

Modèles de prédictions de répartition (niches)

Présentation 1: Concepts, hypothèses, approches et limites

Présentation 2: Utilisation en écologie et verrous méthodologiques

Présentation 3: Cas particulier des invasions biologiques

Modèles de prédictions de répartition (niches)

Présentation 1: Concepts,
hypothèses, approches et limites

Modèles : pourquoi et pour quoi ?

- Pour tester une hypothèse
- Pour décrire (quantifier) la relation entre une variable réponse (y) et une ou plusieurs variables explicatives (x_i)

$$y = f(x_i)$$

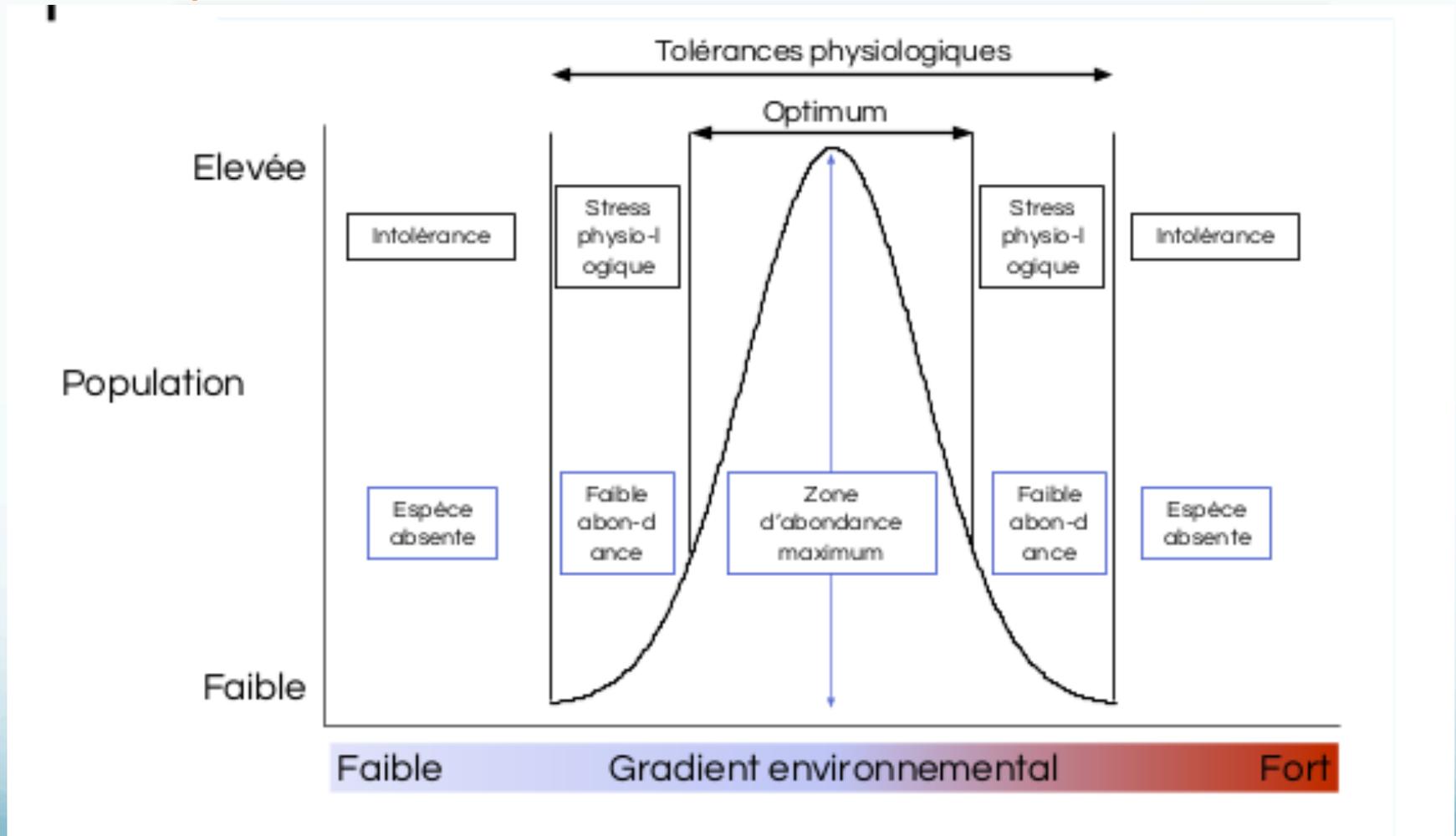
- Pour prédire la valeur la plus probable de la variable réponse à partir de valeurs de x_i :
 - Pour une autre période (modèle temporel)
 - Pour une autre région (modèle spatial)

Exemple : estimer les impacts potentiels des changements globaux sur la biodiversité



- Nécessite :
 - De comprendre les facteurs environnementaux influençant les espèces
 - De les quantifier dans des modèles grâce aux patrons actuels
 - D'utiliser ces modèles pour prédire les patrons futurs

Réponse hypothétique des espèces à un gradient climatique



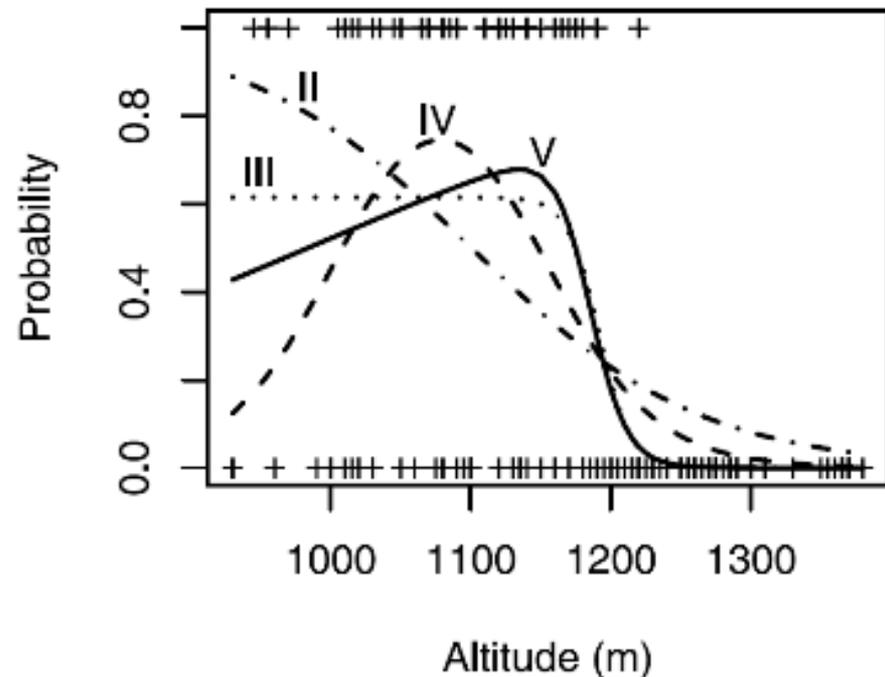
Réponse hypothétique des espèces à un gradient climatique

Différentes formes de réponse selon les variables

Modèles Huisman-Olff-Fresco (HOF)

$$\mu = M \frac{1}{1 + \exp(a + bx)} \frac{1}{1 + \exp(c - dx)}$$

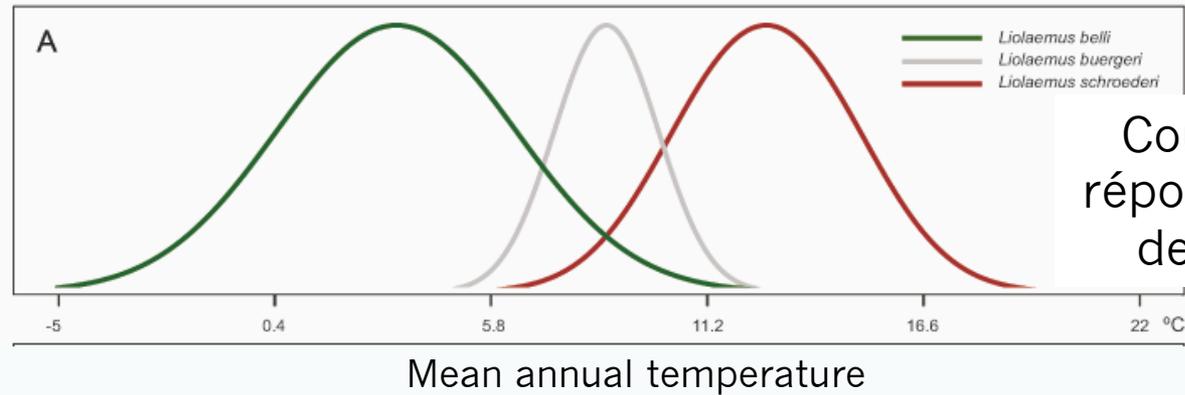
Model		Parameters			
V	Skewed	a	b	c	d
IV	Symmetric	a	b	c	b
III	Plateau	a	b	c	0
II	Monotone	a	b	0	0
I	Flat	a	0	0	0



Réponses des espèces à un gradient climatique



Liolaemus lizard in central Chile

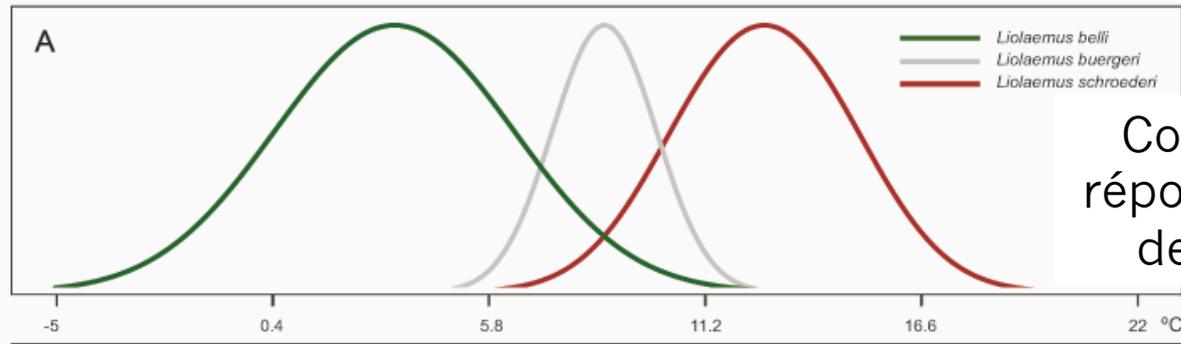


Courbes de réponses issus des SDMs

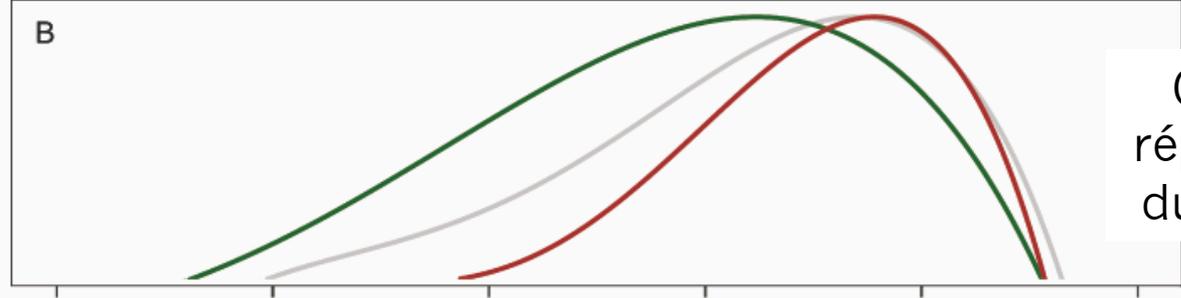
Réponses des espèces à un gradient climatique



Liolaemus lizard in central Chile



Courbes de réponses issus des SDMs

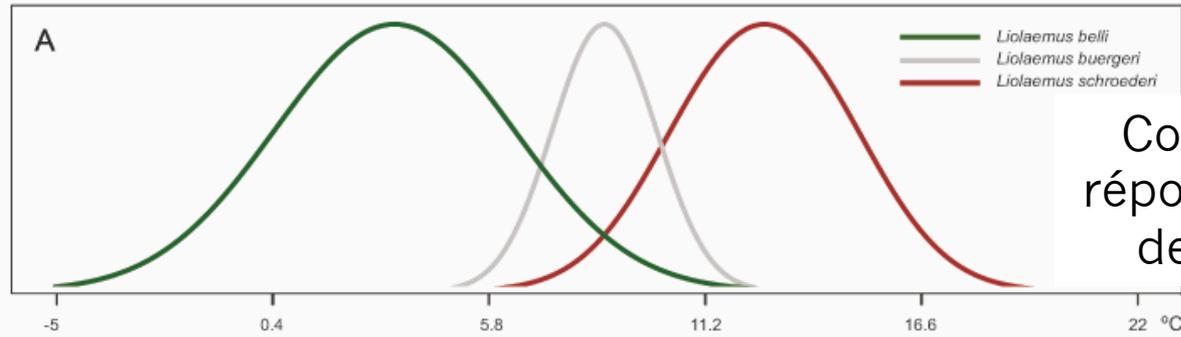


Courbes de réponses issu du laboratoire

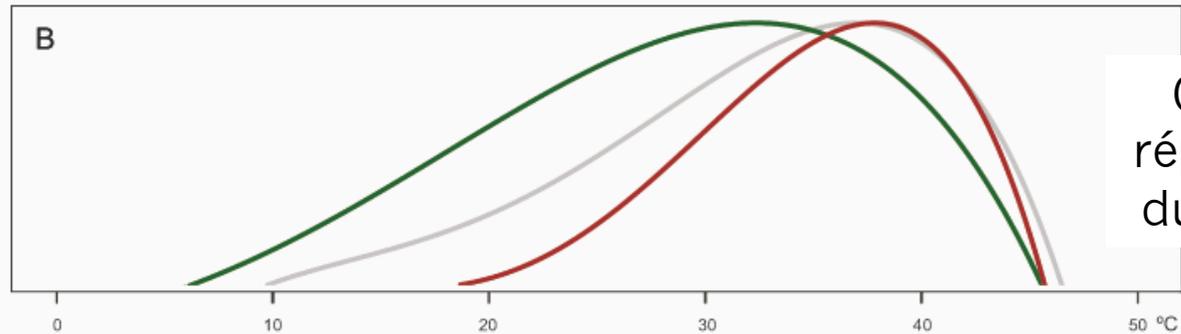
Réponses des espèces à un gradient climatique



Liolaemus lizard in central Chile

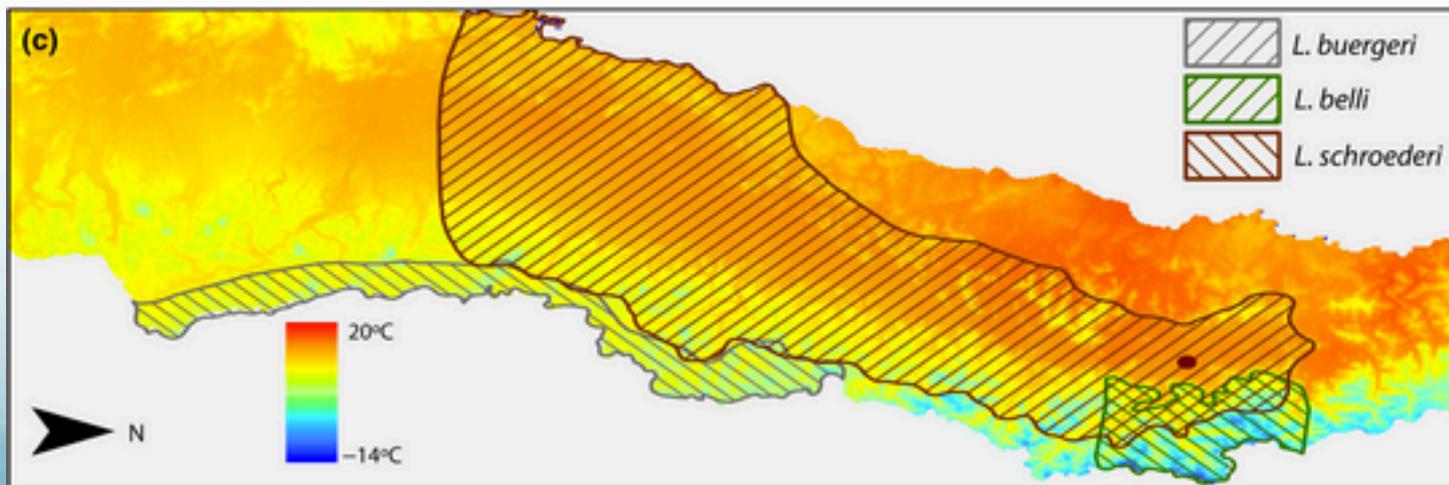


Courbes de réponses issus des SDMs

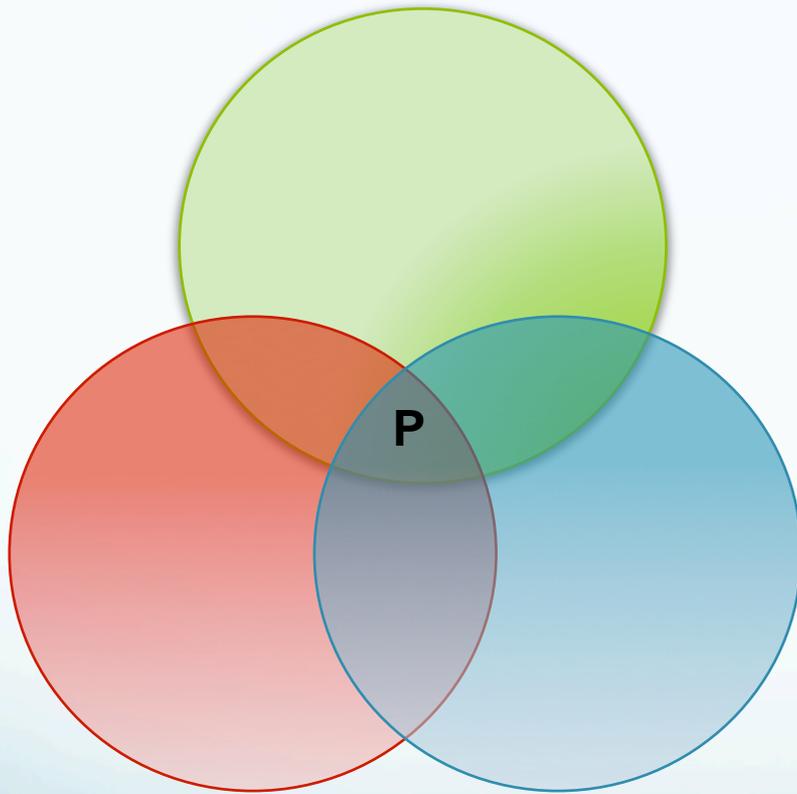


Courbes de réponses issu du laboratoire

Geographical distributions of the three species of *Liolaemus* overlaid on mean annual temperature



Facteurs contrôlant la distribution des espèces



- Climat
- Dispersion
- Interactions biotiques

P is the region that has the right set of biotic and abiotic factors and that is accessible to the species (geographic distribution of the species)

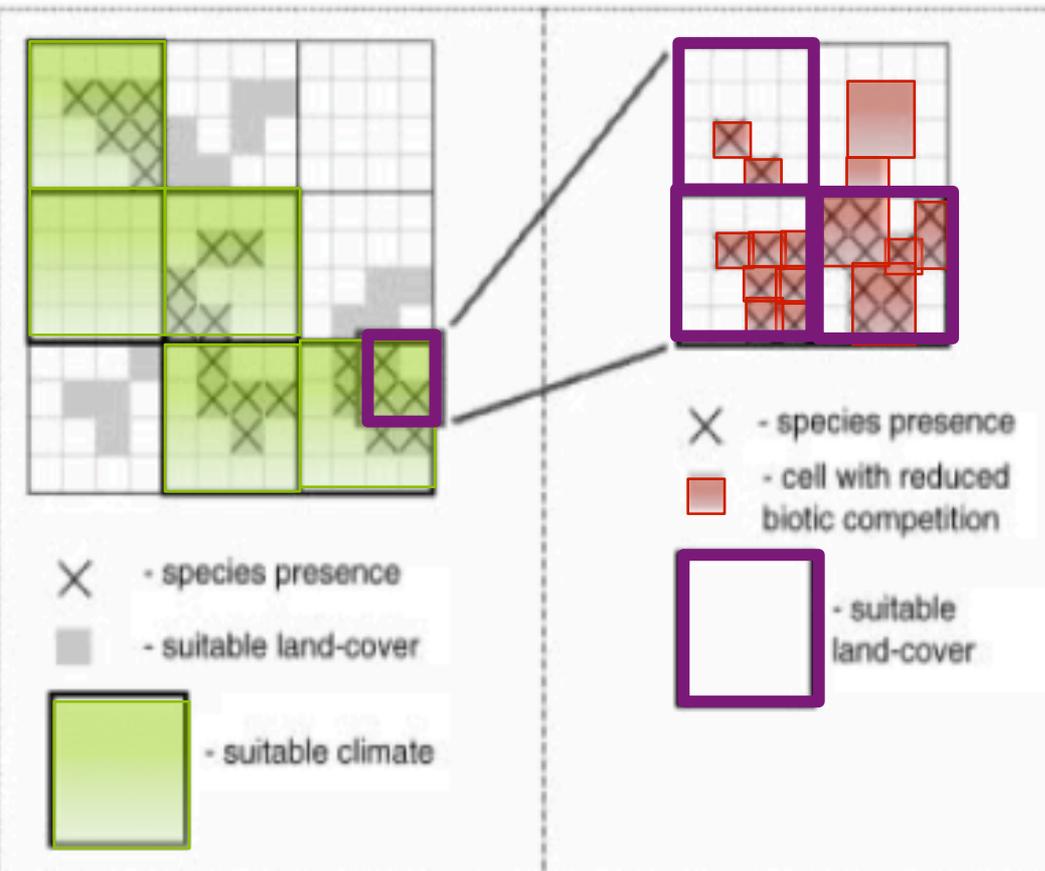
Facteurs contrôlant la distribution des espèces : A quelle échelle ?

-Climat

-Dispersion

-Interactions biotiques

-Habitat



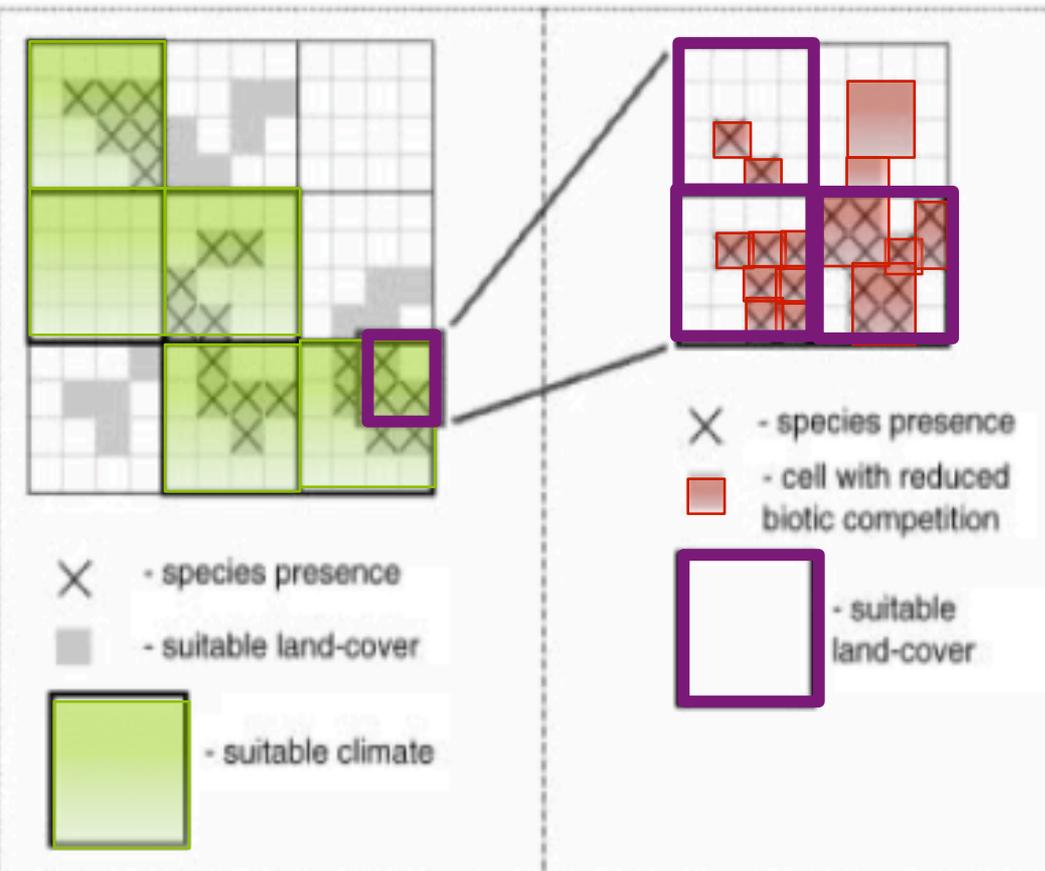
Facteurs contrôlant la distribution des espèces : A quelle échelle ?

-Climat

-Dispersion

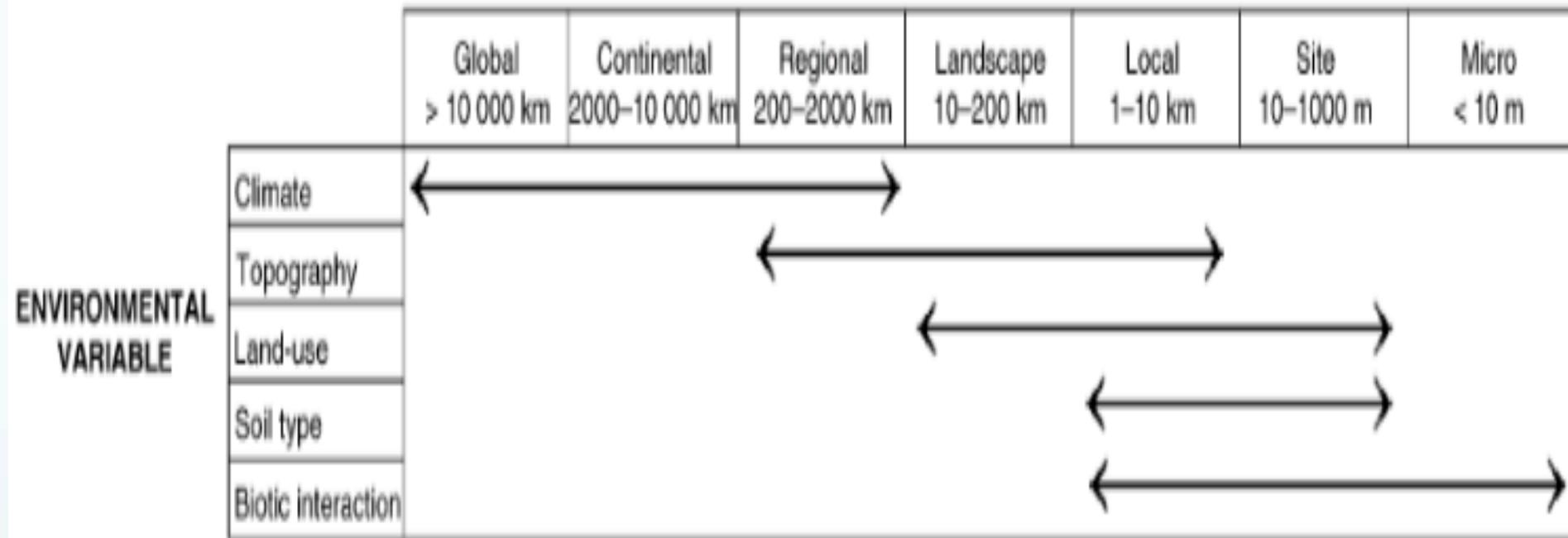
-Interactions biotiques

-Habitat



La distribution d'une espèce est d'abord définie par ses **tolérances climatiques**, si la résolution est de 50 km², alors que lorsqu'on diminue de résolution, (-> 5 km²), le **type d'habitat** peut devenir le facteur prédominant la distribution des espèces. A une résolution encore plus fine (<1 km²), les **interactions biotiques** peuvent, elles aussi, devenir prédominantes.

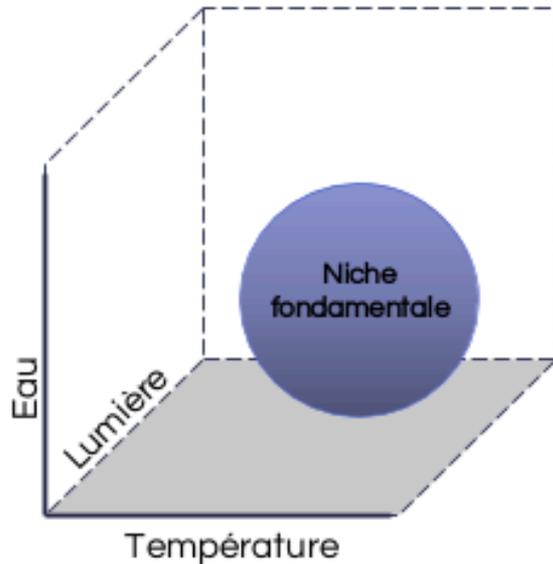
Facteurs contrôlant la distribution des espèces : A quelle échelle ?



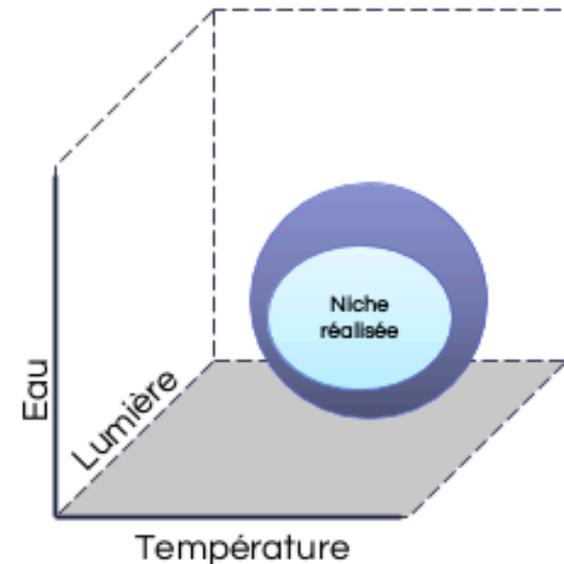
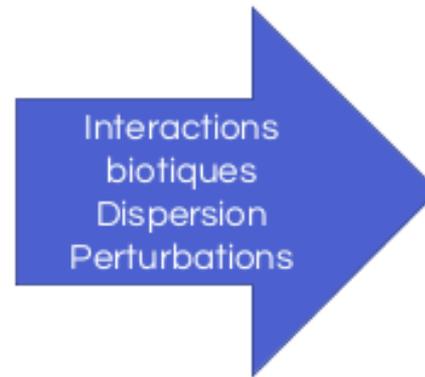
La distribution d'une espèce est d'abord définie par ses tolérances climatiques, si la résolution est de 50 km², alors que lorsqu'on diminue de résolution, (-> 5 km²), le type d'habitat peut devenir le facteur prédominant la distribution des espèces. A une résolution encore plus fine (<1 km²), les interactions biotiques peuvent, elles aussi, devenir prédominantes.

Basés sur un concept clé – la niche fondamentale

Différence entre niche fondamentale et niche réalisée



Ensemble des conditions dans lesquelles l'espèce peut vivre et se reproduire

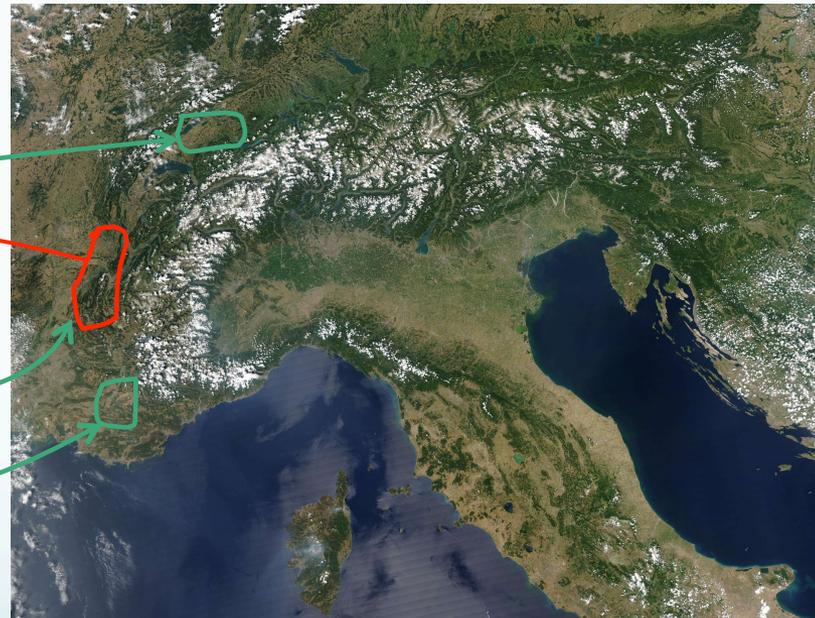
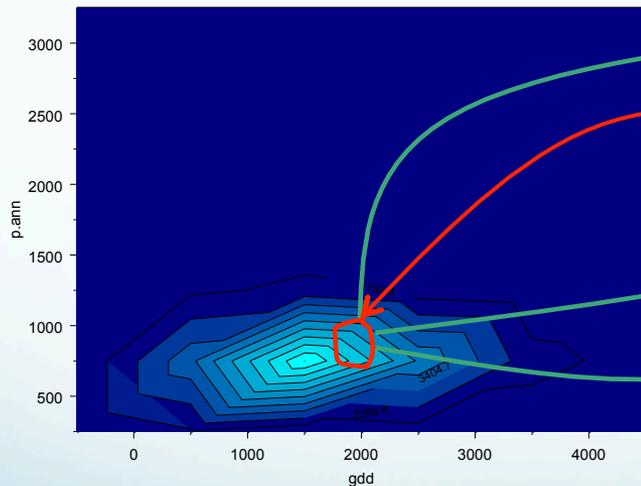


Ensemble des conditions réelles dans lesquelles l'espèce vit et se reproduit

Lien entre niche et distribution

Concept I:

To each region in geographical space there are corresponding ones in environmental space and vice-versa



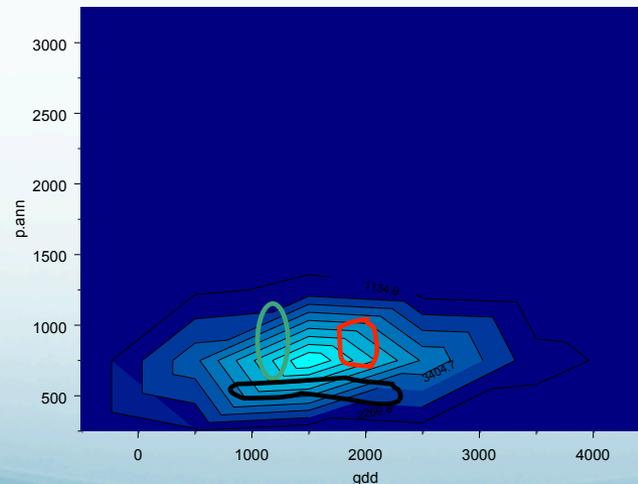
Ecological space (E),
 n dimensions

Geographical space (G),
2 dimensions

Lien entre niche et distribution

Concept II :

- Niches are subsets of spaces of environmental variables corresponding to subsets of geographical space where a species live, or can live
- The *function* that maps from **G** to **E** is defined very simply by the tables in the GIS that link coordinates with environmental values.
- And vice-versa, for the *relation* that maps back **E** to **G**

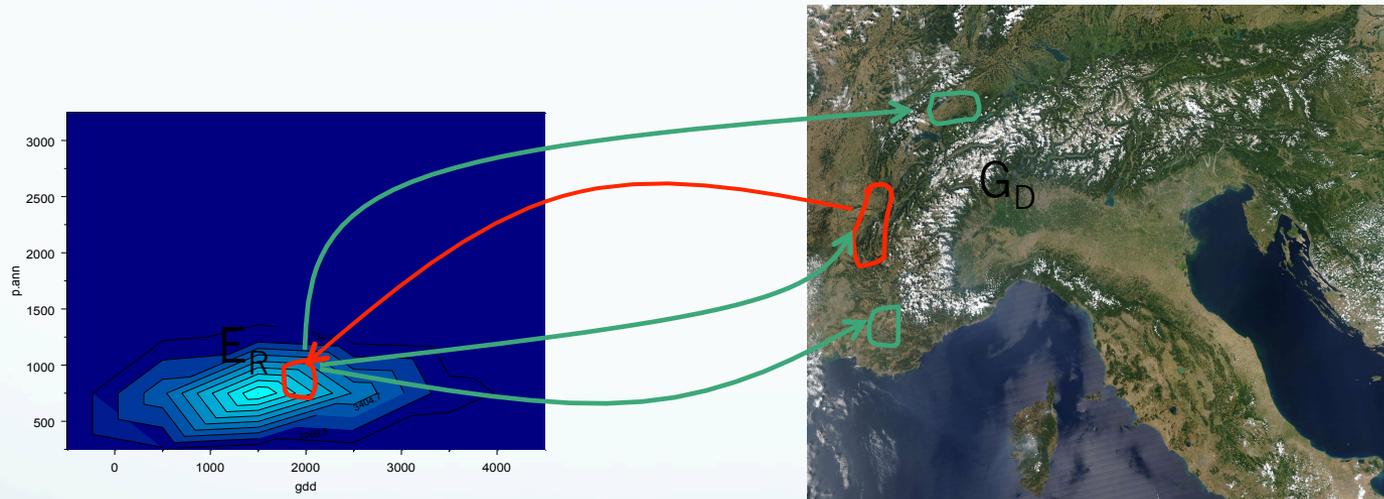


Ecological space (E),
 n dimensions

Lien entre niche et distribution

Hypothesis I: on connait la distribution géographique de l'espèce G_D

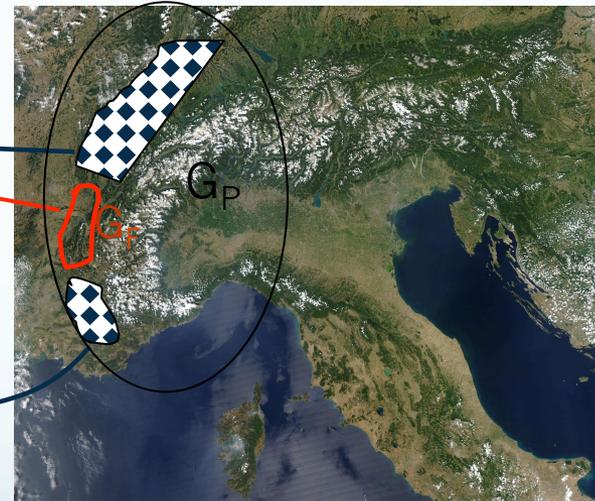
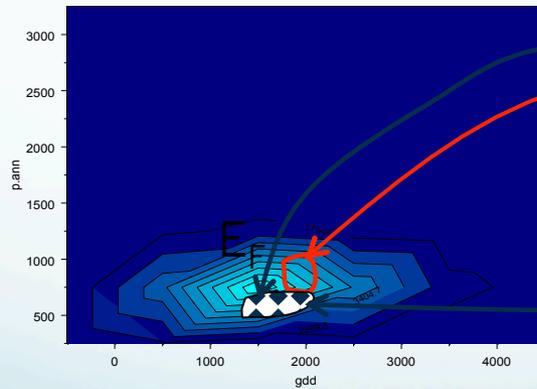
-> Niche réalisée, E_R



Lien entre niche et distribution

Hypothesis II: on connait toutes les régions où le taux de croissance de l'espèce est positif (on s'abstrait des interactions biotiques et de la dispersion) G_P

-> *Environmental subset* de la Niche fondamentale, E_F



Comment on mesure et représente la niche ?

1. On observe où l'espèce se trouve
2. On mesure les variables environnementales à ces endroits
3. On extrapole à partir de quelques "points" une aire "favorable" à partir d'algorithmes statistiques
4. On obtient les valeurs de ces variables environnementales
5. C'est la niche

=> Conceptuellement, chaque algo essaie de "retrouver" la niche réalisée (car les points d'occurrences viennent de la distribution de l'espèce G_D)

=> MAIS, lorsqu'on extrapole E vers G, en général, on ne retrouve pas que G_D car beaucoup d'autres régions ont les mêmes propriétés E.

Hypothèses clés des SDMs

- **Le postulat d'équilibre**

Les modèles de niche font l'hypothèse **que les espèces sont en équilibre avec leur environnement**, c'est-à-dire qu'elles sont présentes dans tous les habitats favorables et absentes de tous les habitats défavorables (Araújo and Pearson, 2005).

Hypothèses clefs des SDMs

- **Le postulat d'équilibre**

Les modèles de niche font l'hypothèse **que les espèces sont en équilibre avec leur environnement**, c'est-à-dire qu'elles sont présentes dans tous les habitats favorables et absentes de tous les habitats défavorables (Araújo and Pearson, 2005).

- **Et de bon échantillonnage :**

Les **données de présences représentent un bon échantillonnage de l'espace environnemental**.

l'échantillonnage des données de présence et des variables environnementales est souvent limité dans le temps et/ou l'espace, les modèles reflètent un instantané de la relation entre l'espèce et son environnement (Guisan and Thuiller, 2005).

Équilibre et échantillonnage

Partial range
filling



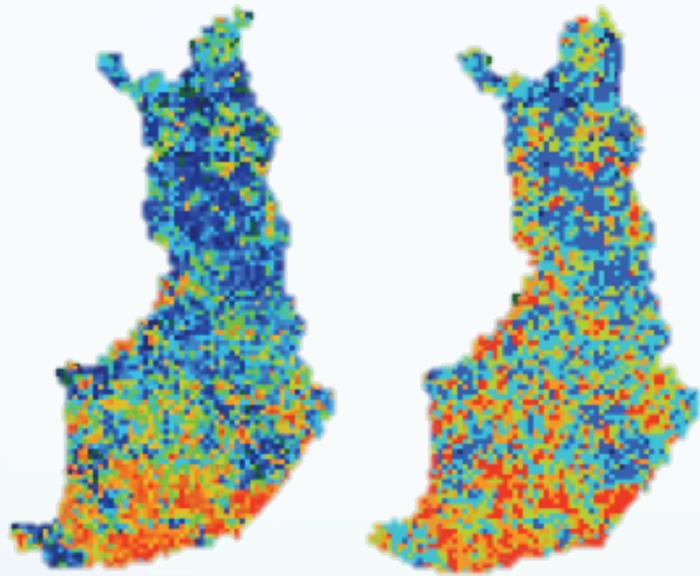
Complete
range filling



38,3% filling niche
->55 arbres européens

Équilibre et échantillonnage

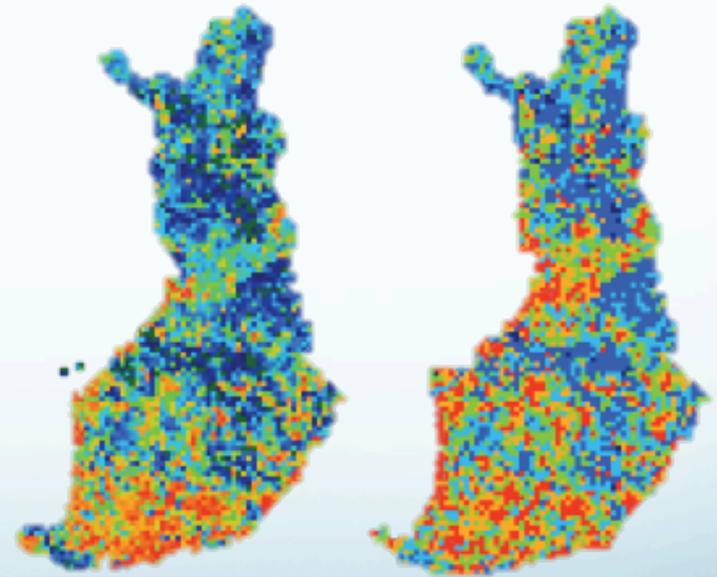
Bird species richness



Recording efforts

1970s

Bird species richness



Recording efforts

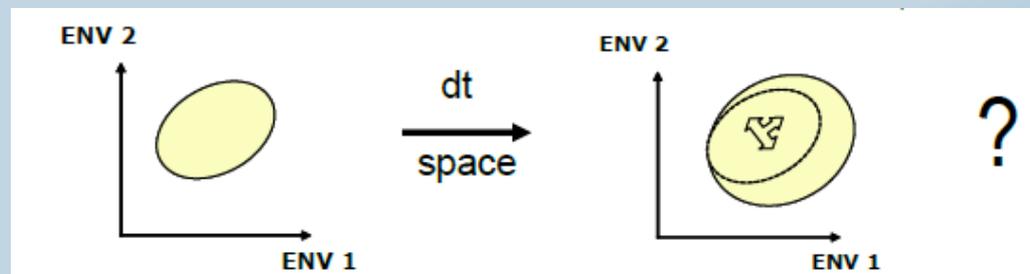
1990s

Synthèse : Équilibre et échantillonnage

- ▶ Deux facteurs déterminent le degré de pertinence d'un modèle de distribution
 - ▶ Le degré d'équilibre de l'espèce
 - ▶ % de la niche réalisée potentielle ?
 - ▶ La représentativité de l'échantillonnage
 - ▶ Etendue de l'espace environnemental représentée ?
 - ▶  pas de lien direct entre échantillonnage dans l'espace géographique et dans l'espace environnemental !
- ▶ Conclusion : bien analyser ses données avant toute modélisation !

En plus de l'équilibre, hypothèse du conservatisme de la niche

- **Niche conservatism** – la niche de l'espèce ne varie ni dans l'espace ni dans le temps



Species distribution models (SDMs) are used to predict the risk of invasion

In space: invasion

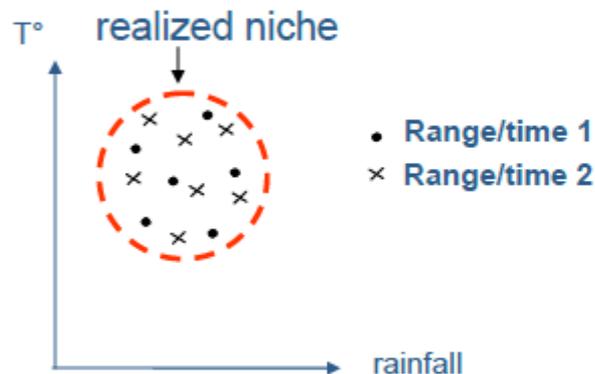
Worldwide risk of invasion by South-African Plants

In time: climate change

When predicting in space or time, the niche of a species is assumed to be conserved

En plus de l'équilibre, hypothèse du conservatisme de la niche

H_0 :
niche(1) = niche(2)

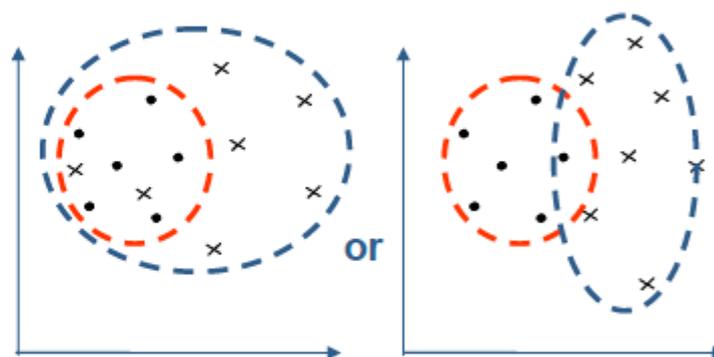


Implication for predictions:

models fitted in area/time 1 can be extrapolated to area/time 2 (e.g.

Peterson 2003 *Quart. Rev. Biol.*)

H_1 :
niche(1) \neq niche(2)



Implication for predictions:

models fitted in area/time 1 cannot predict the (full) potential distribution in area/time 2

Conservatisme de la niche

- Dans l'espace
 - Est ce que la niche réalisée d'une espèce envahissante est conservée durant son invasion ?
 - 2 causes
 - Ecological processes: absence de compétiteurs / prédateurs/ pathogènes dans l'aire envahie
 - **Modification de la niche réalisée**
 - Processus évolutifs: dérive génétique, sélection ...
 - **Modification de la niche fondamentale**

Conservatisme de la niche

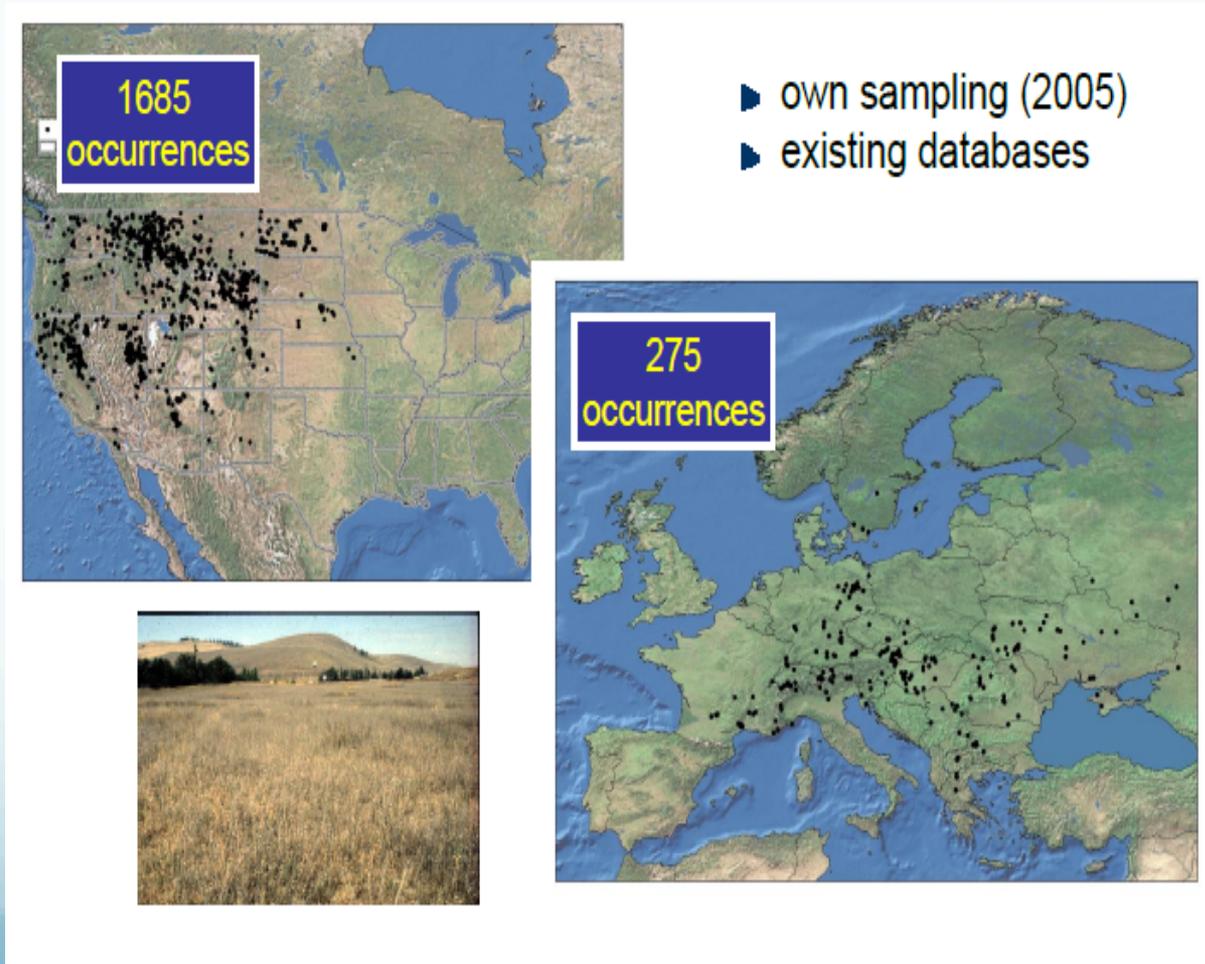
- Dans l'espace



- Introduite en ~1890 in the western part of the US from Europe
- Invaded over 3 millions hectares of cultivated areas
- Cost over \$150 millions a year

Conservatisme de la niche

- Dans l'espace

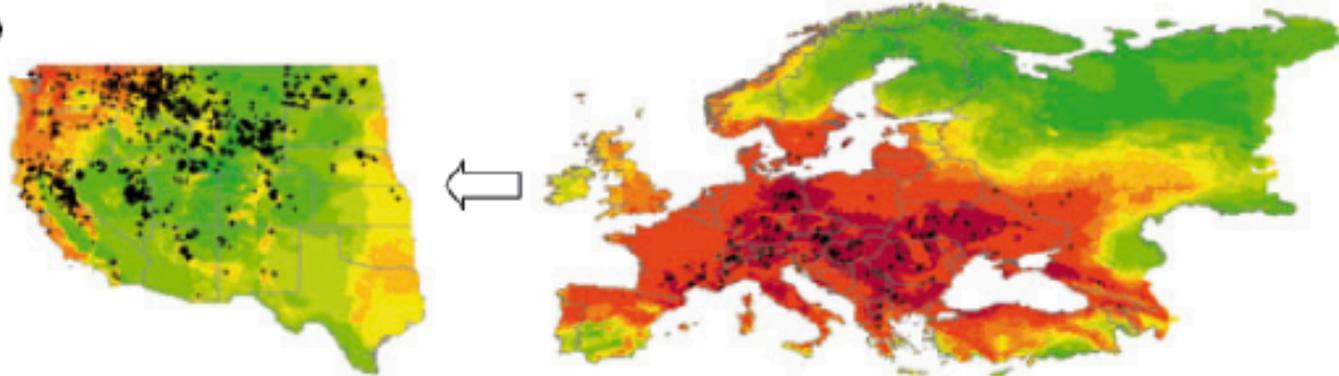


Conservatisme de la niche

- Dans l'espace

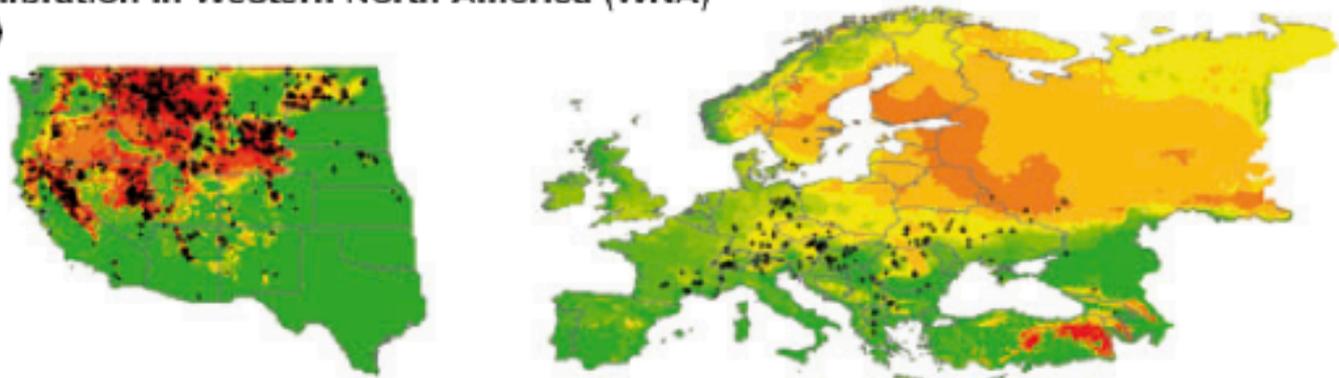
Calibration in Europe (EU)

(a)



Calibration in Western North America (WNA)

(c)



Occurrences of
Centaurea maculosa : +

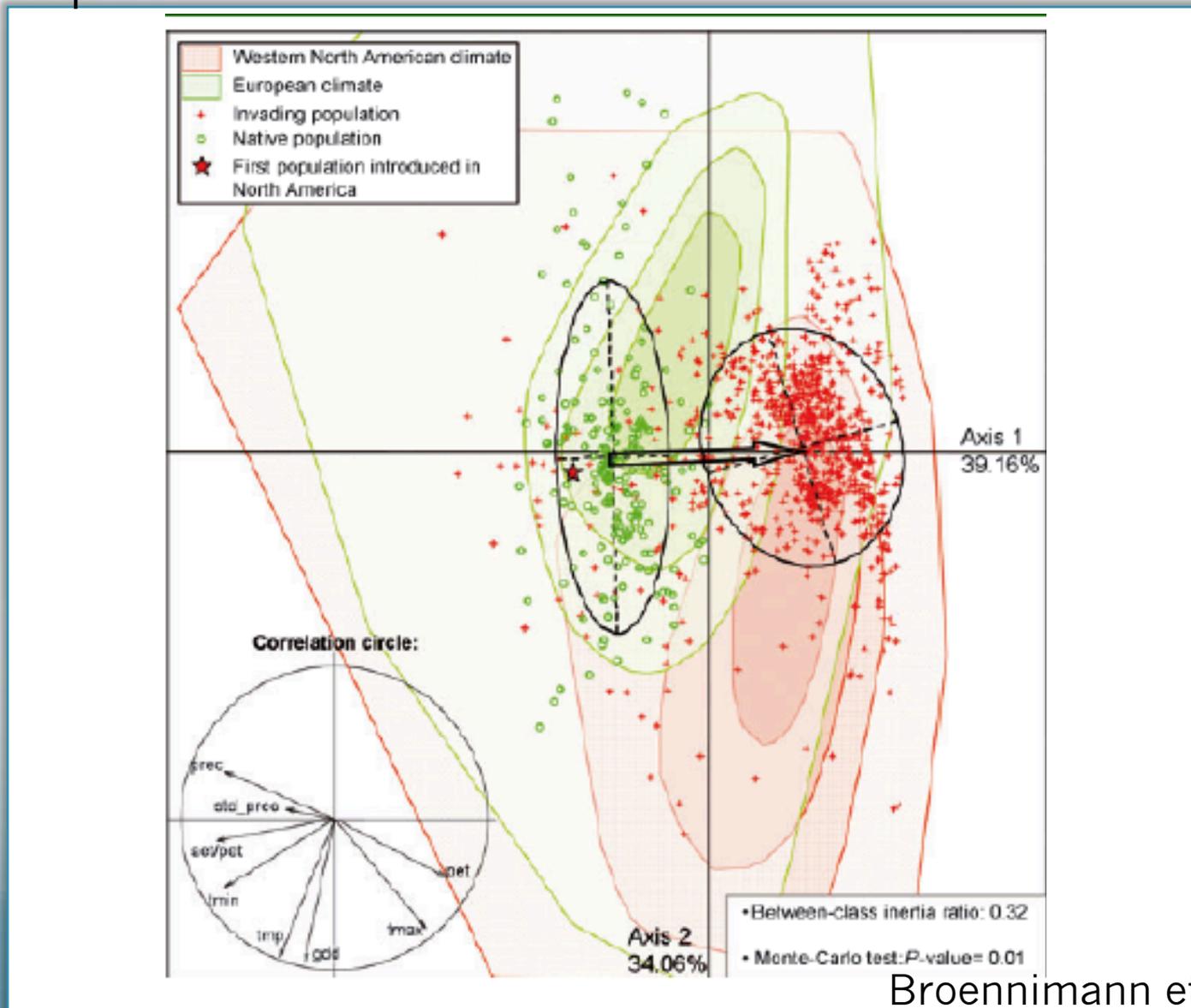
Predicted climatic
suitability :



Model
performance : 0 ←
Counter
prediction

Conservatisme de la niche

- Dans l'espace



Conclusion sur les hypothèses

Idéal

-modéliser la niche fondamentale de l'espèce

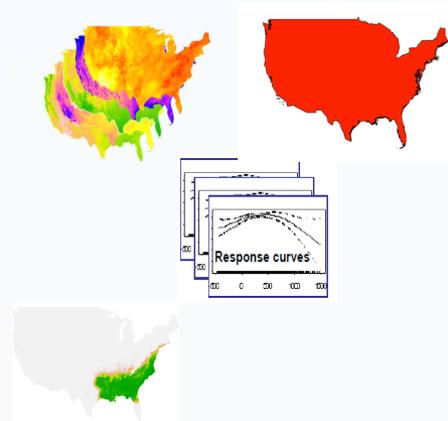
L'objectif n'est pas alors de décrire complètement l'hypervolume dans un espace multidimensionnel mais plutôt d'utiliser comme **approximation de la niche fondamentale (une combinaison de variables environnementales qui définissent la répartition des espèces)** (Tingley et al., 2009).

En reliant de manière empirique des données d'occurrences à des variables environnementales prédictives; on caractérise **essentiellement la niche réalisée de l'espèce** puisque les données d'occurrences sont contraintes par les interactions biotiques.

Enfin, pour approximer au mieux la niche fondamentale, il est recommandé de calibrer le modèle avec des données d'occurrences provenant d'une large étendue géographique, à travers l'ensemble de la répartition de l'espèce (Araújo and Guisan, 2006; Barbet-Massin et al., 2010; Phillips et al., 2006).

Construction des SDMs

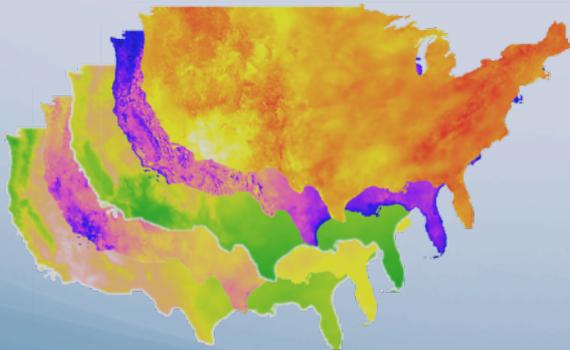
1. Conceptualisation
2. Préparation des données
3. Calibration des modèles
4. Evaluation des modèles
5. Prédiction spatiales
6. Prise en compte de l'incertitude



Principe de la modélisation d'aire de répartition

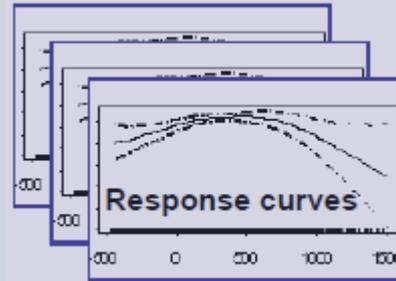
Data collection

présence /
absence

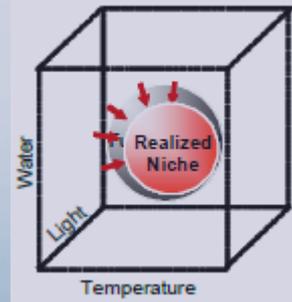


variables climatiques et
d'habitat

Statistical
modeling

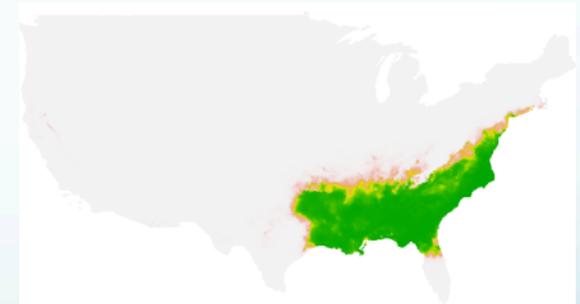


Fitting the niche



Spatial predictions

Distribution spatiale
actuelle modélisée



Modèle de
prédiction de
répartition

Construction des SDMs

1. Conceptualisation
2. Préparation des données
3. Calibration des modèles
4. Evaluation des modèles
5. Prédiction spatiales
6. Evaluation de l'applicabilité des modèles

Préparation des données : les données d'occurrences

Deux types de données utilisées

- Cartes ponctuelles (dot maps)
- Répartition estimée (outline maps)

Préparation des données : les données d'occurrences

Deux types de données utilisées

- Cartes ponctuelles (dot maps)

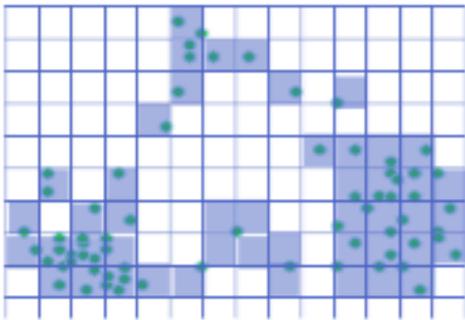


- Ponctuellement (coordonnées GPS)

Préparation des données : les données d'occurrences

Deux types de données utilisées

- Cartes ponctuelles (dot maps)



- Ponctuellement (coordonnées GPS)
- Par unités spatiales homogènes (grilles / rasters)
- Par unités spatiales hétérogènes (limites géopolitiques, unités écologiques)

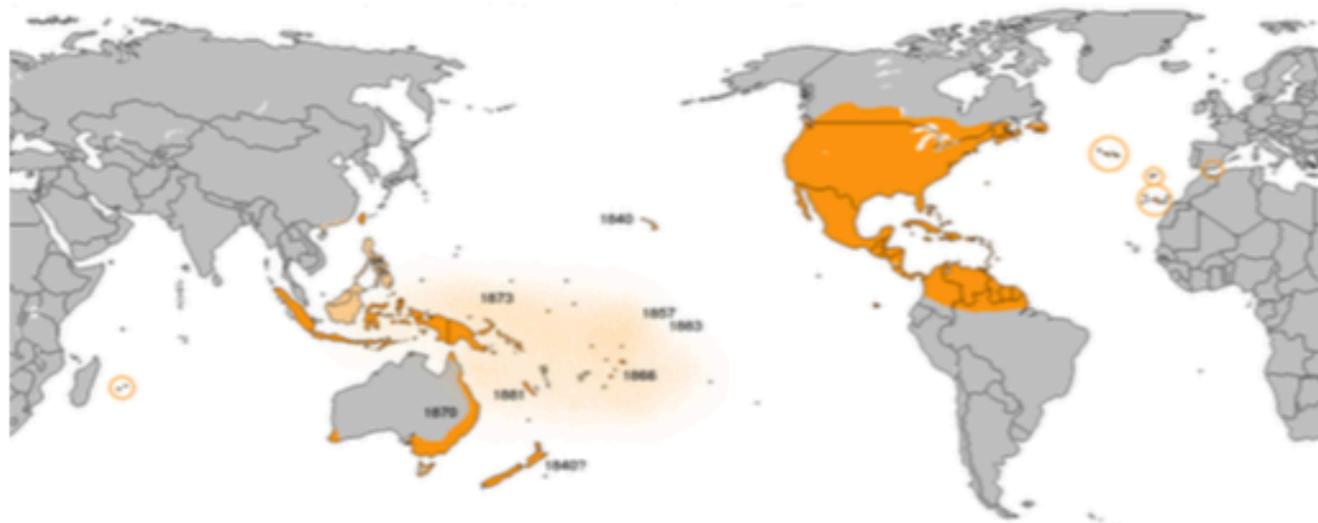
Préparation des données : les données d'occurrences

Deux types de données utilisées

- Cartes ponctuelles (dot maps)
- Répartition estimée (outline maps)



Répartition du monarque



Préparation des données : les données d'occurrences

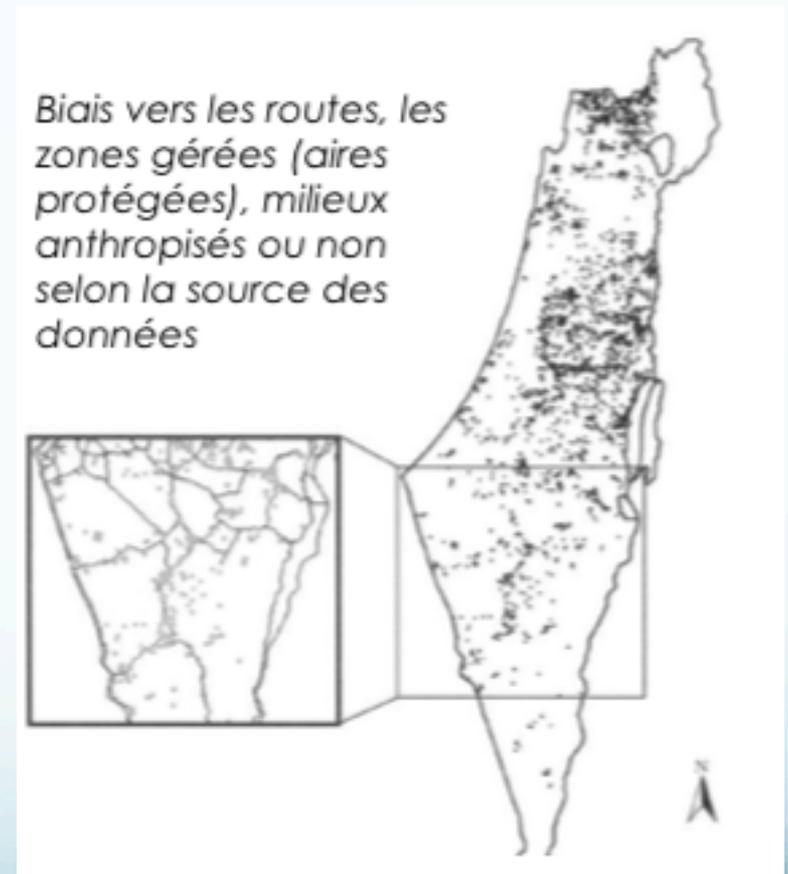
Deux types de données utilisées

- Cartes ponctuelles (dot maps)
- Répartition estimée (outline maps)
 - Limites présumées des aires
 - Basées sur les habitats connus, sur les capacités de dispersion estimées, etc.
 - Précision dépend de la connaissance de l'espèce et de l'expertise de l'auteur

Préparation des données : les données d'occurrences

Type : Présences (+absences ?)

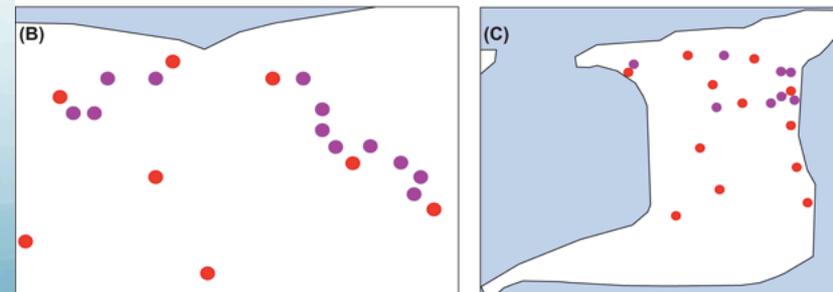
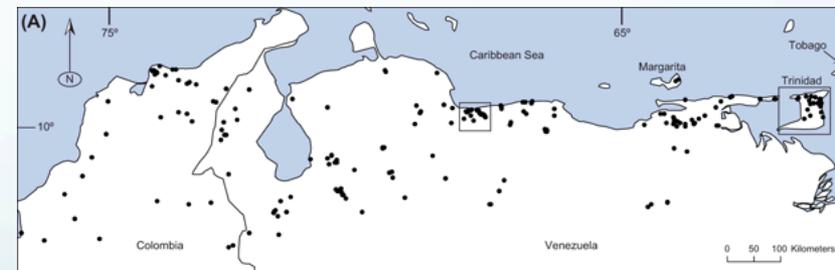
- Erreurs d'identification
- Approximations géographiques
- Etendue des données d'occurrence
- Biais d'échantillonnage
 - Biais spatial : variations dans la probabilité d'échantillonnage dans l'espace géographique (e.g., route, pays)
 - Biais environnemental



Préparation des données : solutions potentielles aux biais d'échantillonnage

- Quantifier l'effort d'échantillonnage et l'intégrer dans la calibration (Phillips *et al.* 2009, *Ecological Applications*)
- **Filtration géographique** : supprimer les doublons dans chaque localisation géographique.
 - 'Rasterisation' des occurrences
 - Package spThin (Aiello-Lammens *et al.* 2015, *Ecography*)

full spatial extent of unthinned records (black circles).



records removed by 'thin' (purple circles) and those retained by the function (red circles).

Préparation des données : solutions potentielles aux biais d'échantillonnage

- Quantifier l'effort d'échantillonnage et l'intégrer dans la calibration (Phillips *et al.* 2009, *Ecological Applications*)
- **Filtration géographique** : supprimer les doublons dans chaque localisation géographique.
 - 'Rasterisation' des occurrences
 - Package spThin
(Aiello-Lammens *et al.* 2015, *Ecography*)
- **Filtration environnementale** : supprimer les doublons dans chaque type d'habitat (e.g. Varela *et al.* 2014, *Ecography*)

Préparation des données :

Synthèse données d'occurrences

- Modèles ont besoin d'infos sur la présence ET l'ensemble des conditions environnementales possibles
 - Données de présence-absence difficiles à obtenir en quantité et qualité
- On utilise des données de présences seulement
- On crée des « pseudo-absences », points échantillonnés au hasard et fournis aux modèles comme « absences »

Methods in Ecology and Evolution

Selecting pseudo-absences for species distribution models: how, where and how many?

Morgane Barbet-Massin , Frédéric Jiguet, Cécile Hélène Albert, Wilfried Thuiller

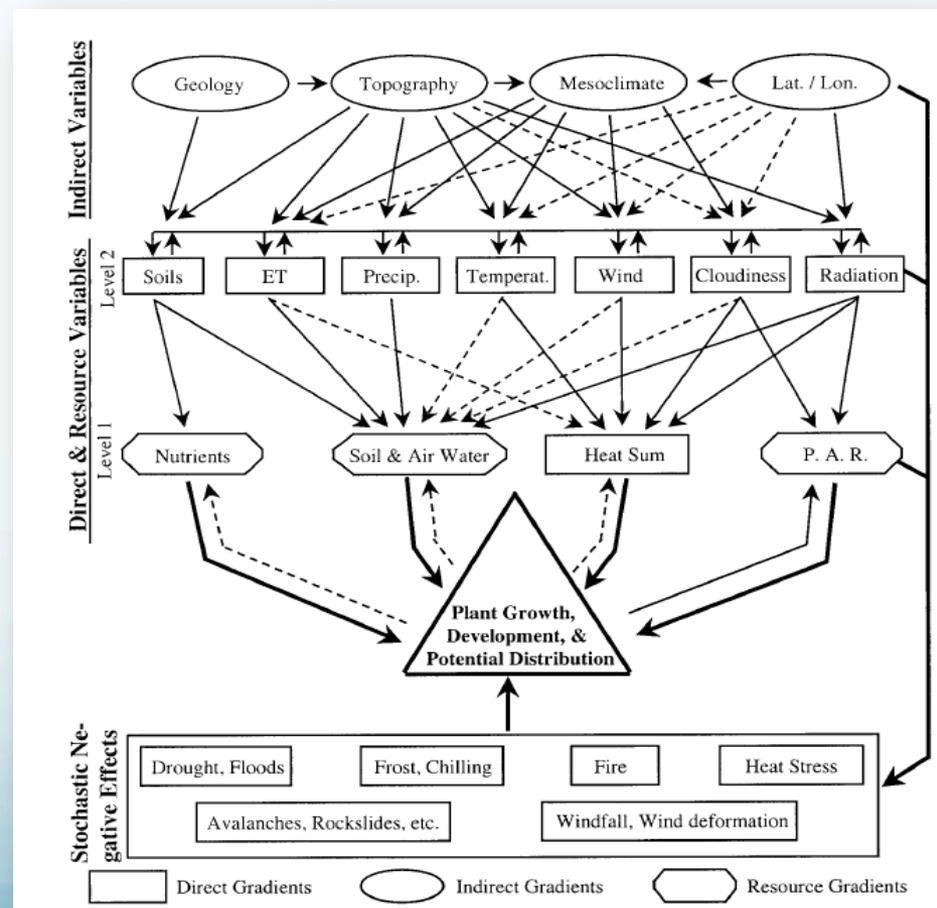
First published: 19 January 2012 [Full publication history](#)



View Issue TOC
Volume 3, Issue 2
April 2012
Pages 327-338

Préparation des données : les données environnementales

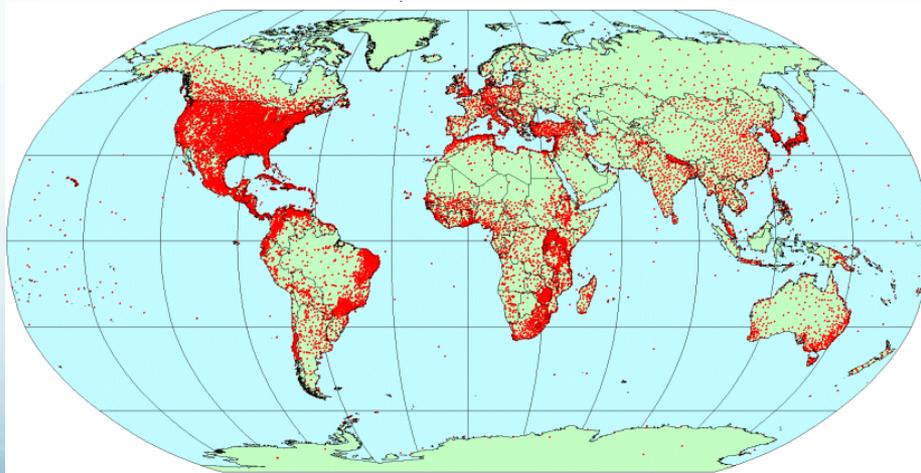
- Prédicteurs pertinents (physiologie, moyenne, extrême ?)
- Types de variables environnementales (effets directs, indirects , ...)



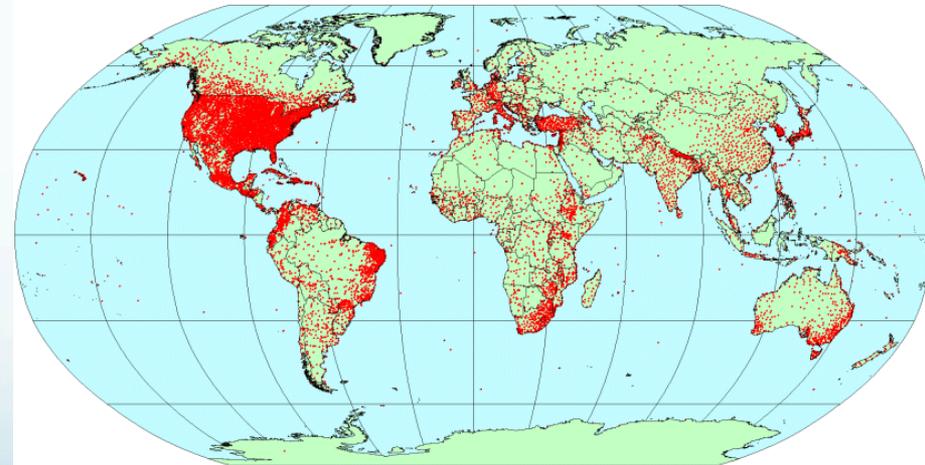
Préparation des données : les données environnementales

Climat

Past
Current
Future



Precipitation



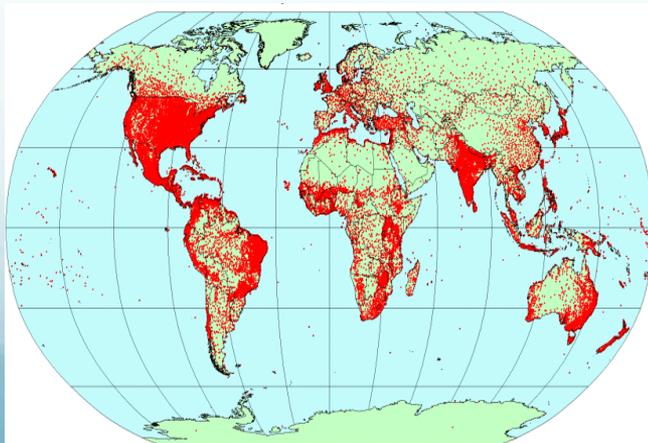
Temperature

Préparation des données : les données environnementales

www.worldclim.org

Milieus terrestres

Données interpolées sur la
base de stations météo



Research Article

 PDF
 Info

WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas

Stephen E. Fick  Robert J. Hijmans

First published: 15 May 2017 [Full publication history](#)

DOI: 10.1002/joc.5086 [View/save citation](#)

Cited by (CrossRef): 0 articles [Check for updates](#)
 Citation tools

 score 24

Funding Information

ABSTRACT

We created a new dataset of spatially interpolated monthly climate data for global land areas at a very high spatial resolution (approximately 1 km²). We included monthly temperature (minimum, maximum and average), precipitation, solar radiation, vapour pressure and wind speed, aggregated across a target temporal range of 1970–2000, using data from between 9000 and 60 000 weather stations. Weather station data were interpolated using thin-plate splines with covariates including elevation, distance to the coast and three satellite-derived covariates: maximum and minimum land surface temperature as well as cloud cover, obtained with the MODIS satellite platform. Interpolation was done for 23 regions of varying size depending on station density. Satellite data improved prediction accuracy for temperature variables 5–15% (0.07–0.17 °C), particularly for areas with a low station

Early View



[Browse Early View Articles](#)
Online Version of Record
published before inclusion
in an issue

Préparation des données : les données environnementales

Software note

PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales



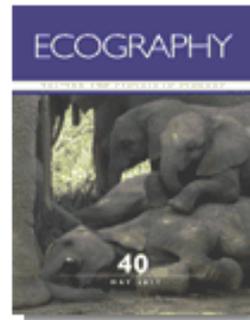
Damien A. Fordham¹, Frédéric Saltré¹,
Sean Haythorne¹, Tom M. L. Wigley¹,
Bette L. Otto-Bliesner², Ka Ching Chan³
and Barry W. Brook³

Version of Record online: 12 MAY 2017

DOI: 10.1111/ecog.03031

© 2017 The Authors

Issue



Ecography

**Early View (Online Version of
Record published before
inclusion in an issue)**

Am score 36

Additional Information (Show All)

[How to Cite](#) | [Author Information](#) | [Publication History](#)

SEARCH

In this issue

[Advanced >](#) [Saved Searches >](#)

ARTICLE TOOLS

- Get PDF (13429K)
- Save to My Profile
- E-mail Link to this Article
- Export Citation for this Article
- Get Citation Alerts
- Request Permissions

Préparation des données : les données environnementales

climex.knmi.nl

Milieus terrestres et
marins

Observations

Modèles climatiques

Séries temporelles

The screenshot shows the KNMI Climate Explorer website. The header includes the KNMI logo, the text "KNMI Climate Explorer", and navigation links for "Climate Explorer", "European Climate Assessment & Data", and "KNMI". A search bar is located in the top right corner. Below the header, there are several tabs: "Help", "News", "About", "Contact", "World weather", "Effects of El Niño", "Seasonal forecasts", and "Climate Change Atlas". The main content area is titled "KNMI Climate Change Atlas" and contains a form for selecting a region, season, dataset, and variable. The form is divided into several sections: "Select a region", "Select a season", "Select a dataset and variable", and "Map options". The "Select a region" section has radio buttons for "IPCC WG1", "IPBES", "countries", "place", and "box", with "World" selected in the dropdown. The "Select a season" section has dropdowns for "First month" (set to "Jan") and "length" (set to "12"). The "Select a dataset and variable" section has a dropdown for "Dataset" (set to "GCM: CMIP5 (IPCC AR5 Atlas subset)") and a dropdown for "Variable" (set to "near-surface temperature"). There are also radio buttons for "absolute" (selected) and "relative changes are shown", and radio buttons for "map" (selected) and "time series". The "Map options" section has a dropdown for "Scenario" (set to "Historical + RCP4.5"), a dropdown for "Measure" (set to "Difference of two periods"), and input fields for "Reference period" (1986-2005) and "Future period" (2081-2100). There is also a dropdown for "Mean/percentiles" (set to "mean"). A "Make map" button is at the bottom of the form, with a note: "May take up to 15 minutes the first time a season / measure is selected". To the right of the form, there is a "Further information" section with links to "Short introduction", "IPCC WG1 AR5 report, notably Annex 1 'Atlas'", "CMIP5 co-ordinated climate model experiments", and "RCP scenarios". Below that is a "Funding" section with links to "KNMI", "SPECS", "Red Cross / Red Crescent Climate Centre", and "Dutch Ministry of Infrastructure and Environment, DGM". At the bottom of the page, there is a "Done" section with a list of updates and bug fixes, including dates and descriptions of changes.

Préparation des données : les données environnementales

www.marspec.org

Milieux marins

Observations *in situ* et
satellite

Download data in GeoTiff or KMZ format (smaller downloads) :

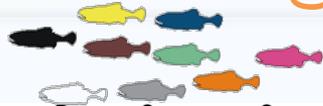
Layer Name	Definition (units)	Scaling Factor	Resolution				Google Earth*
			30 arcseconds	2.5 arcminutes	5 arcminutes	10 arcminutes	
bathymetry	Depth of the seafloor (m)	1x	~1km	~5km	~10km	~20km	KMZ
biogeo01	East/West Aspect (radians)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo02	North/South Aspect (radians)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo03	Plan Curvature	10,000x	~1km	~5km	~10km	~20km	KMZ
biogeo04	Profile Curvature	10,000x	~1km	~5km	~10km	~20km	KMZ
biogeo05	Distance to Shore (km)	1x	~1km	~5km	~10km	~20km	KMZ
biogeo06	Bathymetric Slope (degrees)	10x	~1km	~5km	~10km	~20km	KMZ
biogeo07	Concavity (degrees)	1,000x	~1km	~5km	~10km	~20km	KMZ
biogeo08	Mean Annual SSS (psu)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo09	SSS of the freshest month (psu)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo10	SSS of the saltiest month (psu)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo11	Annual range in SSS (psu)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo12	Annual variance in SSS (psu)	10,000x	~1km	~5km	~10km	~20km	KMZ
biogeo13	Mean Annual SST (°C)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo14	SST of the coldest month (°C)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo15	SST of the warmest month (°C)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo16	Annual range in SST (°C)	100x	~1km	~5km	~10km	~20km	KMZ
biogeo17	Annual variance in SST (°C)	10,000x	~1km	~5km	~10km	~20km	KMZ

SSS = Sea Surface Salinity

SST = Sea Surface Temperature

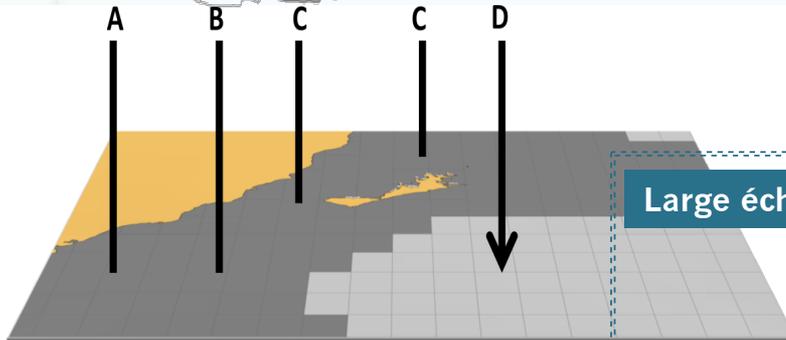
*KMZ files generated from 5 arcminute data - for visualization purposes only

Préparation des données : comment gérer les pbs d'échelles ?

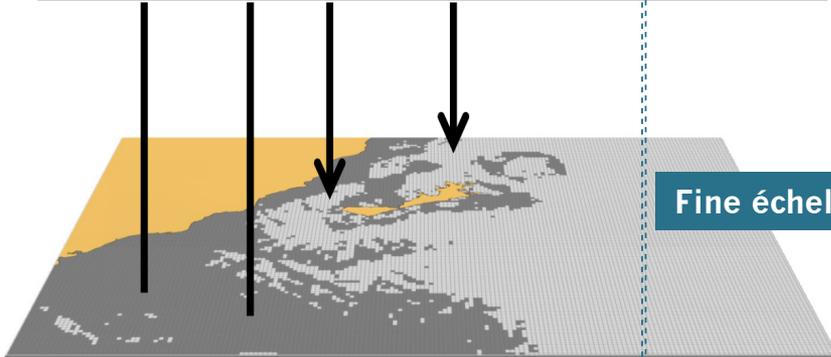


Echelle d'analyse : **possibilité d'adopter une approche multi-échelles**

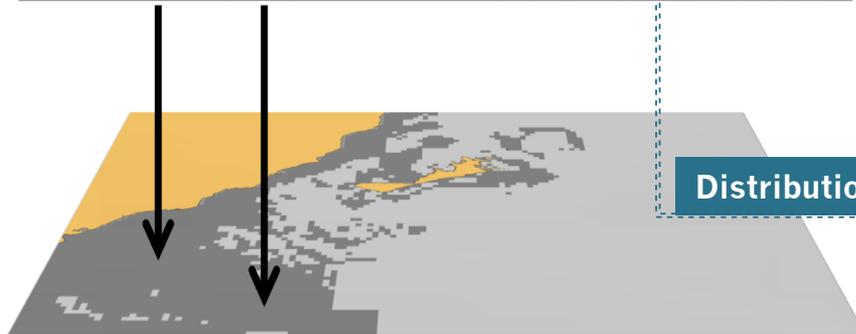
Decreasing spatial resolution



Large échelle : Modélisation des enveloppes climatiques



Fine échelle : Modélisation des habitats

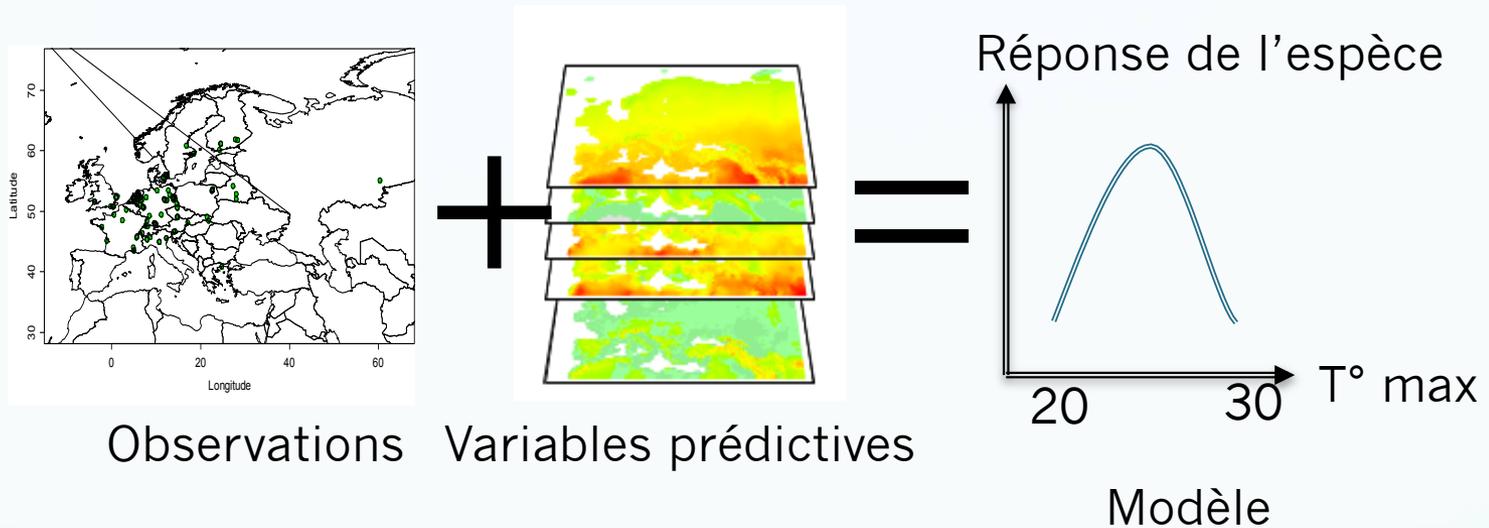


Distribution spatiale potentielle

Construction des SDMs

1. Conceptualisation
2. Préparation des données
3. Calibration des modèles
4. Evaluation des modèles
5. Prédiction spatiales
6. Evaluation de l'applicabilité des modèles

Calibration : principe général



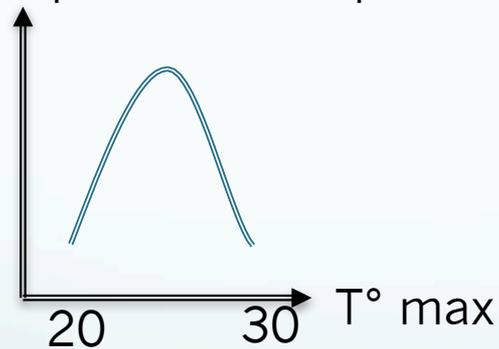
X	Y
1.1454	23.454
65	86
2.5631	24.551
56	45

T° max
24.5
23.5

Calibration : principe général

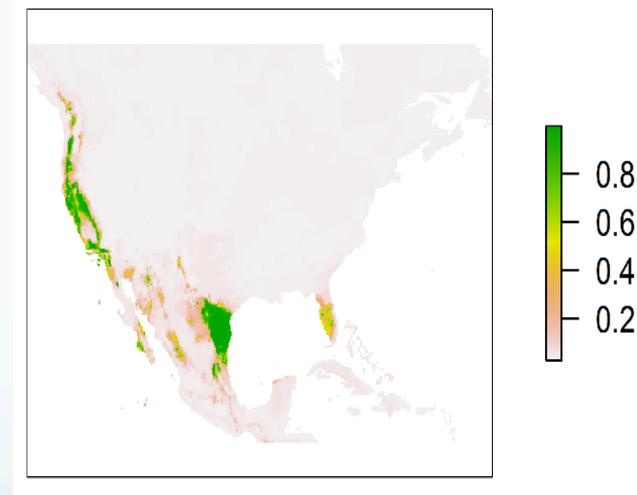
Réponse aux variables environnementales :
probabilité de présence

Réponse de l'espèce



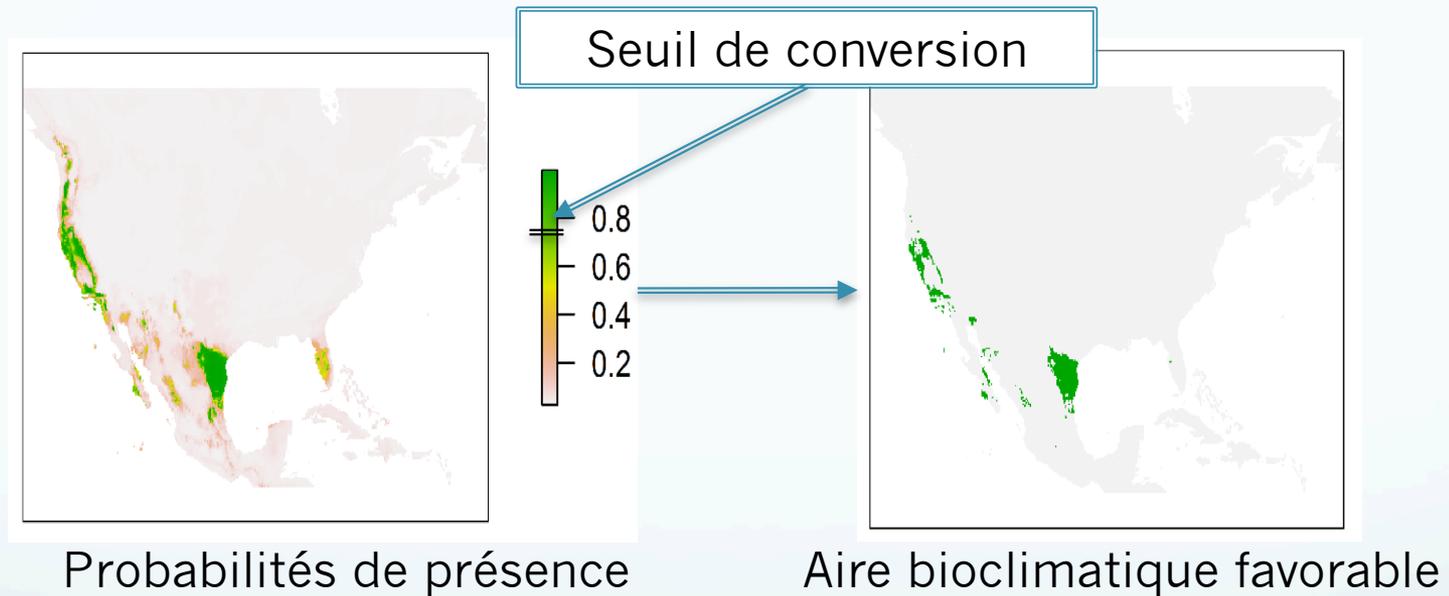
Modèle

Projection
dans l'espace
géographique



Probabilités de présence

Calibration : principe général



Calibration : principe général

“All models are wrong; the question is how wrong do they have to be to not be useful”

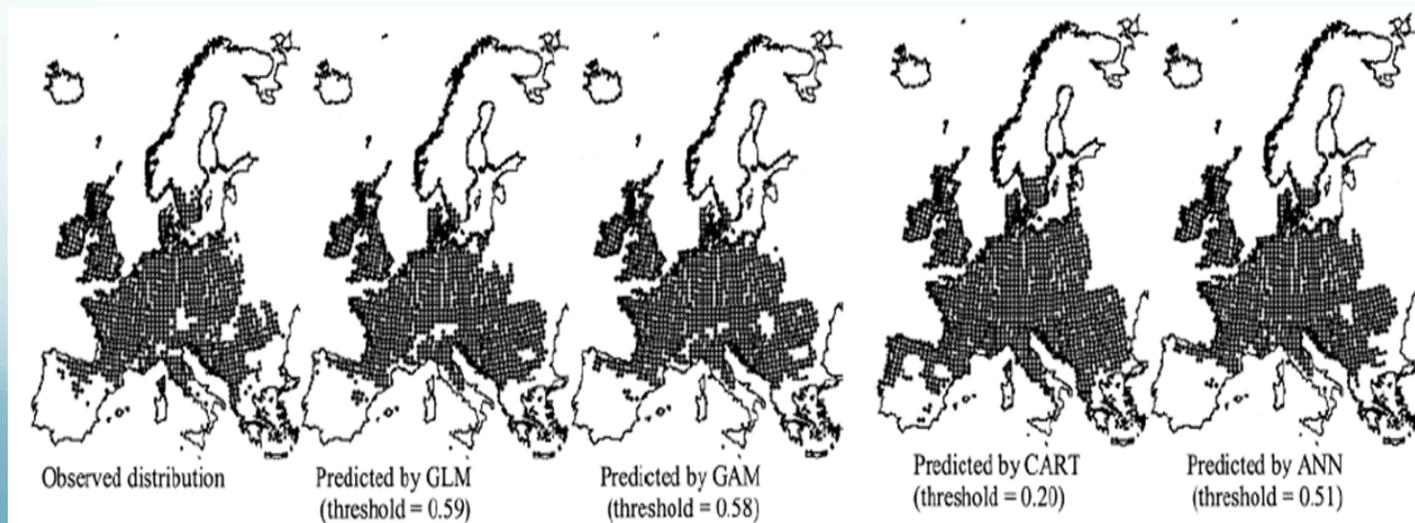
G.E.P. Box

(Savoir à quel point ils sont faux permettra de déterminer s'ils sont utiles)

A quel point les SDMs sont faux ?

Calibration : sources de variations

- Conditions initiales (données présences et pseudo-absences)
 - Plusieurs types de modèles...
 - Modèles de type « regression » e.g. GLM, GAM
 - Modèles de type « classification » e.g. CTA
 - Modèles de type « apprentissage automatique (e.g. RF, ANN, BRT)
- ... donc variabilité des résultats



Calibration : sources de variations

- Conditions initiales (données)
- Plusieurs types de modèles...
 - Modèles de type « regression » (e.g. GLM, GAM)
 - Modèles de type « classification » (e.g. CTA)
 - Modèles de type « apprentissage automatique » (e.g. RF, ANN, BRT)
- ... donc variabilité des résultats
- Protocoles de modélisation (paramétrisations, répétitions de simulations de pseudo-absence, répétitions pour l'évaluation)

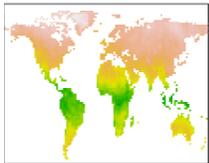
Calibration : sources de variations

100 espèces X 5 PAs X 7 Algorithms X 4 Répétitions

Calibration (70%)

Presence + Pseudo Absences (N=1000-10000)

x5

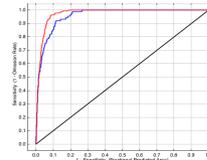


Climate variables (6)
Land use variables (12)

Evaluation (30%)

7 algorithms:
SRE, GBM,
GLM, MARS, RF, FDA,
MaXent
x4

Evaluation of model
performances
TSS, AUC and
Boyce index



Projection (100 %)

Future climate and land use
models and 2 CO₂ emission
scenarios

CGCM, CSIRO, HADCM3 et A1 B2



Consensus model based on
the 5 best models weighted
by their performance en
2050
- 2100

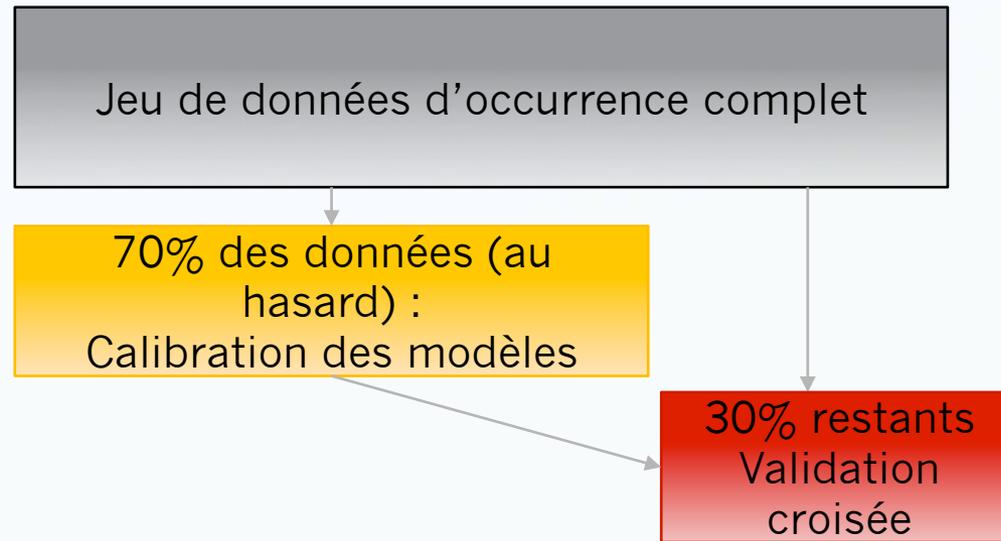
Construction des SDMs

1. Conceptualisation
2. Préparation des données
3. Calibration des modèles
4. Evaluation des modèles
5. Prédiction spatiales
6. Analyse de l'incertitude

Evaluation

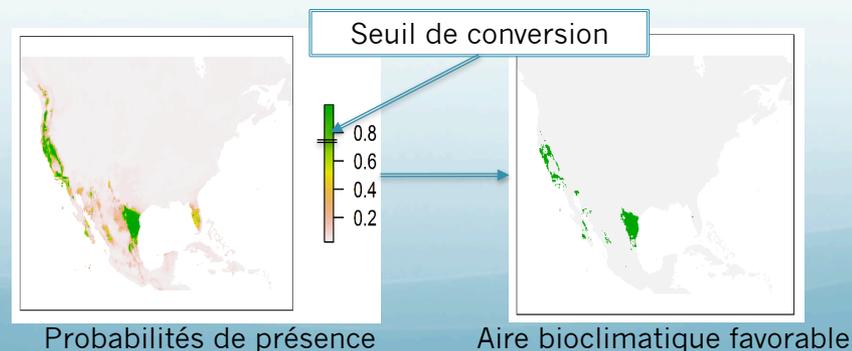
- Calibration et évaluation sur le même jeu de données [**A éviter**]
- Validation croisée [**Approche générale**]
- Validation indépendante [**Meilleure approche – nécessite un jeu de données indépendant**]

Evaluation : validation croisée

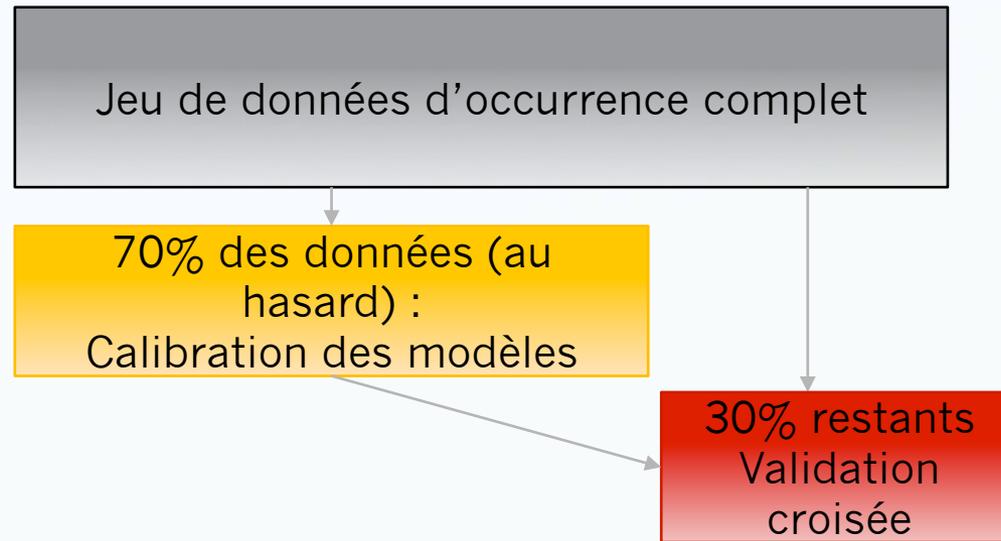


Qu'est ce qu'on teste ?

la capacité des modèles calibrés à reproduire les présences et les pseudo absences (du jeu de données restant)



Evaluation : validation croisée



Qu'est ce qu'on teste ?

la capacité des modèles calibrés à reproduire les présences et les pseudo absences (du jeu de données restant)

- Métriques
 - « seuil-dépendantes » (e.g. TSS, Jaccard) qui vont convertir des variables continues en présence / absence
 - vs.
 - Métriques « seuil-indépendantes » (e.g. Boyce index)

Evaluation : métrique d'évaluation

Métriques seuil-dépendantes :

Observations (CV)	Probabilité prédite
1	0.8
0	0.1
0	0.4
1	0.5

Jeu d'évaluation

Evaluation : métrique d'évaluation

Métriques seuil-dépendantes :

Observations (CV)	Probabilité prédite	Seuil 0	Seuil 0.1	Seuil 0.2
1	0.8	1	1	1
0	0.1	1	1	0
0	0.4	1	1	1
1	0.5	1	1	1

Jeu d'évaluation

Evaluation : métrique d'évaluation

Métriques seuil-dépendantes :

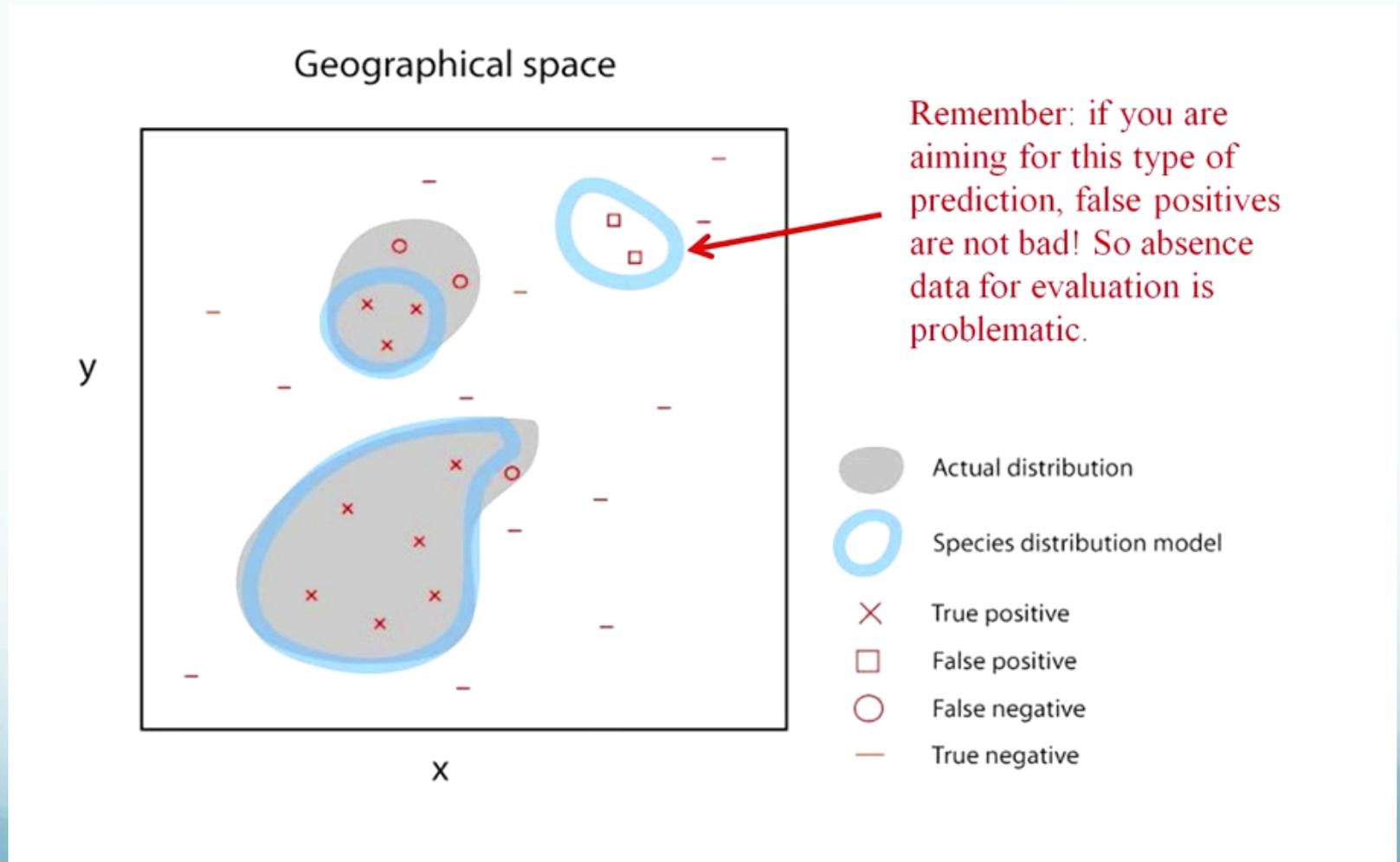
Observations (CV)	Probabilité prédite	Seuil 0	Seuil 0.1	Seuil 0.2	Seuil 0.45	Seuil 1
1	0.8	1	1	1	1	0
0	0.1	1	1	0	0	0
0	0.4	1	1	1	0	0
1	0.5	1	1	1	1	0

Jeu d'évaluation

Matrice de confusion

	Données de présences	Données d'absences (Pseudo-absences)
Présences prédites	Vrais Positifs (TP)	Faux Positifs (FP)
Absences prédites	Faux Négatifs (FN)	Vrais Négatifs (TN)

Evaluation : métrique d'évaluation



Evaluation : métrique d'évaluation

Matrice de confusion

	Données de présences	Données d'absences
Présences prédites	Vrais Positifs (TP)	Faux Positifs (FP)
Absences prédites	Faux Négatifs (FN)	Vrais Négatifs (TN)

Sensitivity = $TP / (TP + FN)$ (**proportion of observed presences correctly predicted**)

Specificity = $TN / (TN + FP)$ **proportion of observed absences correctly predicted**

True Skill Statistic = Sensitivity + Specificity - 1

Recommandé par Allouche *et al.* (2006) comme métrique idéale

Synthèse : Métrique d'évaluation

- Boyce pour présences seules
- Jaccard (indices de similarité) ou TSS pour convertir la probabilité de présence en présence-absence

MAIS Somodi et al., 2017

- Always subjective (TSS >0.6, AUC >0.8)

D'une manière générale, validation croisée rarement indépendante

Mauvaise évaluation = calibration vraiment mauvaise

- Bonne évaluation **n'implique pas forcément** bonne calibration

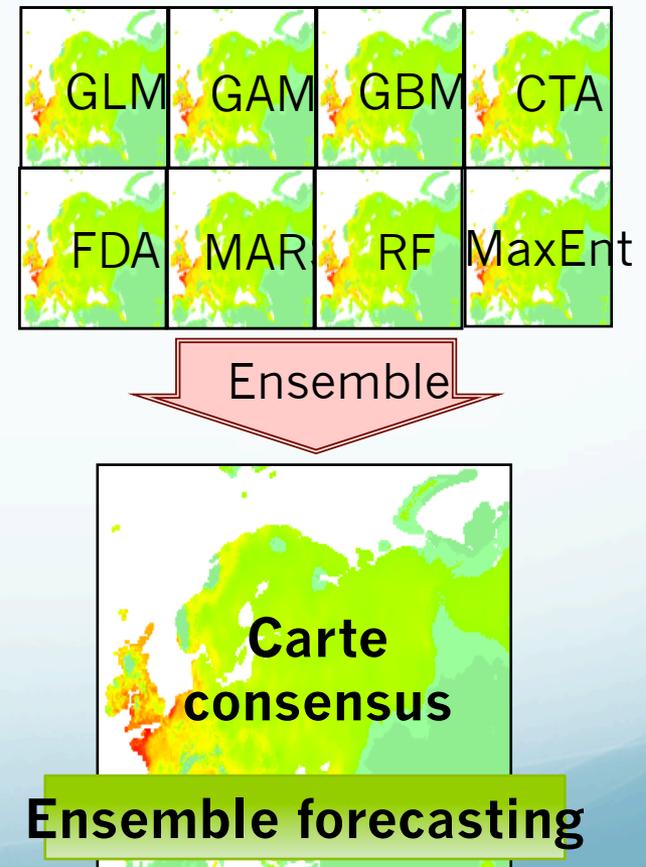
Construction des SDMs

1. Conceptualisation
2. Préparation des données
3. Calibration des modèles
4. Evaluation des modèles
5. Prédiction spatiales
6. Analyses de l'incertitude

Prédictions spatiales

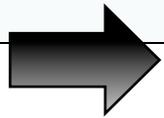
Projection des modèles d'ensemble dans le temps et l'espace :

- Cartes de prédiction spatiale (probabilité de présence, présence-absence)

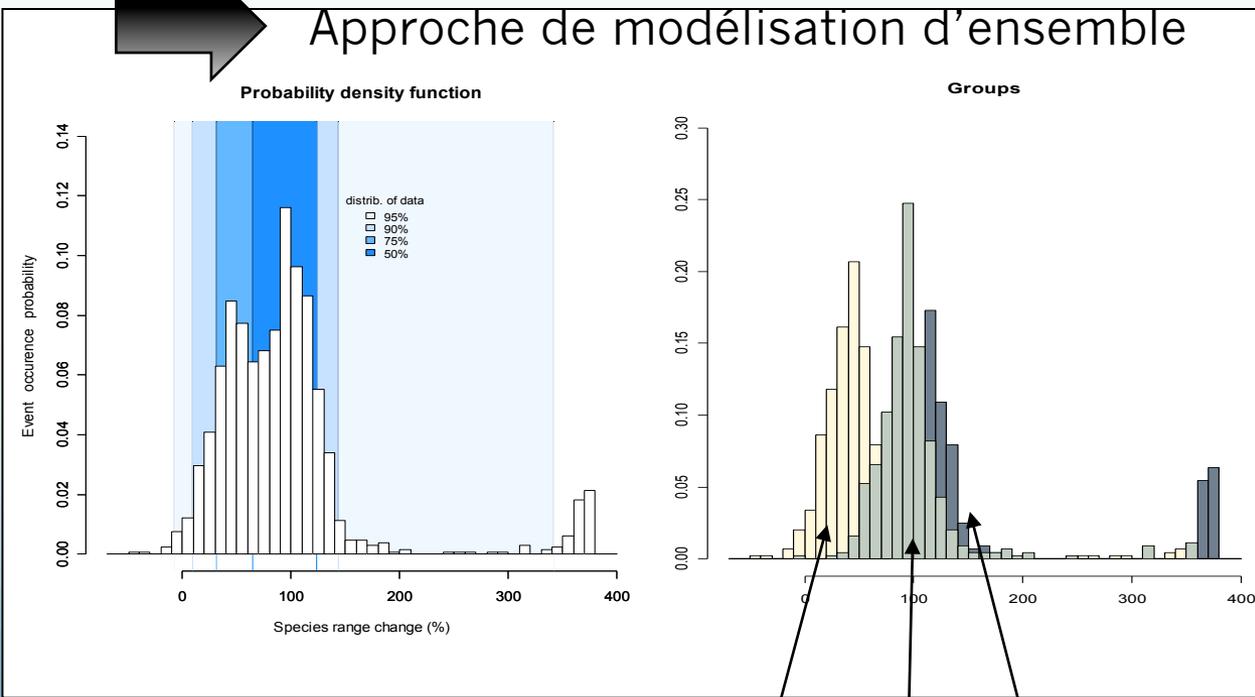


Prédictions spatiales

- Calibrations différentes, parfois divergentes
- Evaluations pas toujours fiables



Approche de modélisation d'ensemble



- Importance:
1. Estimation moyenne = 'meilleure' estimation
 2. **Evaluation de l'incertitude :**
 - Divergence = prédictions non fiables
 - Convergence = prédictions fiablesPossibilité de **soustraire l'incertitude pour ne regarder que les prédictions fiables** (e.g. conservation)

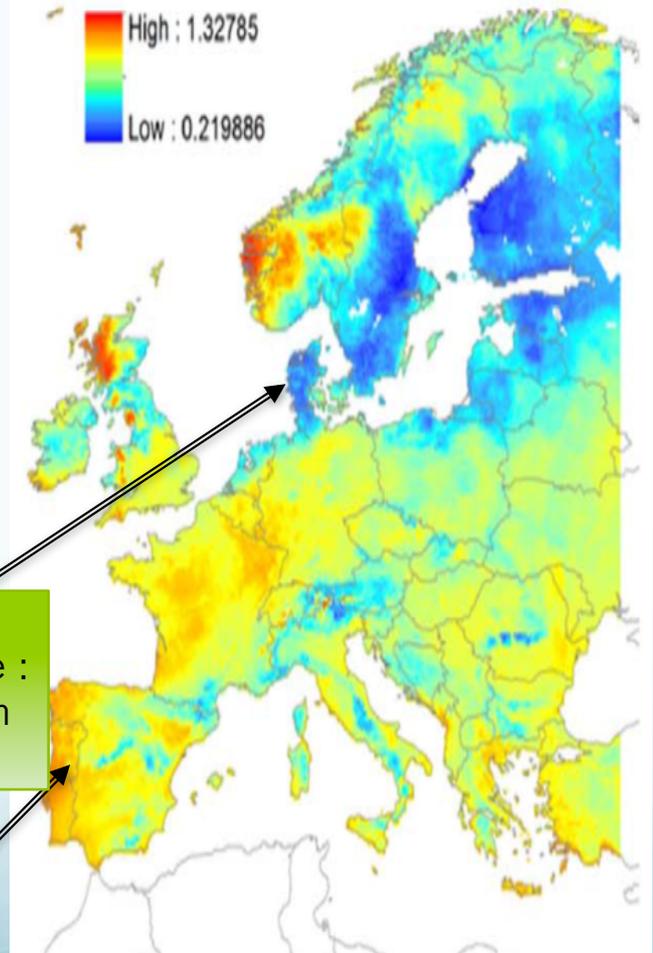
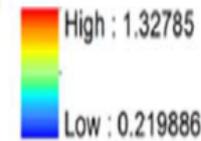
A1
A2
B1

Prédictions spatiales

Projection des modèles d'ensemble dans le temps et l'espace :

- Cartes de prédiction spatiale (probabilité de présence, présence-absence)
- Cartes d'incertitudes (où les modèles sont fiables et où ils ne le sont pas)

Uncertainty from bioclimatic models



Forte
incertitude :
zones non
fiables

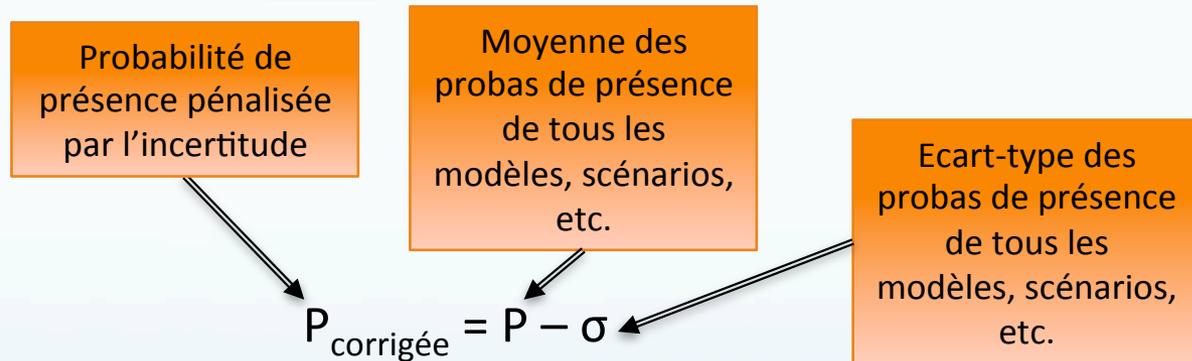
Faible
incertitude :
zones fiables

Ensemble forecasting

Prédiction spatiale prise en compte de l'incertitude

Exemple : recherche de zones **maximisant les chances de survie de l'espèce**

Identification des zones où tous les modèles prédisent le même résultat

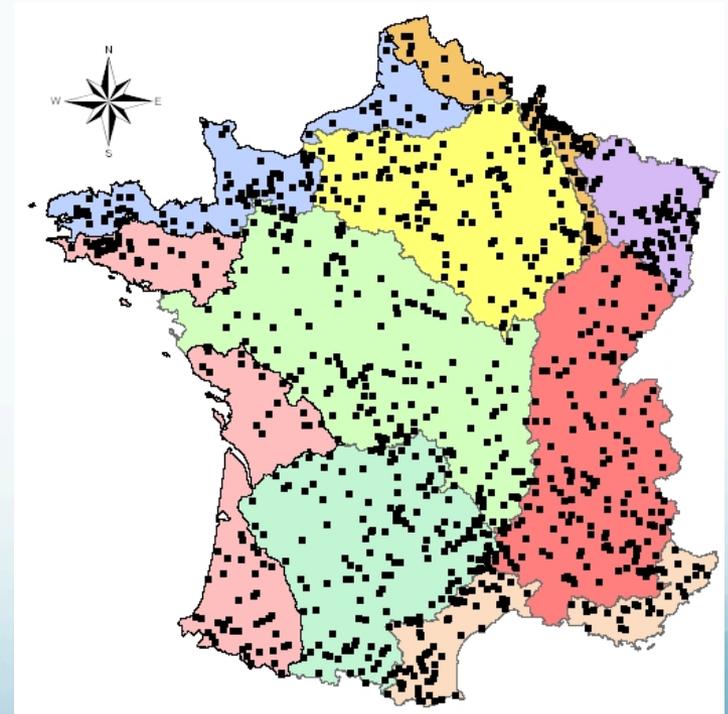


$P = 0$: faible probabilité de présence OU pas de consensus entre les modèles

$P > 0$: probabilité de présence supérieure à zéro, consensus entre les modèles

Prédiction spatiale prise en compte de l'incertitude

- Test with fishes
- 35 fish species in France
- 7 SDMs, 4 global emission scenarios (GES), 3 global climatic models (GCM) and 100 repeated sampling (Data)
- Standard variance partitioning methods

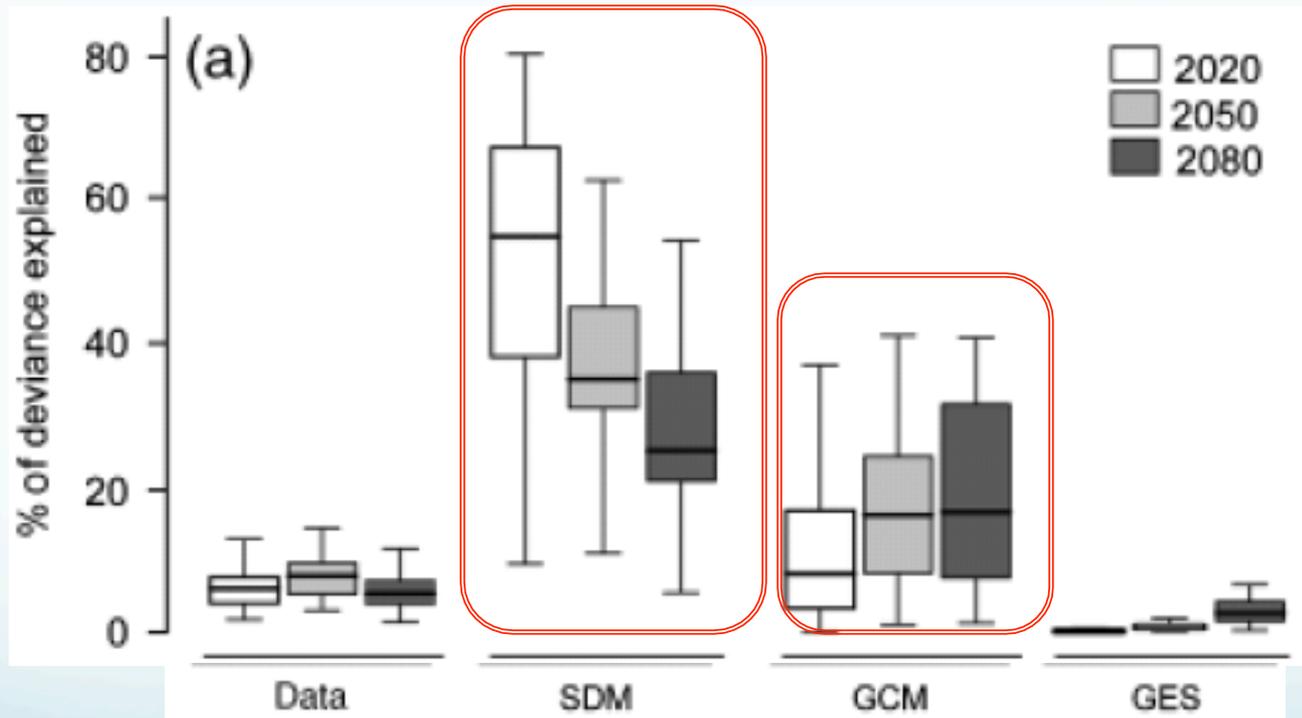


Global Change Biology (2010) 16, 1145–1157, doi: 10.1111/j.1365-2486.2009.02000.x

Uncertainty in ensemble forecasting of species distribution

LAËTITIA BUISSON*†, WILFRIED THUILLER‡, NICOLAS CASAJUS§, SOVAN LEK*
GAËL GRENOUILLET*

Prédiction spatiale prise en compte de l'incertitude



Species range change

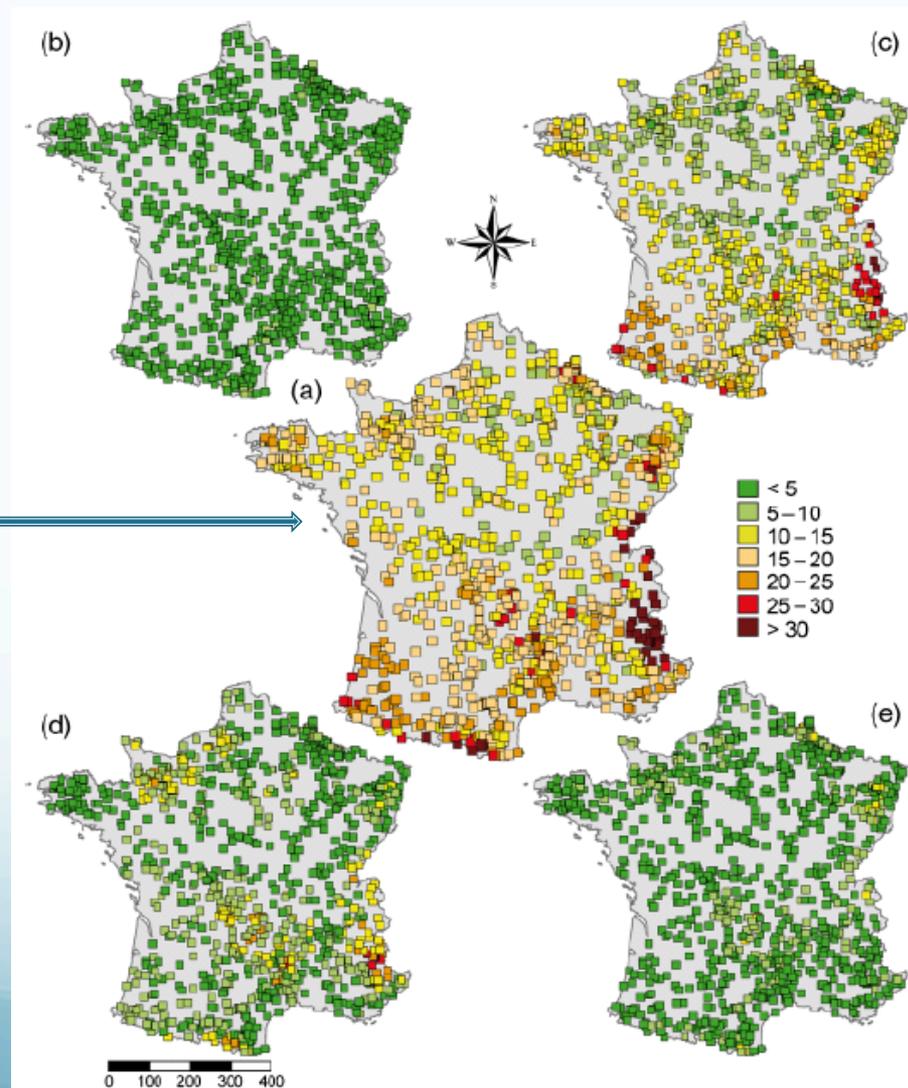
Prédiction spatiale prise en compte de l'incertitude

Sd Data

Sd SDM

Sd TOTAL

Sd GES



Conclusion

- Approches complexes et difficiles à maîtriser
- Attention à la préparation des données
- Ne jamais se fier à une prédiction issue d'un seul modèle, sans incertitude
- Se baser sur la tendance moyenne de plusieurs modèles (ensemble modelling / ensemble forecasting)

Please try to think to all these issues when conceptualizing your analysis, when analysing and interpreting your results and writing your papers

Merci à Boris Leroy, Wilfried Thuiller et Morgane Barbet Massin pour leurs contributions à ce cours