

# Identification spatiale continue des points chauds de diversité beta à l'aide de modèles de distribution d'espèces

Gabriel Dansereau

September 12, 2019

# Plan de la présentation

- ▶ Objectif
- ▶ Description eBird
  - ▶ Exemple données
  - ▶ Checklists
- ▶ Données brutes
  - ▶ Méthode présence-absence
  - ▶ Résultats
- ▶ Modèles de distribution d'espèces (SDM)
  - ▶ Méthode BIOCLIM
  - ▶ Données BIOCLIM
- ▶ À venir
- ▶ Autres points

# Objectif

- ▶ Objectif général: Identification des zones qui contribuent le plus à la diversité beta dans l'espace
  - ▶ Échelle continue, étendue
  - ▶ Diversité beta: composition spécifique, interactions
  - ▶ Prévission changements climatiques
  - ▶ Données science ouverte science citoyenne
- ▶ 1ère partie: Test de la méthode
  - ▶ LCBD sur données brutes
  - ▶ Développement méthode SDM
  - ▶ LCBD sur sorties SDM
  - ▶ Scénarios changements climatiques IPCC
- ▶ 2e partie: Applications aux interactions
- ▶ Pour l'instant: LCBD sur données brutes sorties SDM, SDM de base

# Description eBird

- ▶ Parulines - Famille Parulidae
- ▶ CA, US, MX
- ▶ 23 000 000 observations (uniques)
- ▶ 63 espèces: quartiles = 2, 11 000, 140 000, 370 000, 4 000 000
- ▶ Années: quartiles = 1838, 2012, 2015, 2018, 2019

# Description eBird

- Variables base : espèce, latitude, longitude, date, checklist complète, compte

height	species	latitude	longitude	observationDate	allSpeciesReported	observationCount
1	Setophaga_townsendi	35.32	-120.84	2007-03-17	1	1
2	Geothlypis_trichas	35.32	-120.84	1994-11-06	1	4
3	Oreothlypis_celata	58.35	-134.59	2008-07-22	1	2
4	Setophaga_coronata	58.35	-134.59	2008-07-22	1	1
5	Setophaga_coronata	49.10	-123.17	1979-10-16	1	8
6	Setophaga_coronata	37.90	-75.37	2004-11-23	1	82

- Données échantillonnage: type échantillonnage, durée, distance, nombre d'observateurs

height	protocolType	durationMinutes	effortDistanceKm	numberObservers
1	Area	60		15
2	Historical	30		1
3	Traveling	50	2.09	1
4	Traveling	50	2.09	1
5	Incidental			1
6	Traveling	105	2.58	5

# Données brutes - Méthodes

## Presence-absence par espèce

- ▶ Pixels - Résolution 5 arc-minutes
- ▶ Matrice  $661 \times 1141$  (env. 750 000 pixels)
- ▶ 100 000 sites avec observations sur 400 000 sites

## Matrice Y sites x espèces

- ▶  $100\,000 \times 63$
- ▶ Transformation Hellinger

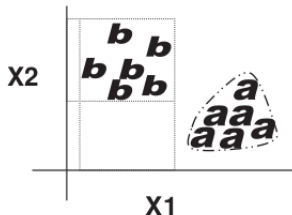
## Calcul LCBD

- ▶  $\text{Var}(Y)$ ,  $SS_{\text{total}}$ ,  $BD_{\text{total}}$ ,  $SS_i$ ,  $LCBD_i$ ...
- ▶ Test permutation (correction?)

# SDM - Données BIOCLIM

heightVariable	Description
1	Annual Mean Temperature
2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
3	Isothermality (BIO2/BIO7) (* 100)
4	Temperature Seasonality (standard deviation *100)
5	Max Temperature of Warmest Month
6	Min Temperature of Coldest Month
7	Temperature Annual Range (BIO5-BIO6)
8	Mean Temperature of Wettest Quarter
9	Mean Temperature of Driest Quarter
10	Mean Temperature of Warmest Quarter
11	Mean Temperature of Coldest Quarter
12	Annual Precipitation
13	Precipitation of Wettest Month
14	Precipitation of Driest Month
15	Precipitation Seasonality (Coefficient of Variation)
16	Precipitation of Wettest Quarter
17	Precipitation of Driest Quarter
18	Precipitation of Warmest Quarter
19	Precipitation of Coldest Quarter

# SDM - Méthode BIOCLIM



*Fig. 8.1.* Graphical representation of environmental envelopes shown for two environmental variables, X1 and X2. The b's represent observations of Species B, enclosed by a box whose boundaries are defined by the maximum and minimum values of X1 and X2 where Species B is observed. This is the boxcar or parallelepiped classifier used in BIOCLIM (typically the box is defined that encompasses 95% of the observations; see text). In contrast, the a's, representing observations of Species A, are encompassed in a minimum bounding box or convex hull, a simple example of the approach used in HABITAT (see text).

**Figure:** BIOCLIM representation (from Franklin 2010)



# SDM - Méthode BIOCLIM

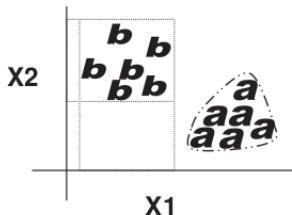
- ▶ Méthode « climate-envelope-model » classique (Nix 1986)
- ▶ Boxcar/parallelepiped/hyper-box
- ▶ Valeur env min-max pour 95% observations

## Fonctionnement

- ▶ Distribution percentile var.env des sites connus
- ▶ Comparaison chaque variable site inconnu, score 0 à 1 ramené entre 0 et 0.5
- ▶ Loi Liebig du minimum, prédiction = plus petit score

## Avantages

- ▶ Simplicité conceptuelle
- ▶ Utilisation très répandue



# À venir

- ▶ Échantillonnage Checklists par pixel
- ▶ Random Forests
- ▶ Prévisions changements climatiques
- ▶ Données d'interactions

# Autres points

- ▶ Formation comité-conseil
- ▶ Travail BIO6077

# SDM - Méthode BIOCLIM

Données brutes - 1 espèce, beaucoup d'observations

Données brutes - 1 espèce, beaucoup d'observations

# Données brutes - 1 espèce, moins d'observations

# Données brutes - Matrice $Y$ (non triée)



# Données brutes - Matrice $Y$ (triée par richesse, puis fréquence)

# Données brutes- Richesse spécifique (nombre d'espèces)

# Données brutes - Diversité spécifique

# Données brutes - Diversité spécifique

# Données brutes- LCBD

# Données brutes - LCBD significatives

# Données brutes - Relation LCBD-richesse

SDM - 1 espèce, beaucoup d'observations



## SDM - 1 espèce, moins d'observations

# SDM - Matrice $Y$ (non triée)

## SDM - Matrice $Y$ (triée par richesse, puis fréquence)

## SDM- Richesse spécifique (nombre d'espèces)

# SDM - Diversité spécifique

# SDM - Diversité spécifique

# SDM - LCBD

# SDM - LCBD significatives



# SDM - Relation LCBD-richesse