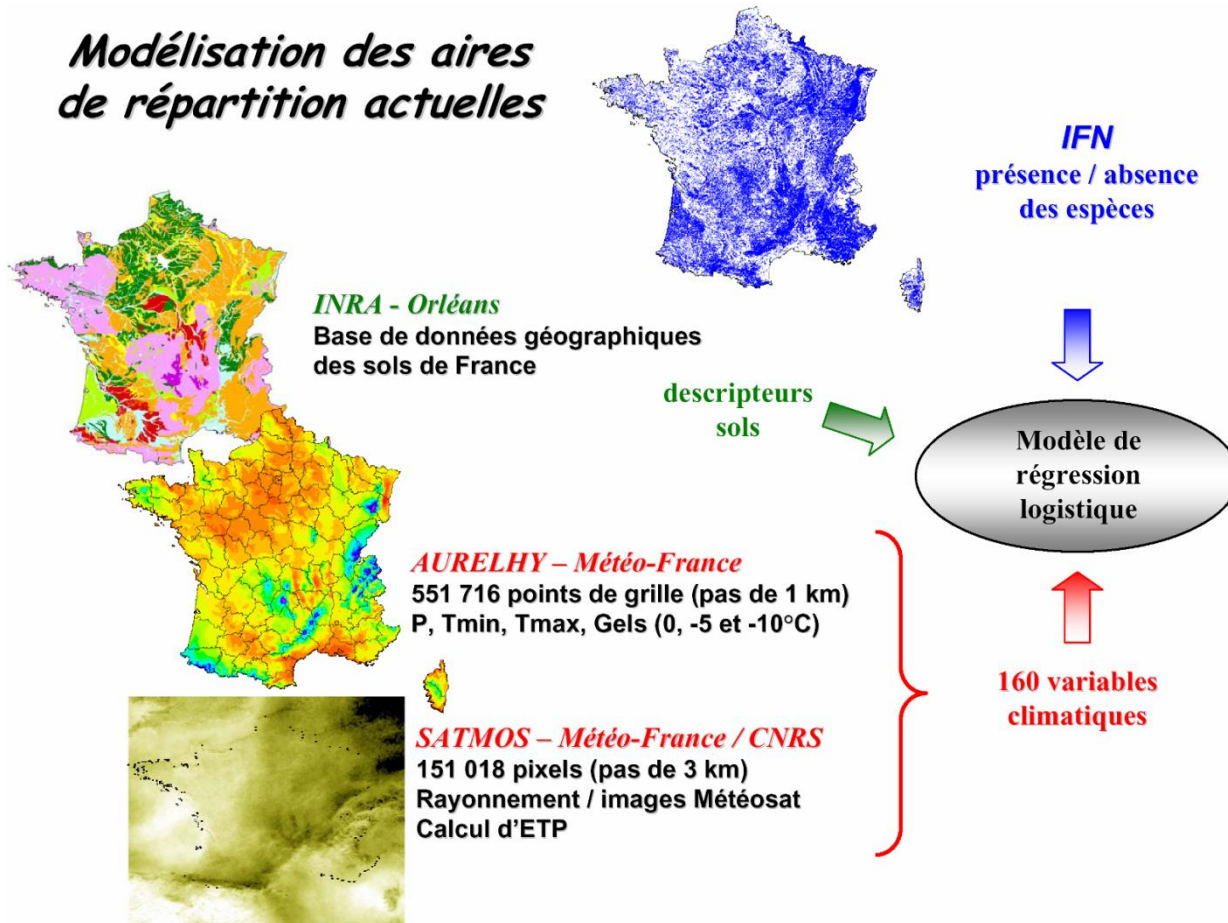
*Dupouey & Chuine*

**INRAE**

Master 1 BICG - FMB  
Frédéric Archaux

# ➤ Exemple d'utilisation de modèle corrélatif

*Modélisation des aires de répartition actuelles*



$$y = \frac{1}{1 + e^{-(\alpha + \beta x)}}$$

Dupouey & Chuine

INRAE

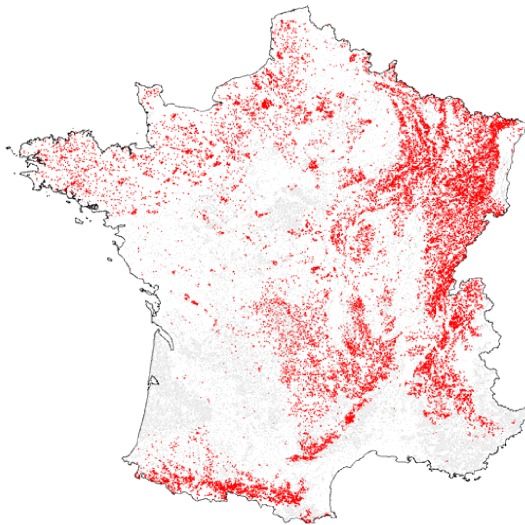
Master 1 BICG - FMB  
Frédéric Archaux



# ➤ Modélisation de la distribution du Hêtre

Hêtre (*Fagus sylvatica*)

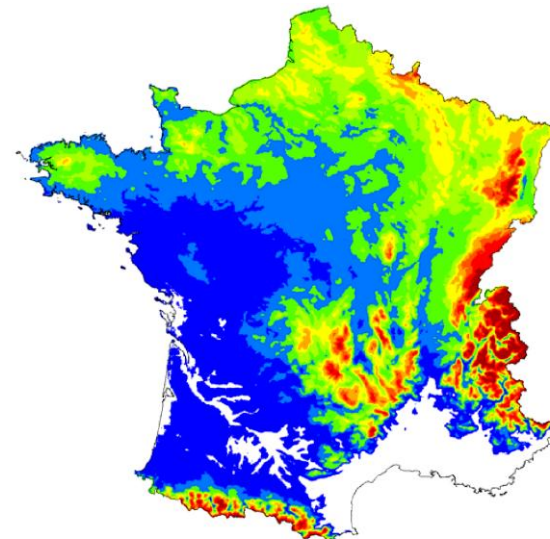
Distribution observée  
(données IFN)



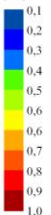
Distribution modélisée

Sensibilité :

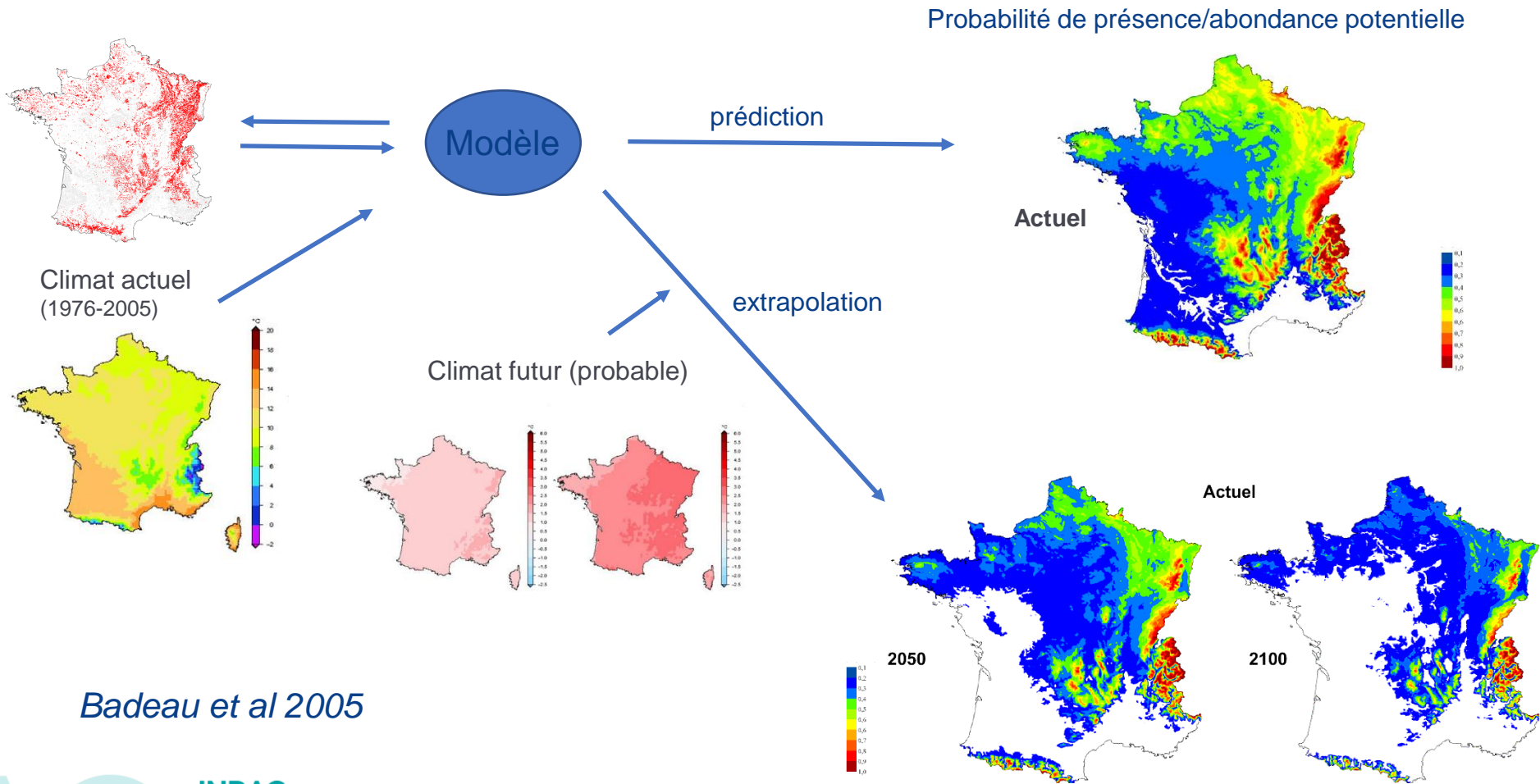
- au déficit hydrique juin-juillet
- aux fortes T°C octobre



Probabilité de présence



# ➤ Modèle de la "distribution future" du Hêtre



*Badeau et al 2005*

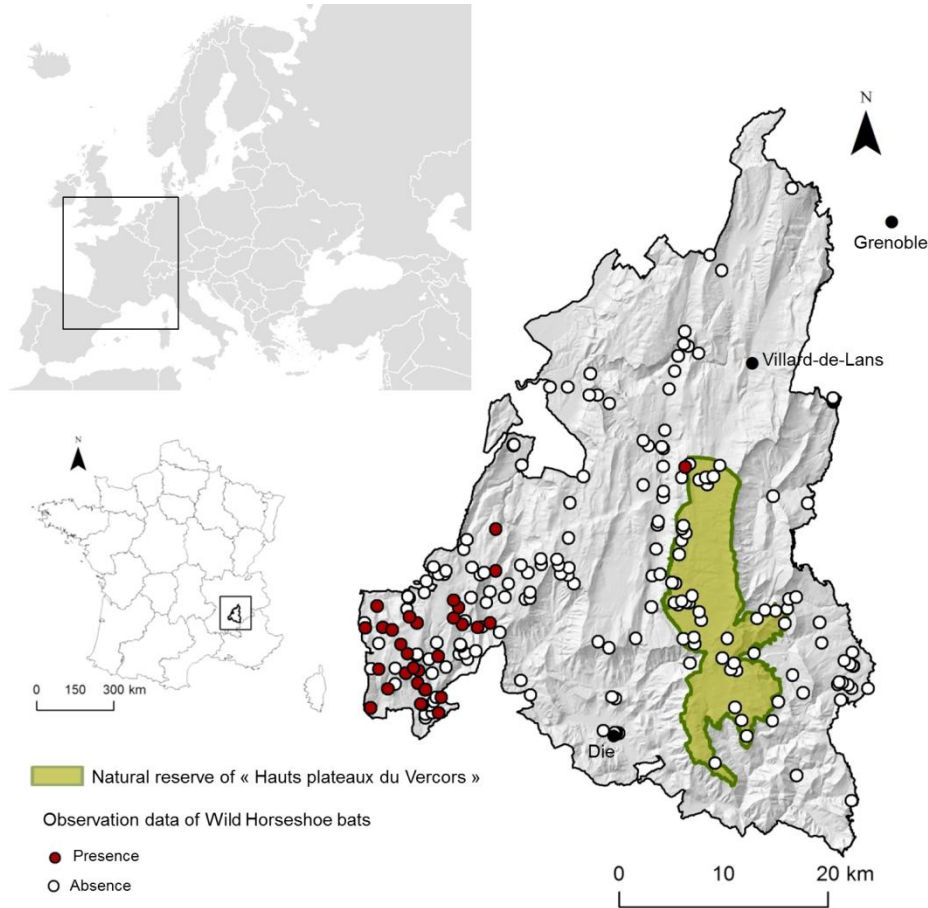


**INRAE**

Master 1 BICG - FMB

Frédéric Archaux

# ➤ Modèle corrélatif intégrant la connectivité paysagère

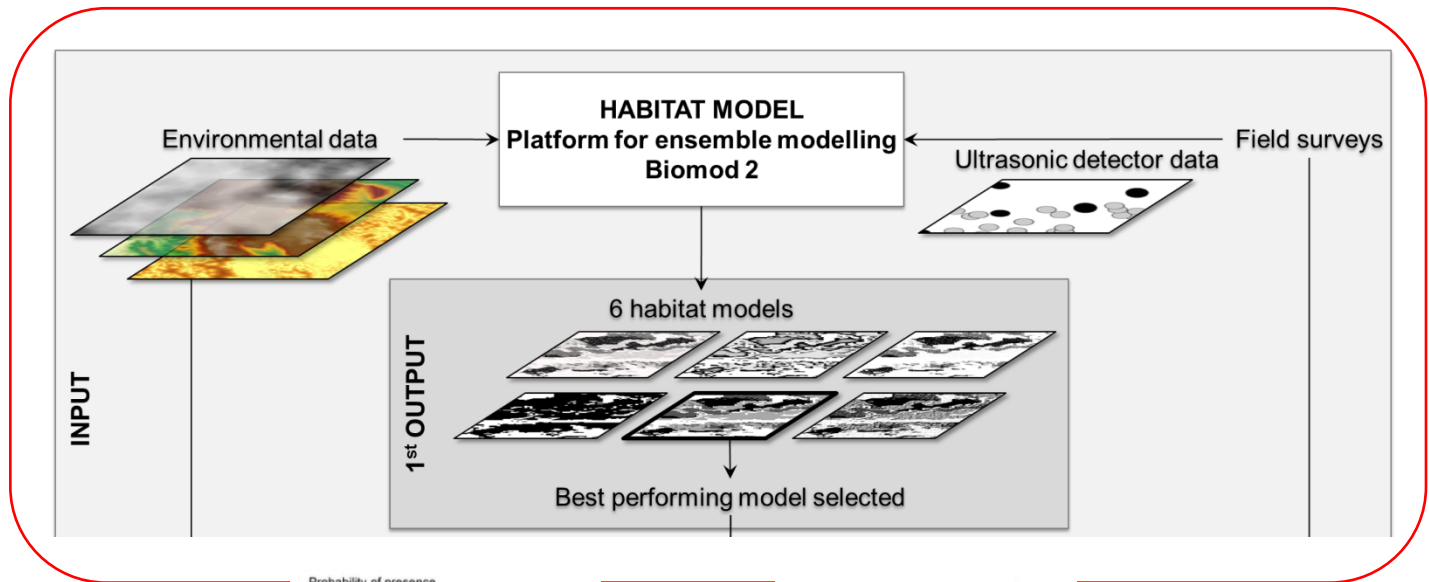
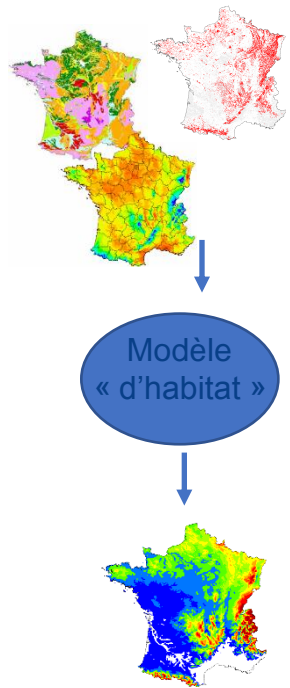


## Rhinolophes

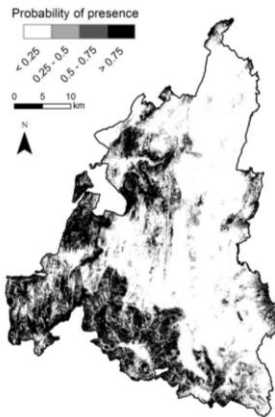


# ➤ Modèle corrélatif intégrant la connectivité paysagère

Première étape : modèle d'habitat



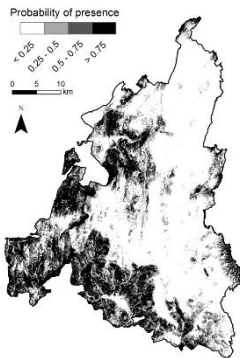
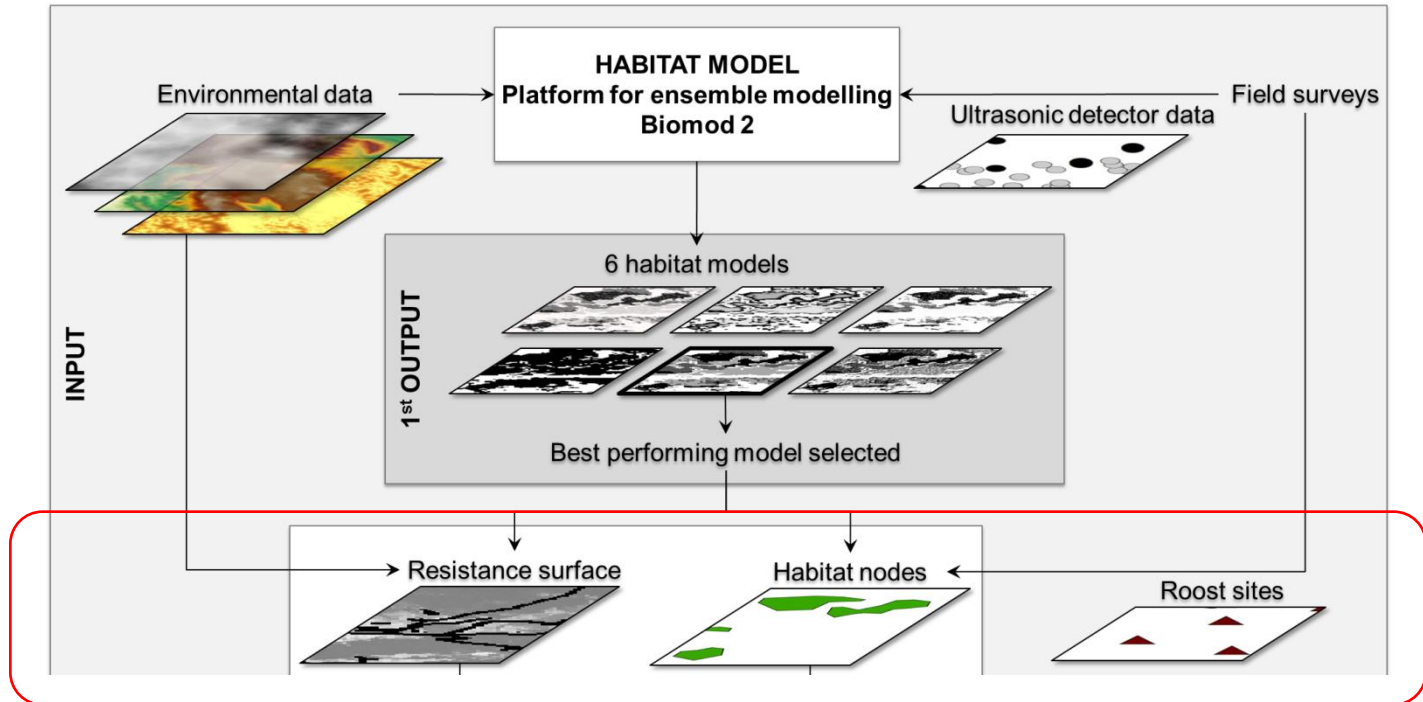
Petit rhinolophe  
(*R. hipposideros*)



Grand rhinolophe  
(*R. ferrumequinum*)



# ➤ Modèle corrélatif intégrant la connectivité paysagère



## Resistance surface

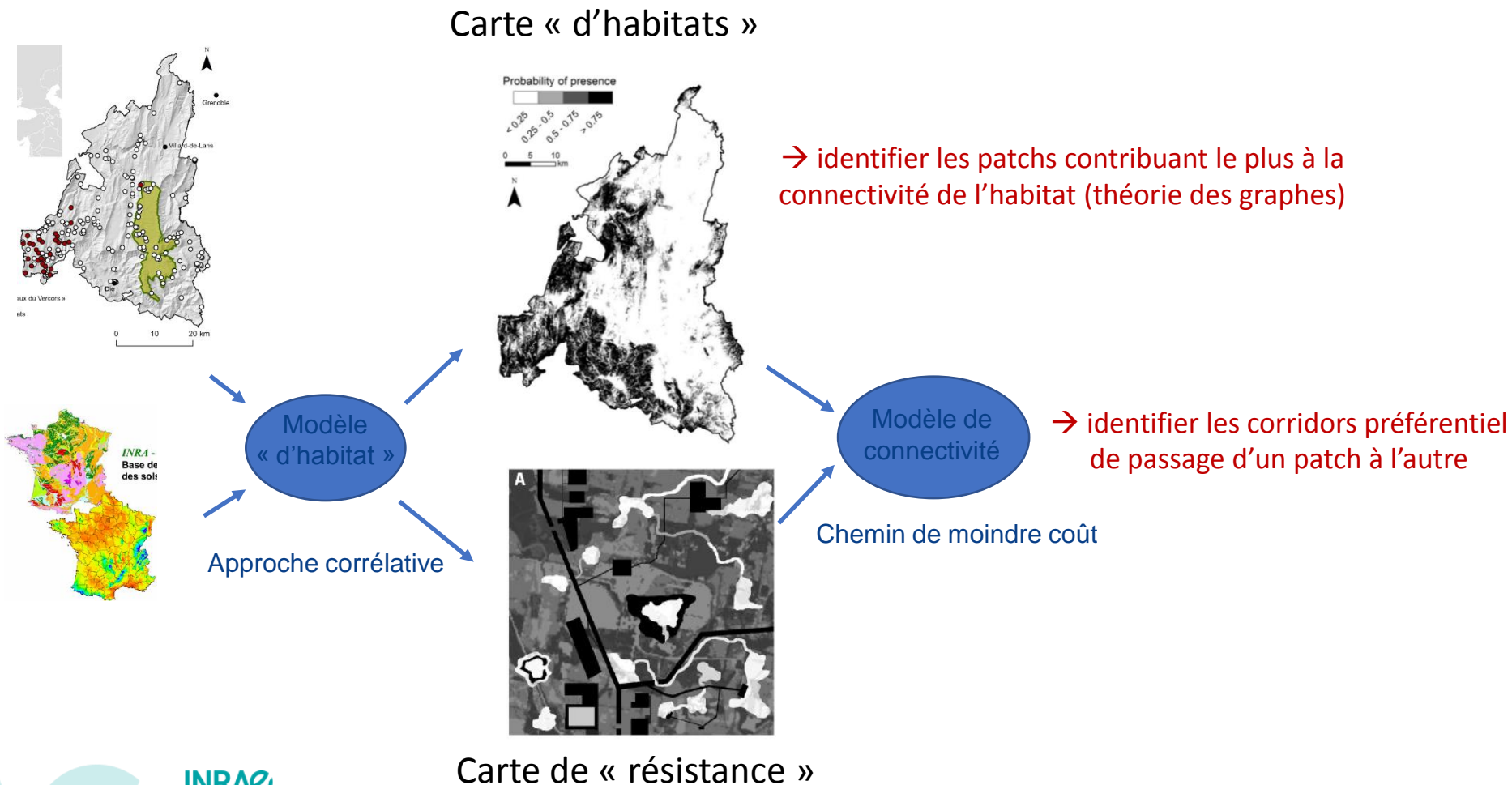
→ postulat : les chauves-souris se déplacent plus facilement dans les meilleurs habitats  
= la résistance au mouvement est d'autant plus faible que la probabilité de présence (P) est forte

## Habitat nodes

→ postulat : habitat si  $P > P_{\text{seuil}}$ , non-habitat si  $P < P_{\text{seuil}}$

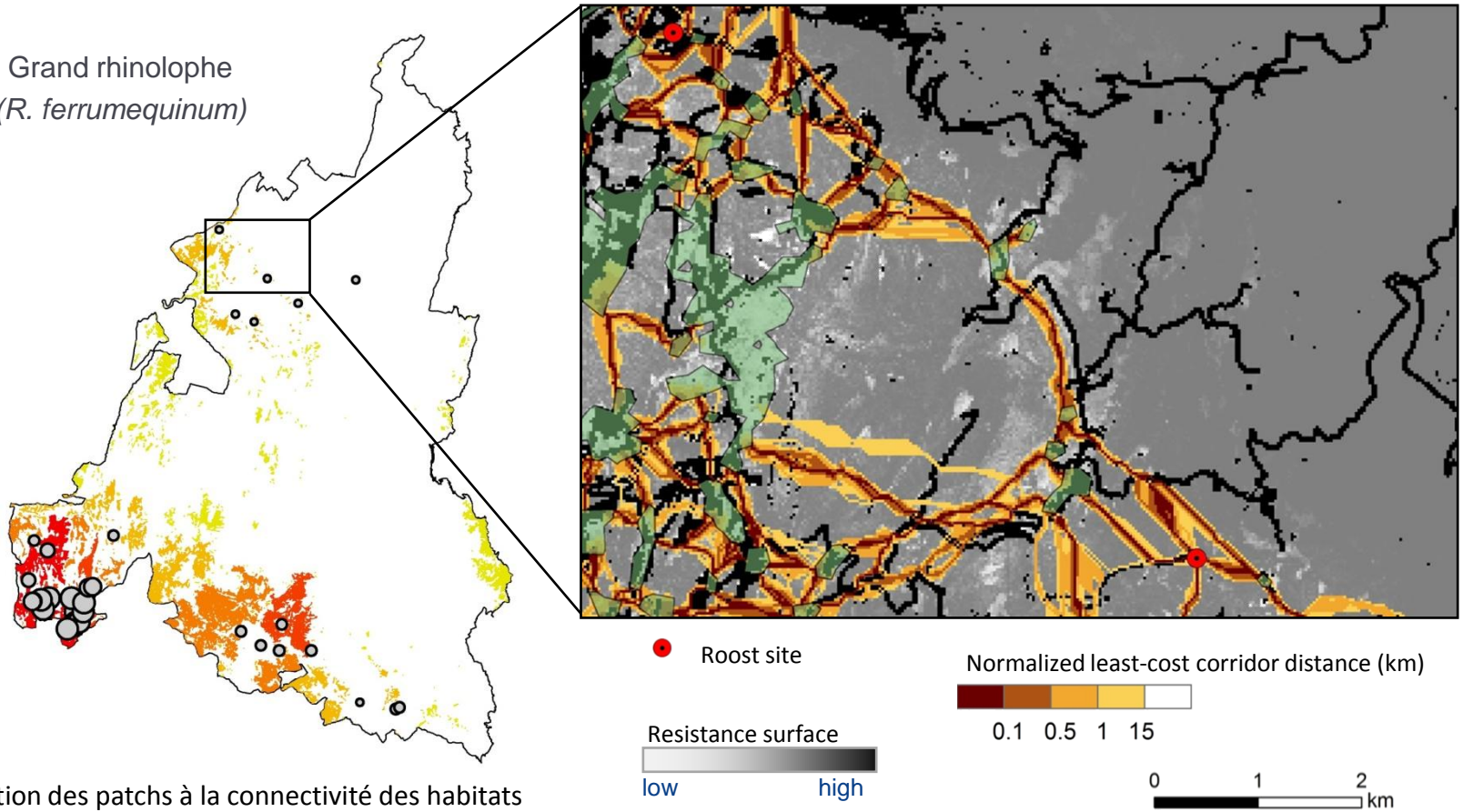


# ➤ Modèle corrélatif intégrant la connectivité paysagère

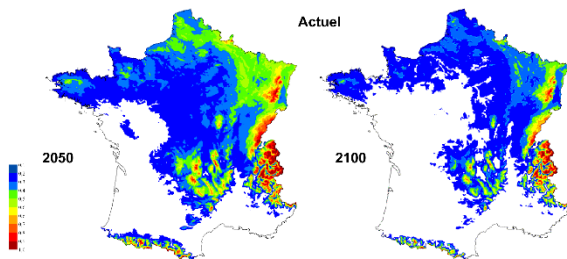
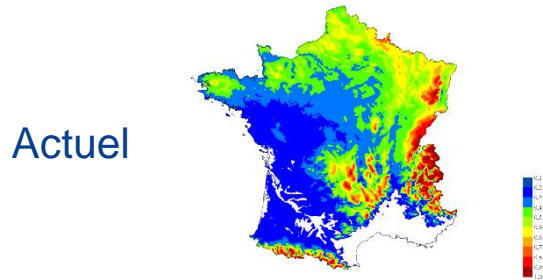


# ➤ Modèle corrélatif intégrant la connectivité paysagère

Grand rhinolophe  
(*R. ferrumequinum*)



# ➤ Limites des approches corrélatives

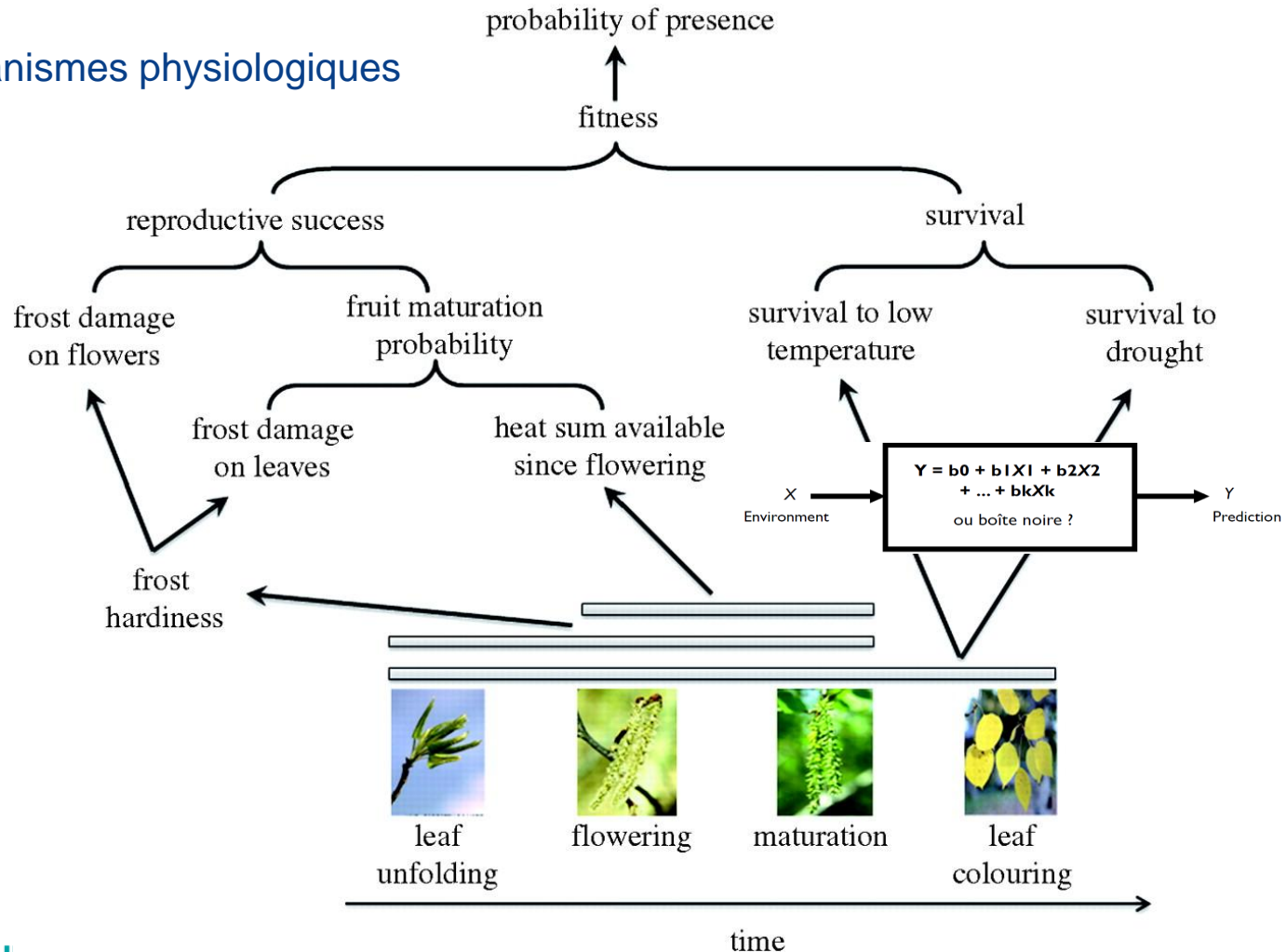


1. Hypothèse d'équilibre avec l'environnement
2. Forte dépendance au contexte
3. Hypothèse de relation invariante avec l'environnement
  - concurrents, pathogènes
  - réponse microévolutive
  - nouvelles conditions (CO<sub>2</sub>)
4. Hypothèse de dispersion sans contrainte



# ➤ Approches mécanistes

Mécanismes physiologiques



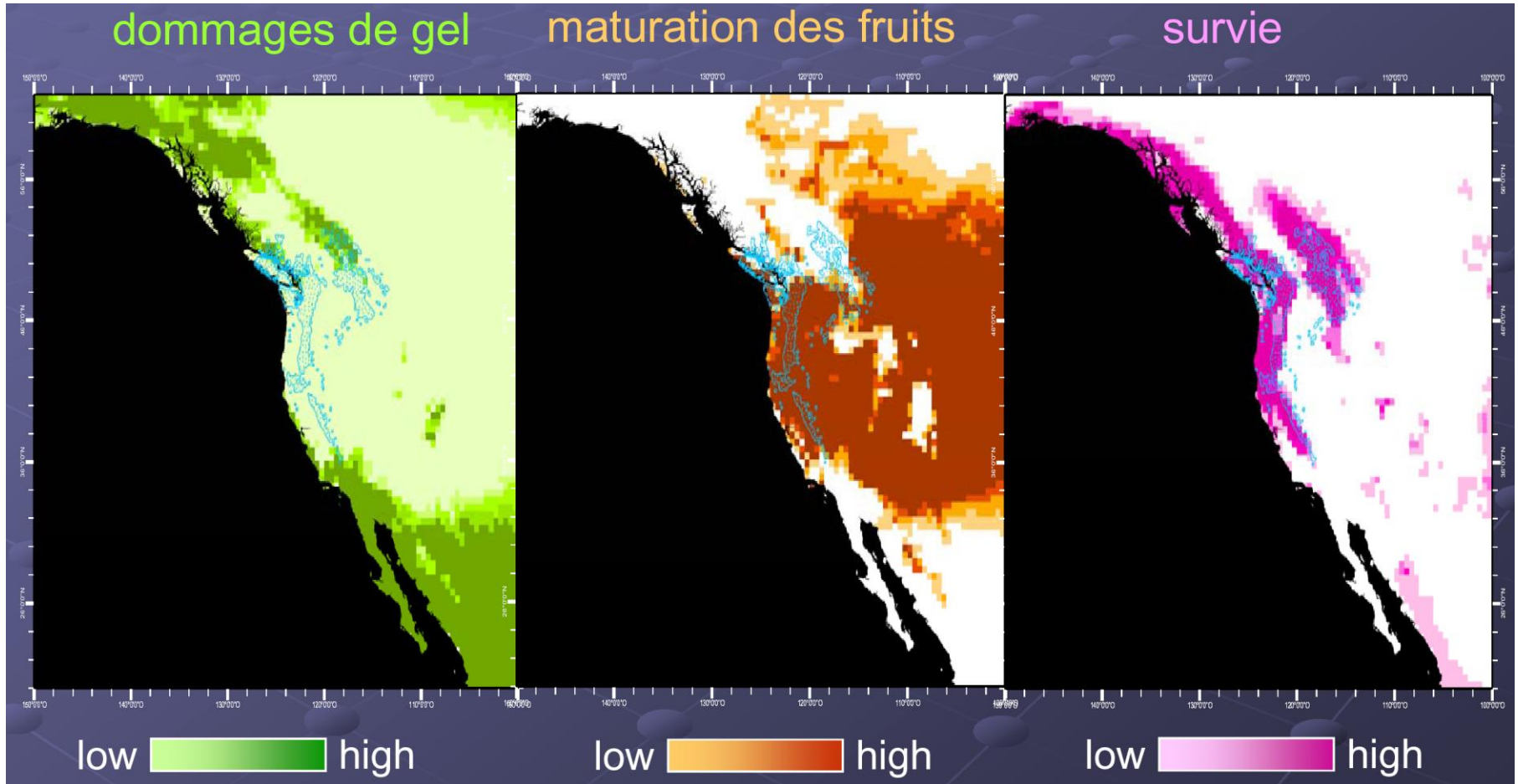
# ➤ Exemple d'approche écophysiological

Mécanismes physiologiques

dommages de gel

maturation des fruits

survie

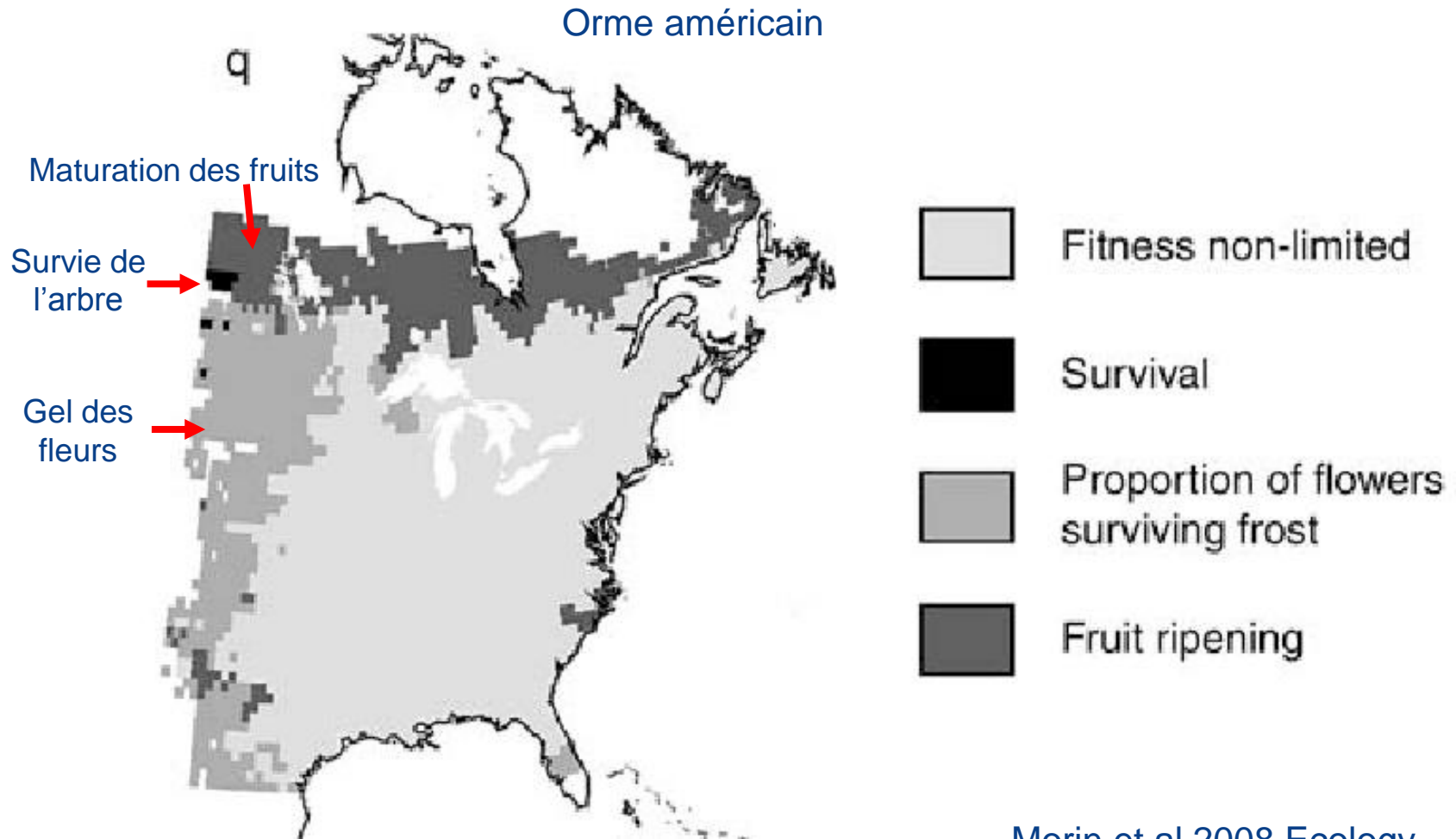


INRAE

Master 1 BICG - FMB

Frédéric Archaux

# ➤ Exemple d'approche écophysiological

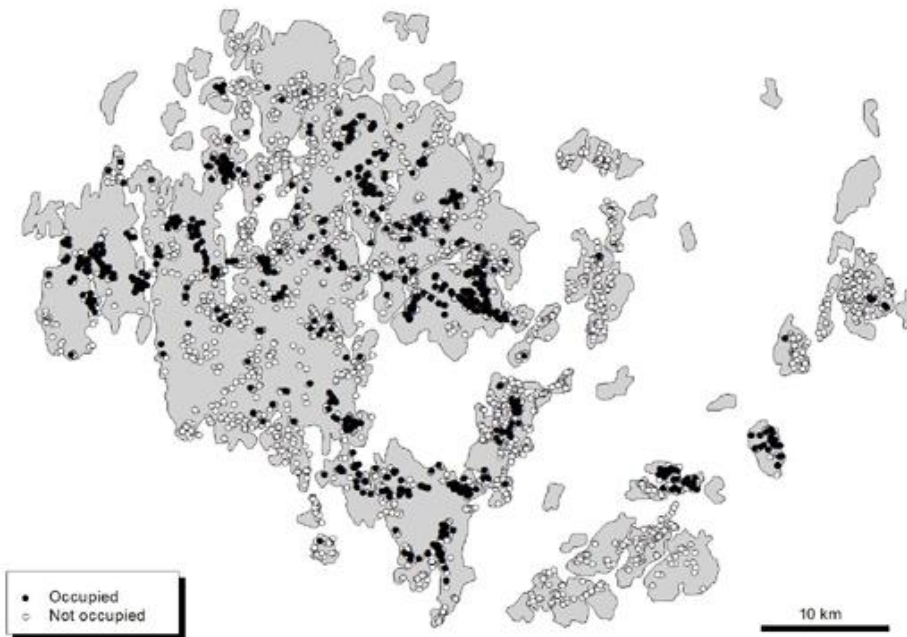


Morin et al 2008 Ecology

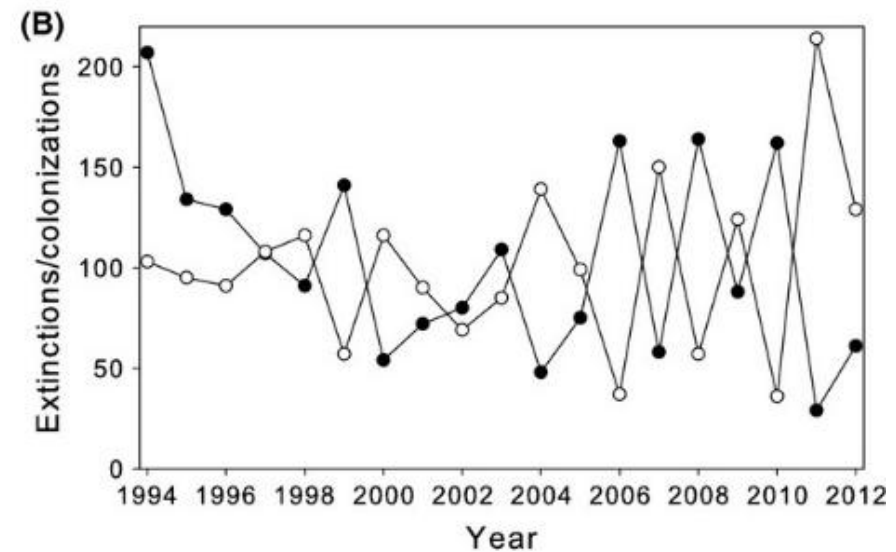
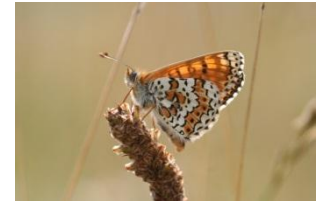
# ➤ Modèle de métapopulation

Une **population** est un ensemble d'individus

Une métapopulation est un **ensemble de populations interconnectées par la dispersion**



Mélitée du plantain  
(*Melitaea cinxia*)

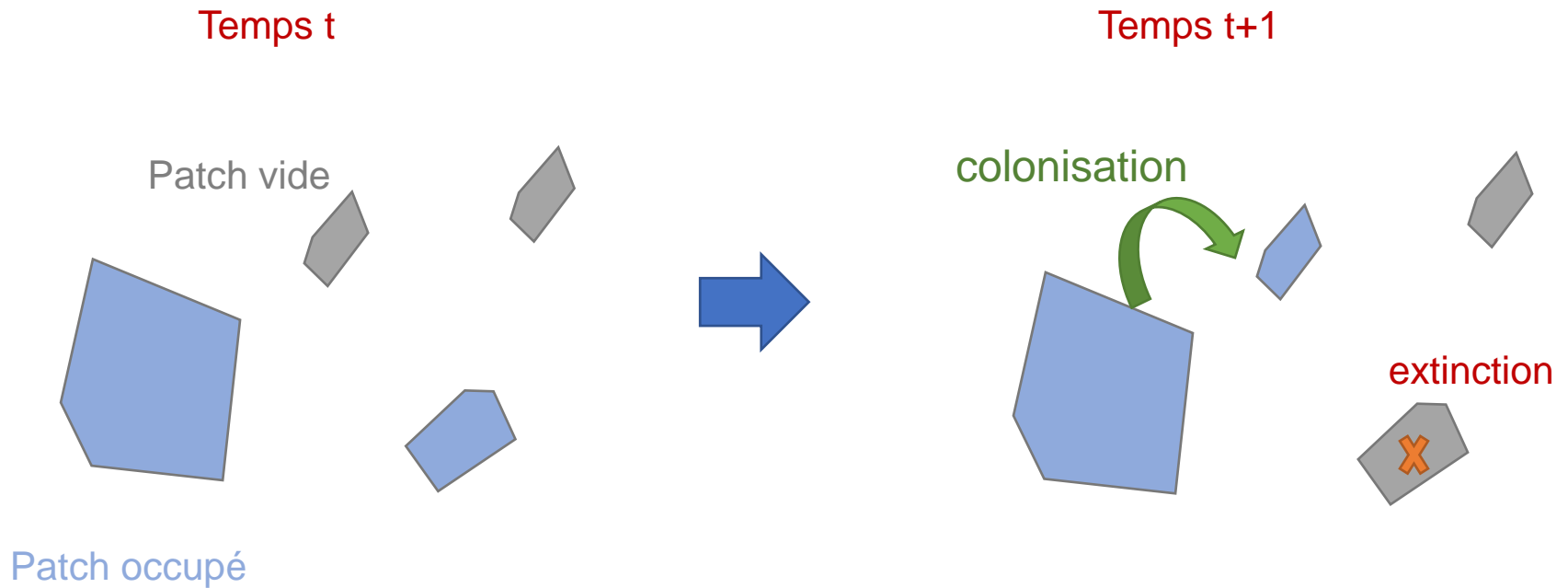


INRAE

Master 1 BICG - FMB

Frédéric Archaux

# ➤ Définition de la métapopulation



## ➤ Hypothèses classiques des modèles métapop

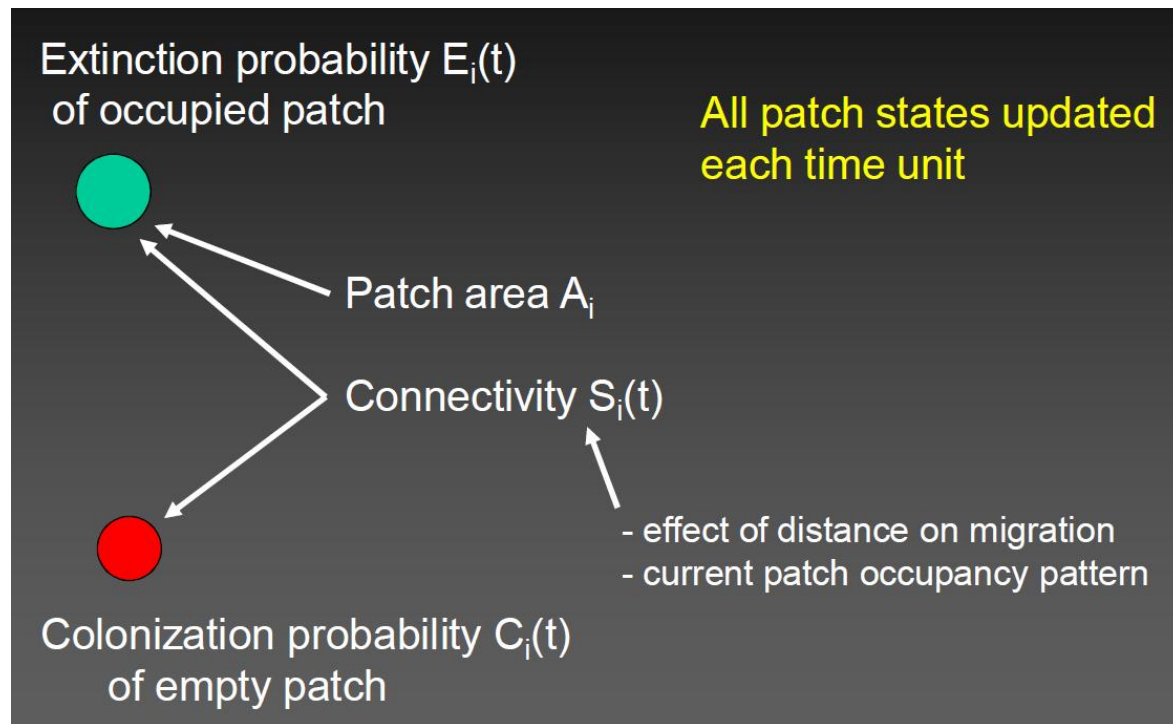
- Pas de modélisation de la dynamique locale (dans le patch)
- Présence/absence dans des patchs (pas d'abondance)
- Les patchs ont une taille, pas une forme
- La qualité des patchs constante dans le temps
- La matrice homogène
- La stochasticité régionale souvent ignorée





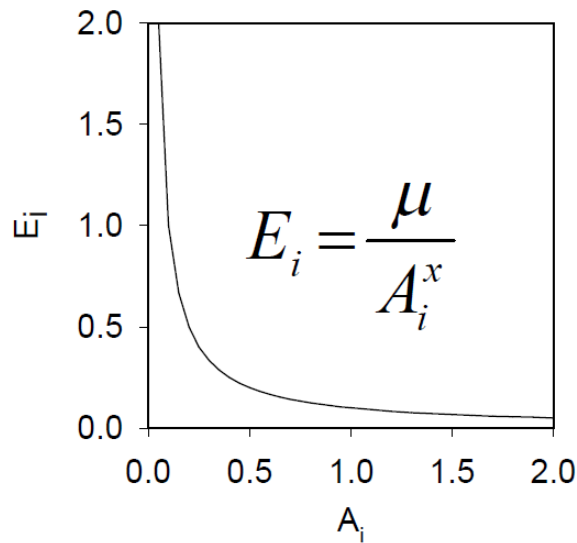
# ➤ Hypothèses classiques des modèles métapop

Stochastic Patch Occupancy Model = modèle de métapopulation



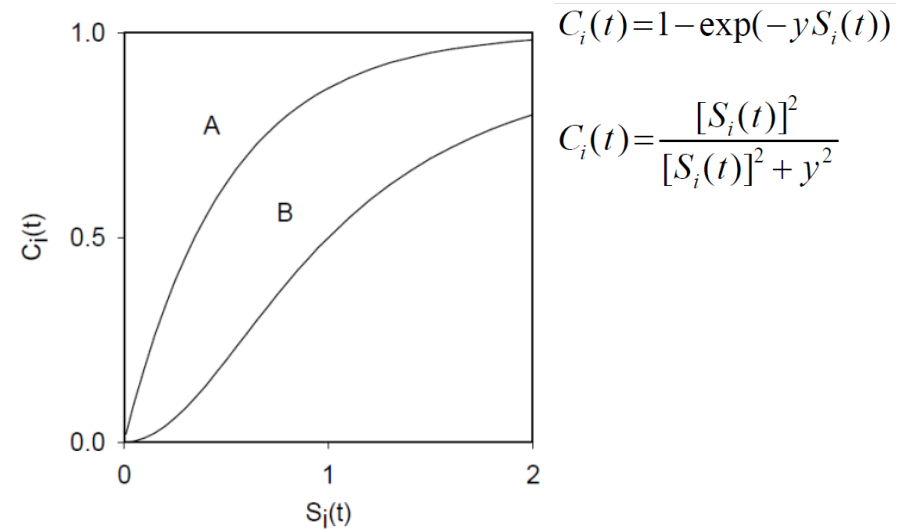
# ➤ Dynamique extinction - colonisation

$E_i$  proba d'extinction dans le patch  $i$



$A_i$  surface du patch  $i$

$C_i$  proba de colonisation dans le patch  $i$



$S_i(t)$  la connectivité du patch au temps  $t$

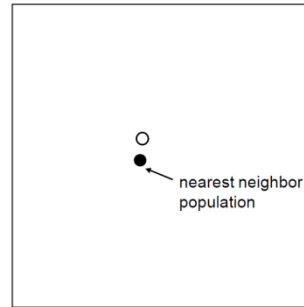
# ➤ Connectivité dans les modèles métapop

La connectivité  $S_i(t)$  dépend du nombre et des distances entre patches

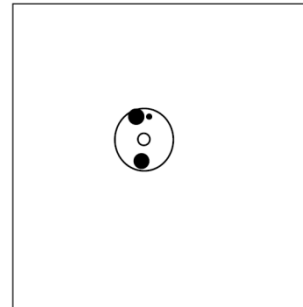


Choix multiples pour sélectionner :

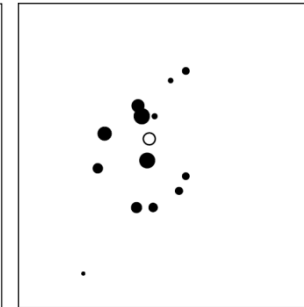
- les patches



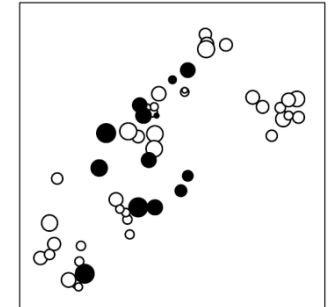
plus proche voisin



buffer

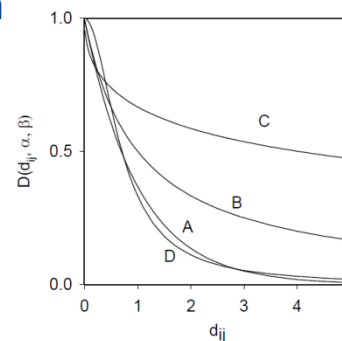


mesures IFM



théorie des graphes

- La métriques de distance entre patch



$$D(d_{ij}, \alpha) = \exp(-\alpha d_{ij})$$

$$D(d_{ij}, \alpha, \beta) = \frac{1}{1 + \alpha d_{ij}^\beta}$$

$d_{ij}$  distance euclidienne entre patches i et j



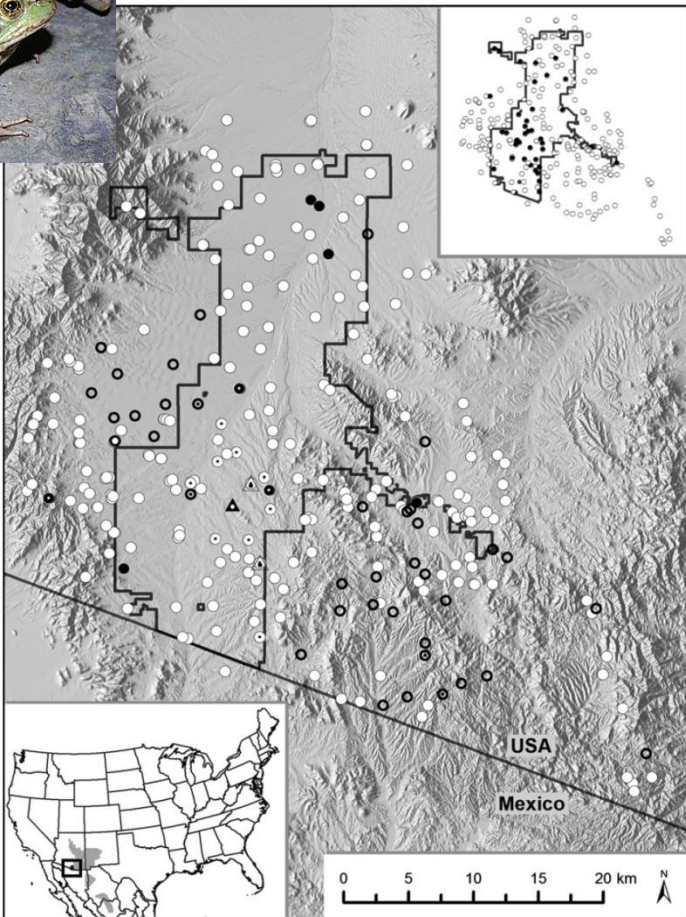
INRAE

Master 1 BICG - FMB

Frédéric Archaux

# ➤ Exemple d'application de modèle métapop

## Chiricahua leopard frog (*Lithobates chiricahuensis*)



Point d'eau

○ intermittent

○ semi-intermittent

● permanent



Sites de réintroduction

Journal of Applied Ecology

Journal of Applied Ecology 2015, 52, 1325–1333

doi: 10.1111/1365-2664.12481

Spatial occupancy models for predicting metapopulation dynamics and viability following reintroduction

Richard B. Chandler<sup>1\*</sup>, Erin Muths<sup>2</sup>, Brent H. Sigafus<sup>3</sup>, Cecil R. Schwalbe<sup>3</sup>, Christopher J. Jarchow<sup>4</sup> and Blake R. Hossack<sup>5</sup>

Occupation du site *i* l'année *k*  
(=0 ou 1)



$$z_{i,k} \sim \text{Bernoulli}(\gamma_{i,k-1}(1 - z_{i,k-1}) + \phi_{i,k-1}z_{i,k-1})$$

Proba de colonisation du site *i* entre l'année *k-1* et *k*

$$\gamma_{i,k} = 1 - \left\{ \prod_{j=1}^M 1 - \rho_{ij,k} \right\}$$

$$\rho_{i,j,k} = \rho_0 \exp(-d_{i,j}^2 / (2\sigma^2)) z_{j,k-1}$$

Occupation du site *i* l'année *k-1*  
(= 0 ou 1)



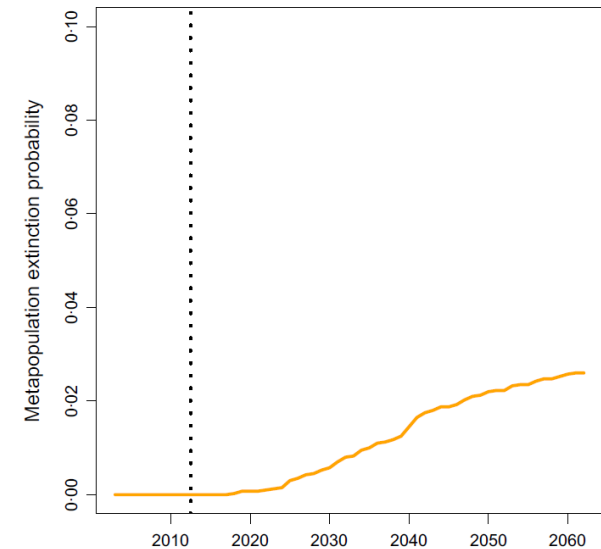
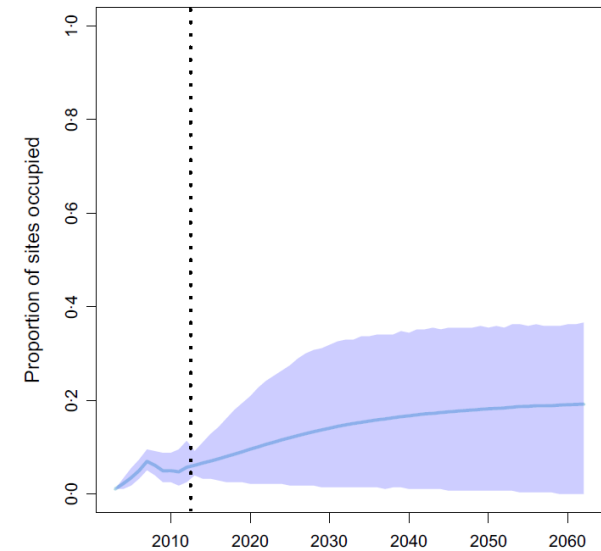
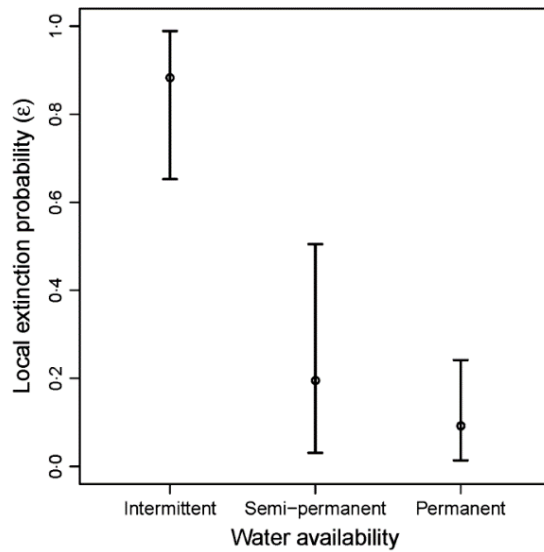
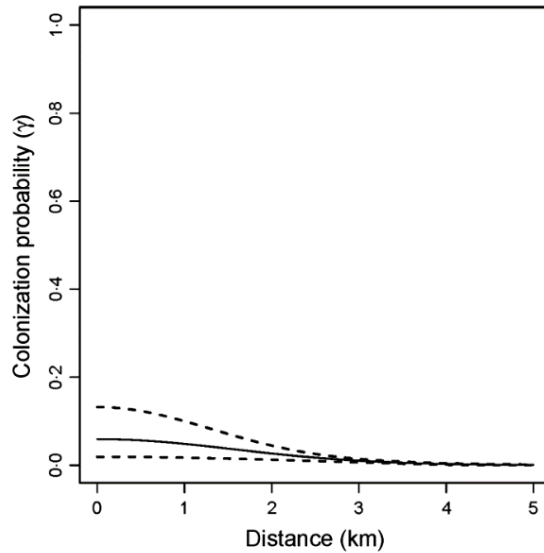
Proba de non-extinction du site *i* entre l'année *k-1* et *k*

Prise en compte de la détection ( $p_{ijk}$ )

$$y_{i,j,k} \sim \text{Bernoulli}(z_{i,k} \times p_{i,j,k})$$

$$\text{logit}(p_{i,j,k}) = \beta_0 + \beta_1 \text{TEMP}_{i,j,k} + \beta_2 \text{PRECIP}_{i,j,k}$$

# ➤ Exemple d'application de modèle métapop



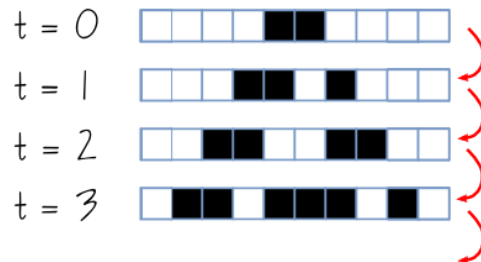
# > Automates cellulaires

Découpage de l'espace en une grille de cellules (pixels)



Chaque pixel possède des caractéristiques  
Chaque pixel est occupé ou non au temps  $t$

Processus à l'échelle du pixel dépendant des pixels voisins



Extinction

Productivité

Colonisation

Règles (dont modèles)



# ➤ Exemple d'application d'automate cellulaire

## Règles (dont modèles)

### Survie/mortalité

Age reproduction >20 ans

Survie = f(favorabilité de l'habitat)

Fécondité = f(favorabilité de l'habitat)

Global Change Biology

Global Change Biology (2012) 18, 2648–2660, doi: 10.1111/j.1365-2486.2012.02679.x

**Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: the case of *Quercus pubescens* in France**

ROMAIN BERTRAND\*†, VINCENT PEREZ\*† and JEAN-CLAUDE GÉGOUT\*†

### Dispersion

$$P_{Sdisp}(d) = e^{(d - Psize) \times \left( \ln \left( \frac{1-k}{EffDist} \right) \right)} - e^{(d \times \ln \left( \frac{1-k}{EffDist} \right))}$$

distance (pointing to  $d$ )  
taille du pixel=1 km (pointing to  $Psize$ )

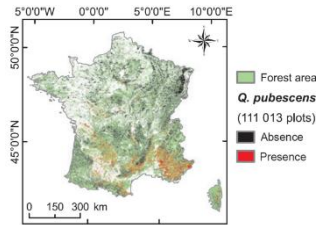
Colonisation = f(artificialisation + glaciers du pixel)

Usages des sols fixes dans le temps

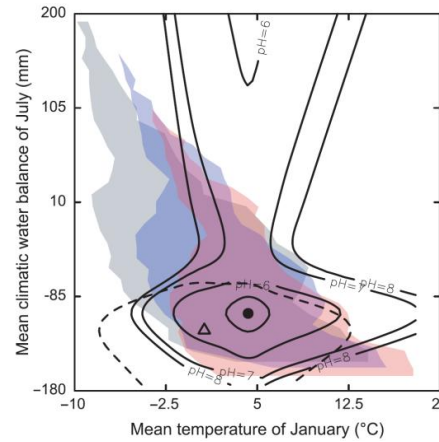
# ➤ Exemple d'application d'automate cellulaire



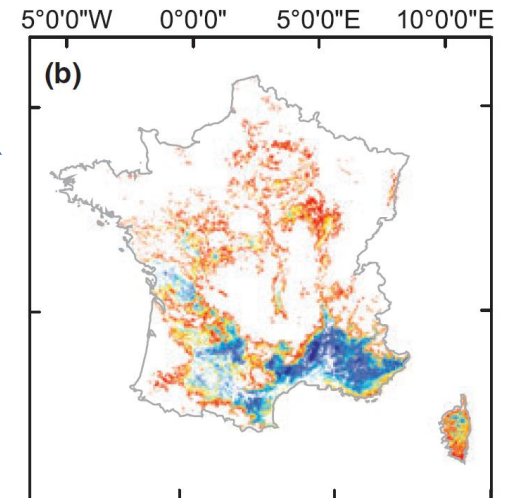
## Données IFN-IGN



## Modèle (corrélatif) de niche



## Extrapolation spatiale

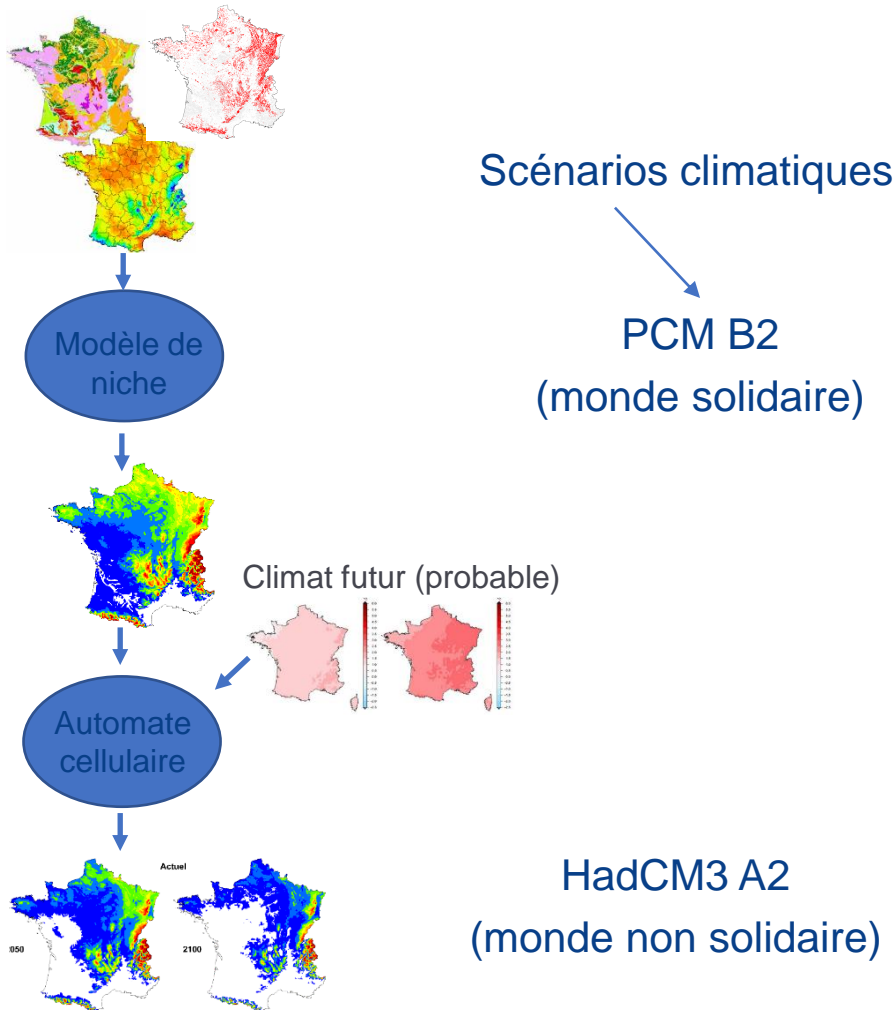


$$g(P_{Qpub}) = s_3(Tm_1) + s_4(CWB_7) + s_5(pH)$$

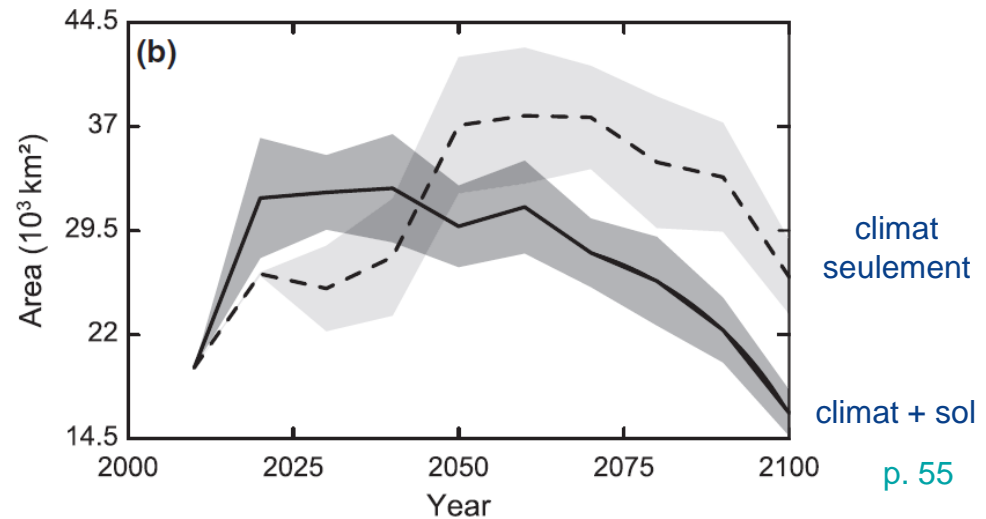
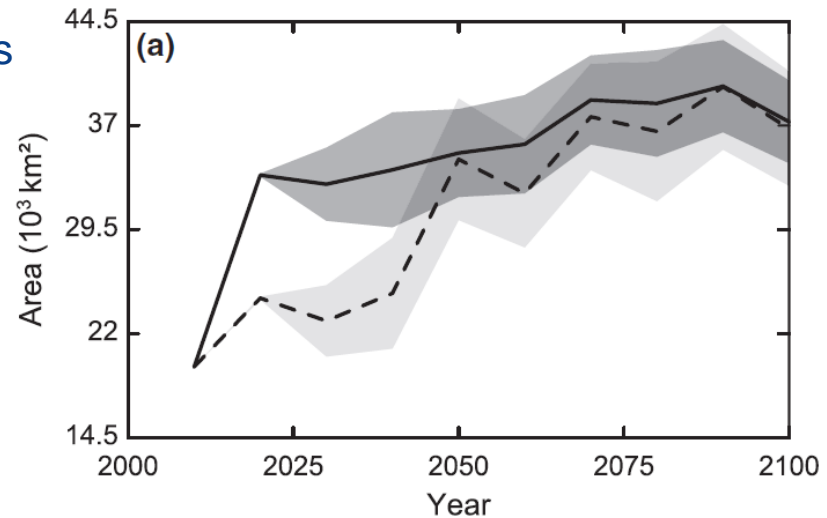
proba de présence

relation non linéaire

# ➤ Exemple d'application d'automate cellulaire



## Prédictions



# ➤ Modèle individu-centré

## Dénomination

- individual-based model (IBM)
- agent-based model (ABM)

## Principe

- Modéliser le comportement/cycle de vie (de chacun) des individus



## ➤ Exemple de modèle individu-centré

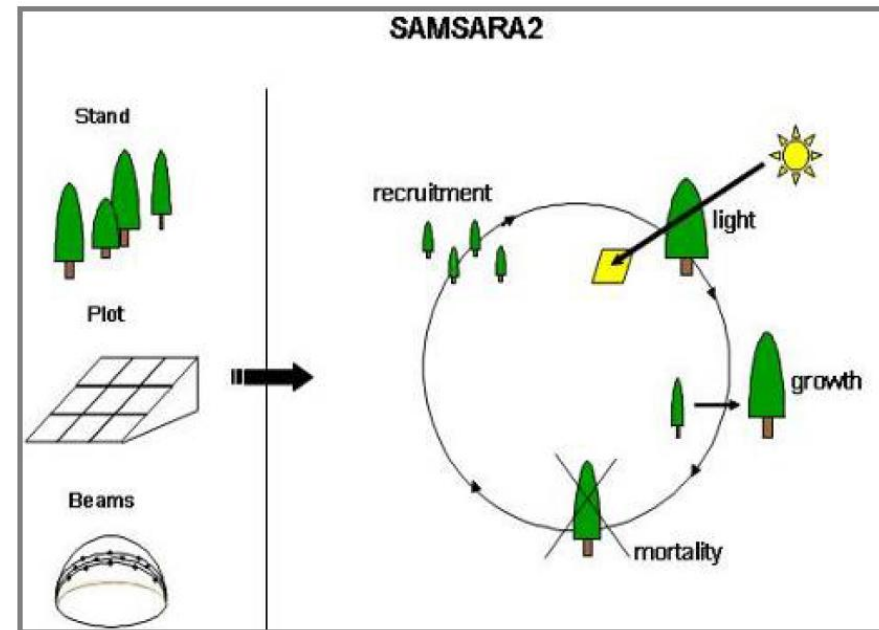
Modèle de dynamique pour les **sapinières-pessières irrégulières** de montagne SAMSARA2 (Courbaud et al.)

### • Caractéristiques :

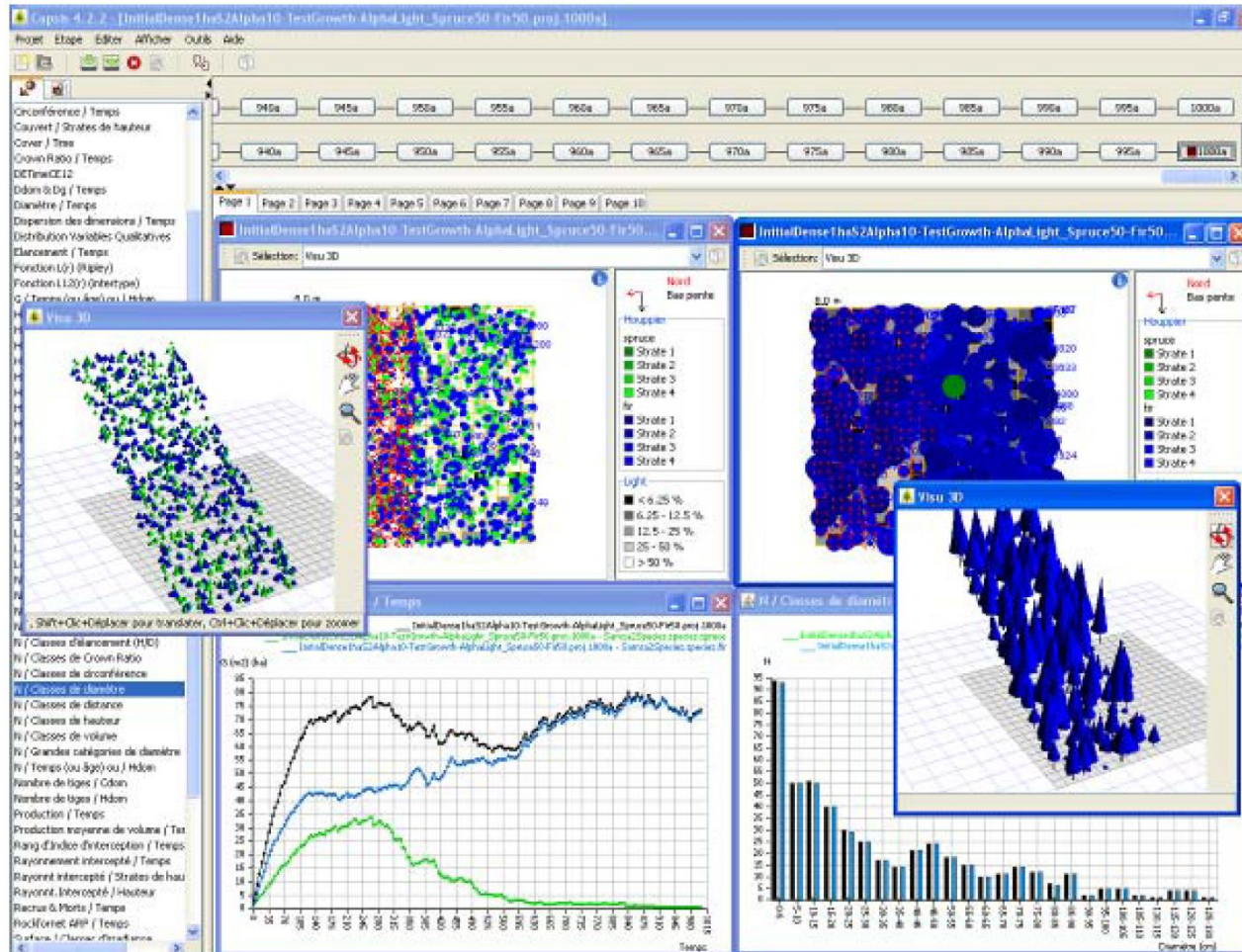
- individu-centré
- spatialement explicite
- module de transmission de la lumière

### • Modèles :

- croissance
- mortalité
- recrutement

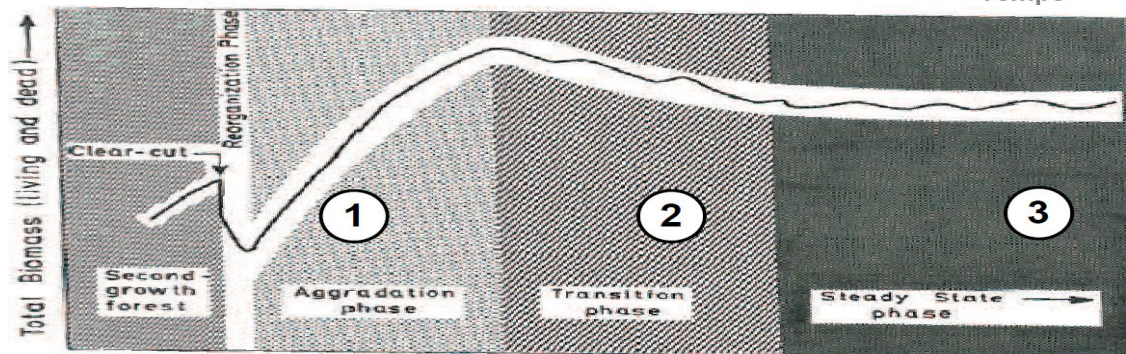
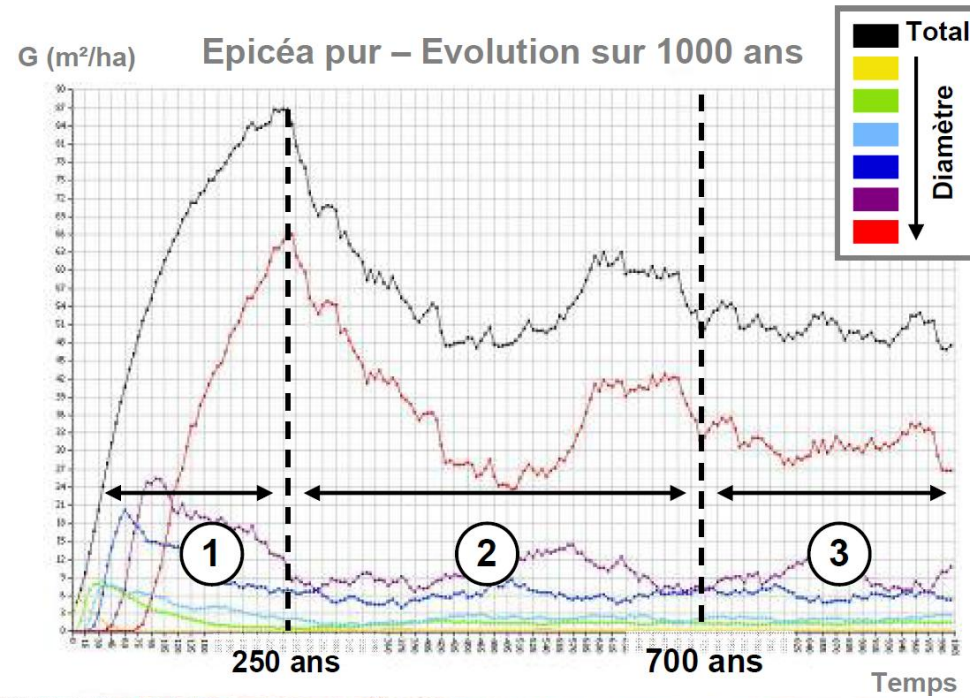


# ➤ Exemple de modèle individu-centré





# ➤ Exemple de modèle individu-centré



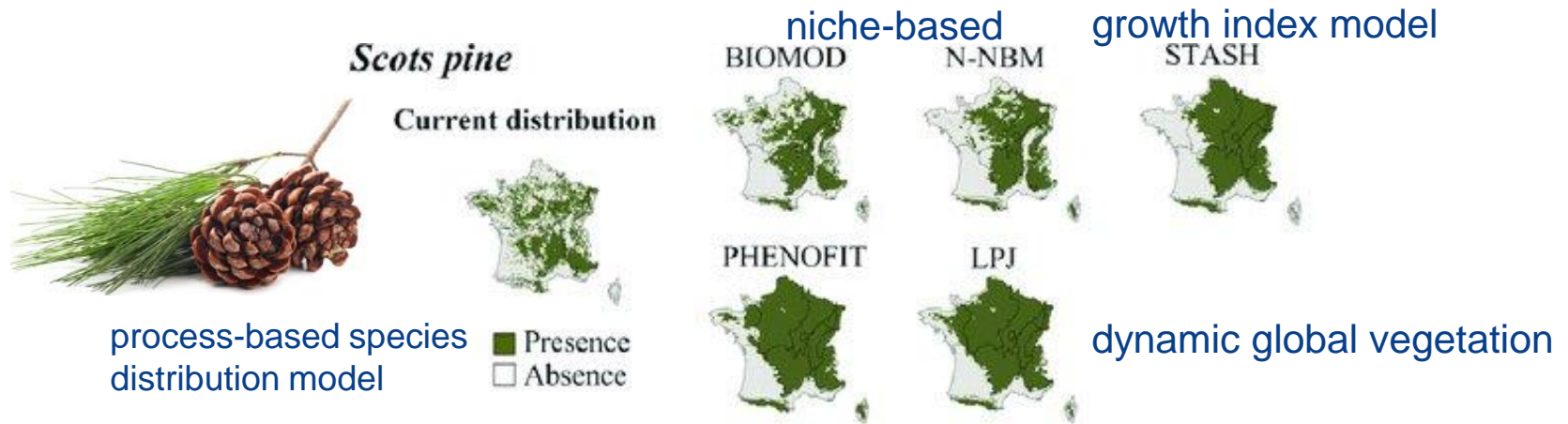
(Bormann et Likens 1979)

## ➤ Coûts-bénéfices des modèles mécanistes

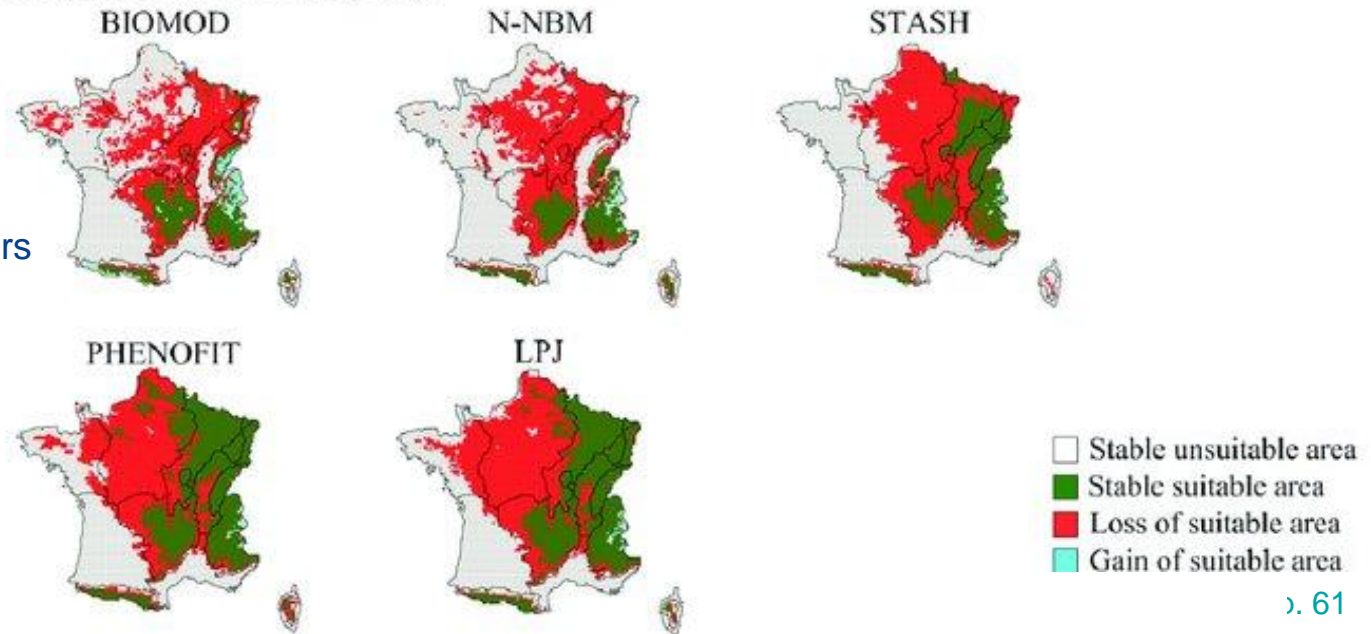
	Mécaniste	Empirique
Besoin en données	Fort	Faible
Temps de de calibration et de traitement	Long	Court
Nombre d'espèces traitées	Faible	Elevé
Puissance explicative	Forte	Faible
Projection future	Possible	Erronée
Validation	Indépendante	Croisée
Approximation de la niche...	Fondamentale	Réalisée



# ➤ Inférence multi-modèles (1)



## Predicted future distribution (2055)

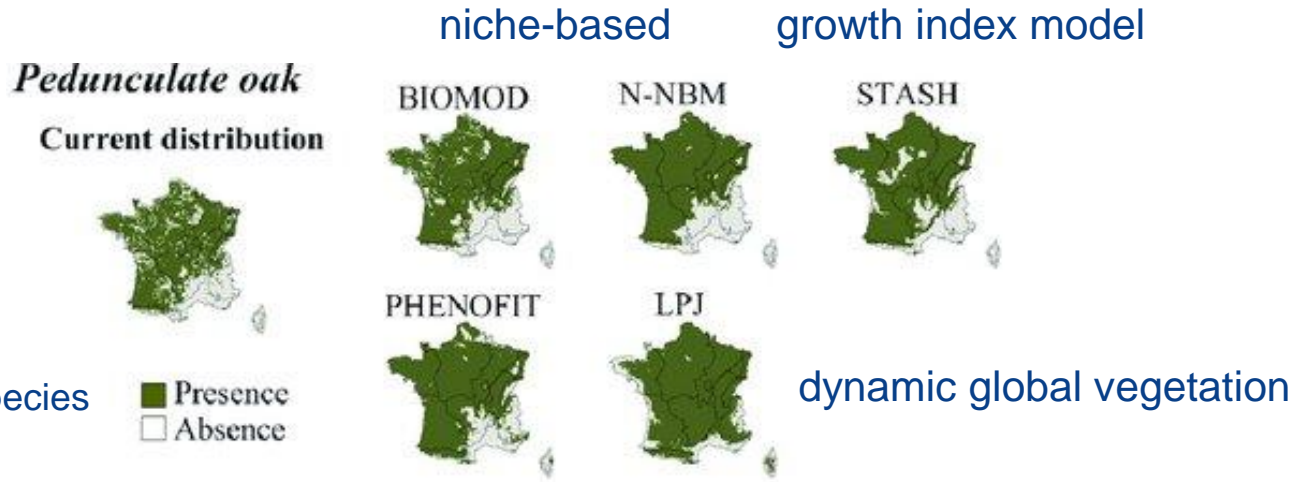


Cheaiba et al 2012 Ecol Letters

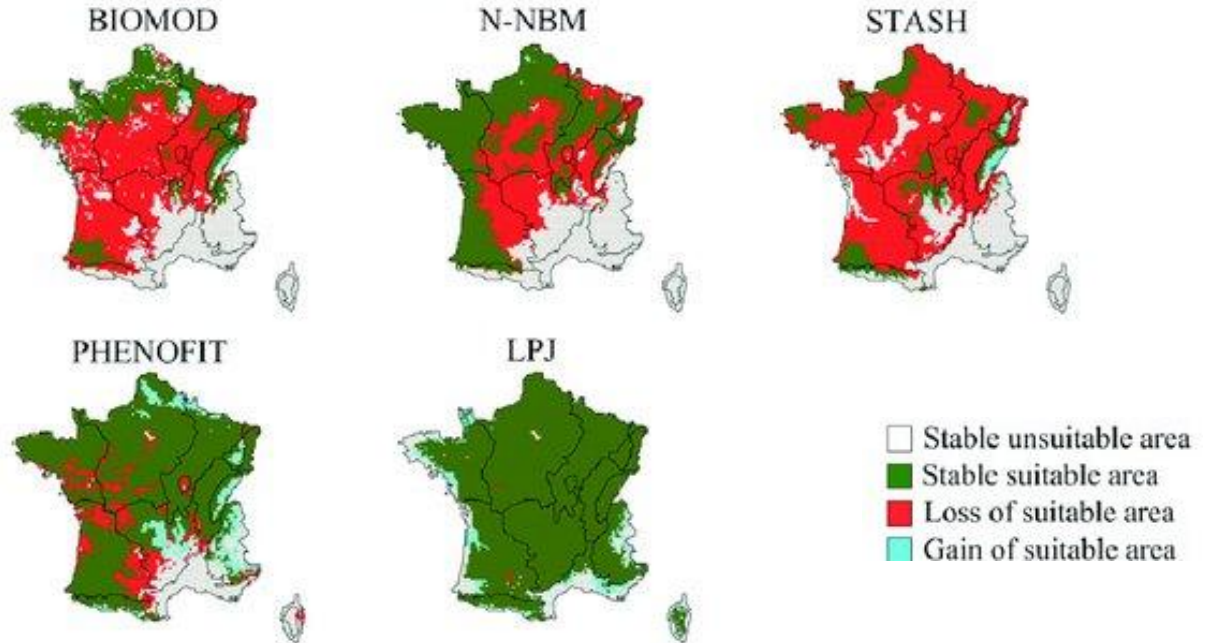
INRAE

Master 1 BICG - FI  
Frédéric Archaux





Predicted future distribution (2055)



INRAE

Master 1 BICG - FMB

Frédéric Archaux

## ➤ Messages clés

Modèles = outils classiques de l'écologie (emboîtements de modèles...)

Modèles = simplification de la réalité

Modèles = le plus complexe n'est pas forcément le meilleur/plus utile

Modèles = objectifs, présupposés, domaine de validité, incertitudes

