



# Questions méthodologiques autour des SP

- ▶ Définitions Sciences Participatives (et ce que l'on ne traitera pas)
- ▶ Schémas Rôles Acteurs / SP
- ▶ Recommandations pour les programmes de SP
- ▶ Analyse exploratoire et difficultés
  - ▶ Ex 1 : APECS    Ex 2 : BioObs plongeurs
- ▶ Types de données
- ▶ Statistique en SP
- ▶ Biais potentiels à travers 2 exemples
  - ▶ Ex 3 : Pl@ntNet
  - ▶ Ex 4 : Papillons Parc Nat. des Ecrins
- ▶ Questions ouvertes ou en cours



# Définitions SP

Les sciences participatives sont définies comme les formes de **production** de connaissances **scientifiques** auxquelles des **acteurs non-scientifiques ou professionnels**, qu'il s'agisse d'individus ou de groupes, participent de **façon active et délibérée** (Houllier, Merilhou-Goudard 2016)

- ▶ Très grande diversité et difficultés à s'y retrouver (terminologie fluctuante : S. citoyennes, S. contributives, S. communautaires, ...)
- ▶ Un continuum qui va du crowdsourcing jusqu'à la recherche action (RAP)



# Définitions SP

Les sciences participatives sont définies comme les formes de **production** de connaissances **scientifiques** auxquelles des **acteurs non-scientifiques ou professionnels**, qu'il s'agisse d'individus ou de groupes, participent de **façon active et délibérée** (Houllier, Merilhou-Goudard 2016)

- ▶ Très grande diversité et difficultés à s'y retrouver (terminologie fluctuante : S. citoyennes, S. contributives, S. communautaires, ...)
- ▶ Un continuum qui va du crowdsourcing jusqu'à la recherche action (RAP)



# Sciences participatives de type "Naturalistes"

- ▶ Programme ayant pour but :
  - ▶ un accroissement des connaissances sur une espèce, une zone, l'environnement
  - ▶ dans le domaine de la biodiversité et l'écologie
- ▶ Grande hétérogénéité, et un continuum
- ▶ Crowdsourcing, Appis type Plantnet, ObsEnMer ou BioObs, plus ou moins opportuniste, observateurs plus ou moins impliqués,
- ▶ Protocole light ou absent à très contraignant (Spipoll, STOC, 100 à 300 part.)
- ▶ Types de données variés (Présence, comptages, présence/absence, traces de présence, ...)
- ▶ et souvent assemblage de plusieurs types de données



# Sciences participatives de type "Naturalistes"

- ▶ Programme ayant pour but :
  - ▶ un accroissement des connaissances sur une espèce, une zone, l'environnement
  - ▶ dans le domaine de la biodiversité et l'écologie
- ▶ Grande hétérogénéité, et un continuum
- ▶ Crowdsourcing, Appis type Plantnet, ObsEnMer ou BioObs, plus ou moins opportuniste, observateurs plus ou moins impliqués,
- ▶ Protocole light ou absent à très contraignant (Spipoll, STOC, 100 à 300 part.)
- ▶ Types de données variés (Présence, comptages, présence/absence, traces de présence, ...)
- ▶ et souvent assemblage de plusieurs types de données



# De multiples acteurs et rôles

- ▶ Des individus : citoyens, professionnels
- ▶ Des scientifiques de toutes disciplines : écologie, environnement, sociologie
- ▶ Des structures de tous types : associations naturalistes, militantes, professionnelles, Des ONG, des structures éducatives (Ecoles, Formations)
- ▶ Des rôles essentiels : recueil des données, gestion archivage des données
- ▶ Des rôles importants : gouvernance du programme, animation, communication, utilisation et valorisation
- ▶ Des rôles complémentaires : validation, analyses, formation,

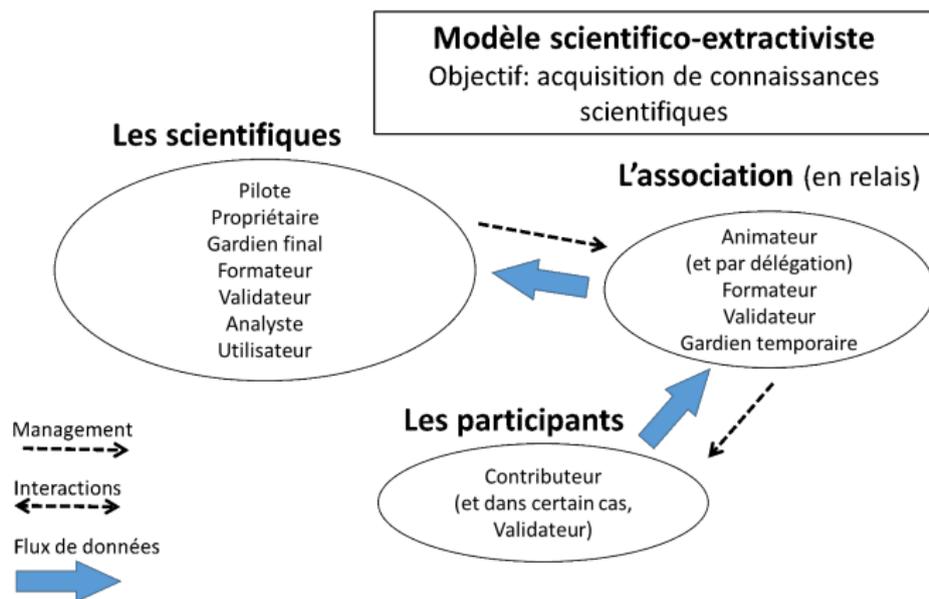


# De multiples acteurs et rôles

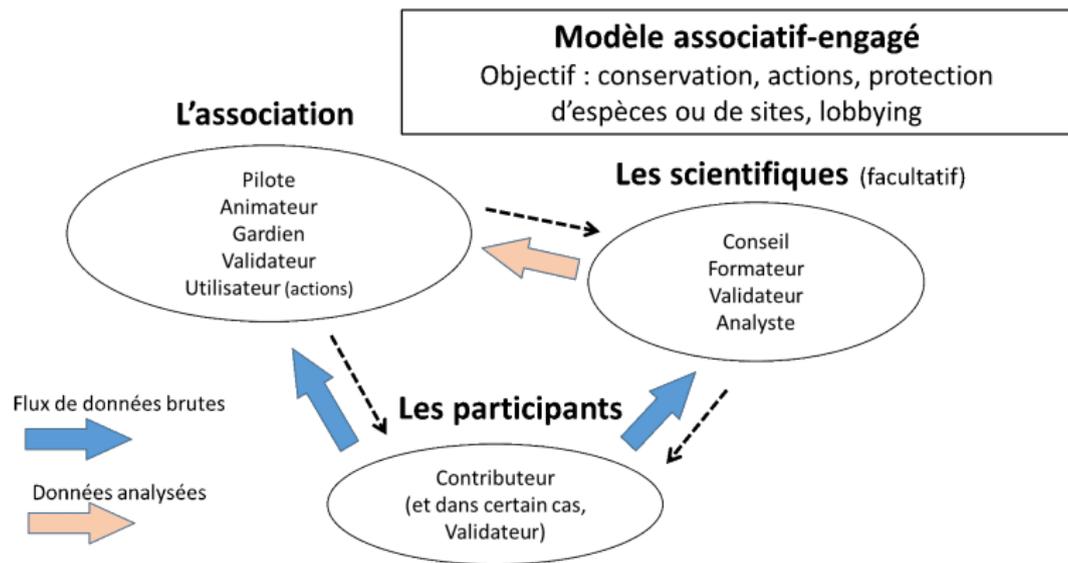
- ▶ Des individus : citoyens, professionnels
- ▶ Des scientifiques de toutes disciplines : écologie, environnement, sociologie
- ▶ Des structures de tous types : associations naturalistes, militantes, professionnelles, Des ONG, des structures éducatives (Ecoles, Formations)
- ▶ Des rôles essentiels : recueil des données, gestion archivage des données
- ▶ Des rôles importants : gouvernance du programme, animation, communication, utilisation et valorisation
- ▶ Des rôles complémentaires : validation, analyses, formation,



# Des fonctionnements très différents



# Des fonctionnements très différents



# Dix principes pour bien gérer un programme de Sciences Participatives

- ▶ 1 : citizen science projects actively involve citizens in scientific endeavour that generates new knowledge or understanding
- ▶ 2 : citizen science projects have a genuine science outcome
- ▶ 3 : both the professional scientists and the citizen benefit from taking part
- ▶ 4 : citizen scientists may participate in various stages of the scientific process
- ▶ 5 : citizen scientists receive feedback from the project
- ▶ 6 : citizen science is considered a research approach like any other, with limitations and biases that should be considered and controlled for

From : Kelly R, Fleming A, Pecl GT, von Gönner J, Bonn A. 2020 Citizen science and marine conservation : a global review. Phil. Trans. R. Soc. B 375 : 20190461. <http://dx.doi.org/10.1098/rstb.2019.0461>

# Dix principes pour bien gérer un programme de Sciences Participatives

- ▶ 7 : Citizen science project data are made publicly available, and results are published in an open-access format
- ▶ 8 : Citizen scientists are acknowledged in projects results and publications
- ▶ 9 : Citizen science programs are evaluated for their scientific output, data quality, participant experience and wider societal or policy impact
- ▶ 10 : The leaders of citizen science projects take into consideration legal and ethical issues (copyright, intellectual property, data sharing agreements, confidentiality, environmental impacts,...





# Exemple 1 : A.P.E.C.S (étude et conservation sélaciens)

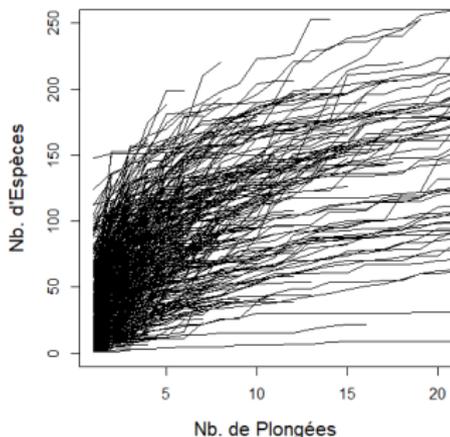
- ▶ Un programme historique (depuis 1998) : **Le Requin Pélerin**
- ▶ Données : opportunistes, sans effort d'observation connu,
- ▶ Comptage, lieu, date et heure, comportement, conditions d'obs.
- ▶ Données assez rares



- ▶ Données intéressantes (car rares)
- ▶ Analyse statistique difficile (impossible?)
- ▶ Nombreux biais potentiels : échantillonnage spatial, hétérogénéité effort, détection, identification
- ▶ EN RESTER A UNE ANALYSE EXPLORATOIRE DES DONNEES
- ▶ Autres programmes (CapOeRa : capsules oeufs de raies) tout aussi difficiles

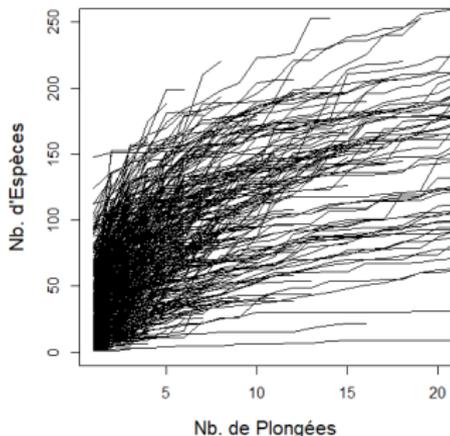
## Exemple 2 : BioObs, Observation faune sous marine par des plongeurs

- ▶ 3000 sites, 500000 observations, 5500 plongeurs
  - ▶ Données opportunistes (présence seule) ou par Checklist (présence vs non détection) Effort mal quantifié (la plongée)
  - ▶ Procédure de validation (et note de doute d'identification)
  - ▶ Données transmises à INPN (PatriNat)
- 
- ▶ Pas de continuité spatiale  
-> analyse site par site
  - ▶ Indicateur de biodiversité dès 15 plongées/site/an et un minimum de mélange (5 à 10 plongeurs/sorties différentes)
  - ▶ Cartographie peu pertinente



## Exemple 2 : BioObs, Observation faune sous marine par des plongeurs

- ▶ 3000 sites, 500000 observations, 5500 plongeurs
  - ▶ Données opportunistes (présence seule) ou par Checklist (présence vs non détection) Effort mal quantifié (la plongée)
  - ▶ Procédure de validation (et note de doute d'identification)
  - ▶ Données transmises à INPN (PatriNat)
- 
- ▶ Pas de continuité spatiale  
-> analyse site par site
  - ▶ Indicateur de biodiversité dès 15 plongées/site/an et un minimum de mélange (5 à 10 plongeurs/sorties différentes)
  - ▶ Cartographie peu pertinente



# Types de données en SP

Programme Science Participative → données → modélisation (selon Q°)

Données peuvent être issues d'observateurs citoyens / observateurs professionnels ou de collectifs

Type de données	Echantillonnage
présences seules	opportuniste ou préférentiel
présences/absences	exhaustif
abondances (détection par distance sampling)	protocolé (dans le mode d'obs.) ou protocolé spatialement (comptages)



# Statistique en Science Participative

Programme Science Participative → données → modélisation (selon Q°)

Contexte de la modélisation de la distribution d'espèces (SDM) à partir de données issues de la SP

## QUESTIONS

- Conservation des espèces
- Evolution de la biodiversité
- Dynamiques de populations
- Adaptation aux changements globaux
- Invasions biologiques
- Changements usages du sol
- Réduction des habitats

## TYPES DE MODELE

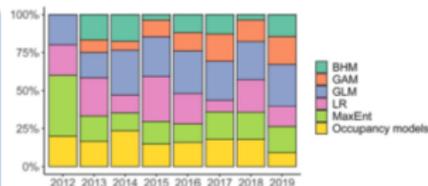
- Statique
- Dynamique
- Avec covariables (environnement comme usage du sol, conditions climatiques, pressions anthropiques...)

## TYPES DE DONNEES

- Présences seules
- Présences/absences
- Abondances

## METHODOLOGIES STATS

Linear regression approaches  
Maximum entropy (MaxEnt) -> PO  
GLM -> PA, Abundance  
Occupancy models -> PA  
Generalized additive models (GAM)  
Bayesian hierarchical models (BHM)



Feldman et al. (2021)

# Biais liés aux données de Sciences Participatives

1. **Biais spatio-temporel d'échantillonnage.**
2. **Biais de quantification de l'effort d'échantillonnage**
3. **Biais spatio-temporel de détection**
4. **Biais taxonomique**



# 1. Biais spatio-temporel d'échantillonnage

- ▶ Effort d'échantillonnage hétérogène dans l'espace/le temps.
- ▶ Affecte toutes les espèces ciblées par l'observateur également.
- ▶ Dépend du protocole, donc du type de données.
  - ▶ Non problématique si données standardisées (présence absence, ou détection/non-détection), si l'échelle des covariables est pertinente.
  - ▶ Problématique en présence-seule sans protocole (opportuniste)
  - ▶ Cas particulier : échantillonnage préférentiel



# 1. Biais spatio-temporel d'échantillonnage

## ► Biais n°1 avec PO.

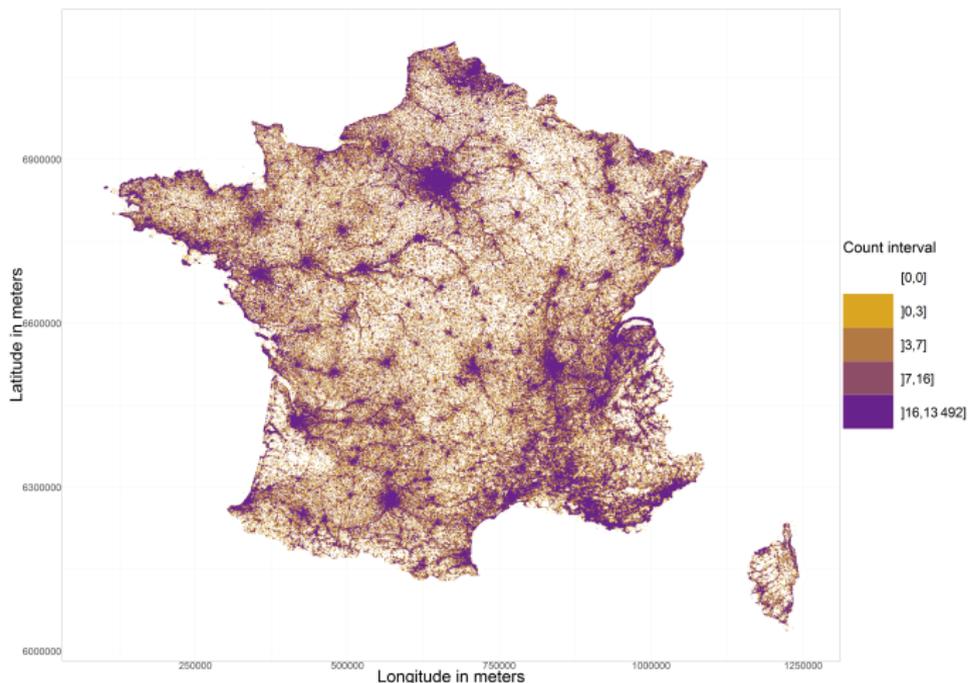


Figure – Counts of PI@ntNet records per km<sup>2</sup> up to 2019

# 1. Biais spatio-temporel d'échantillonnage

## Solutions ?

- ▶ Utilisation d'un proxy de l'effort d'échantillonnage :
  - ▶ ex : Phillips et al., 2009 (TGB) : Utilisent les observations d'un groupe d'espèces (le Target-Group) comme un proxy de l'effort d'échantillonnage.
- ▶ Modélisation jointe de l'effort d'échantillonnage et abondances (cf papillons plus tard)
- ▶ Bradter et al., 2018 : Inférence d'absences via des informations sur le comportement de chaque observateur. Les données PO opportunistes deviennent ainsi PA.



## 2. Biais de quantification de l'effort d'échantillonnage

- ▶ Diverses mesures de l'**effort d'échantillonnage** : nombre d'observations TG, temps d'écoute STOC...
- ▶ Problème : Mauvaise quantification/spécification
- ▶ Exemple : Correction TGB induit biais SI les espèces composant le TG ont des distributions corrélées (Botella et al., 2020).

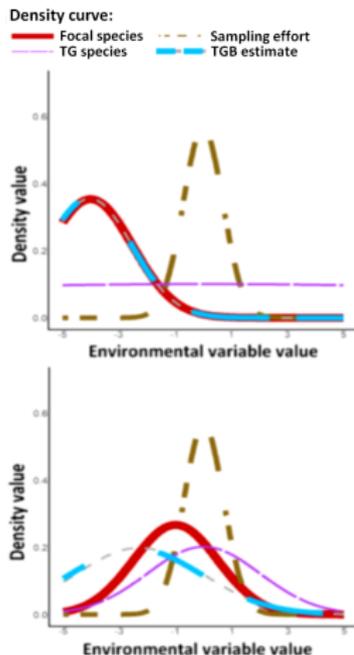


Figure – Distribution d'une espèce vs gradient environnemental, effort d'échantillonnage et estimateur TGB selon deux TG.

### 3. Biais spatio-temporel de détection

**La détectabilité** d'une espèce peut varier dans l'espace, le temps ou l'environnement (occurrence = présence x effort x détectabilité).

Ex1 : Une espèce peu détectable dans certaines conditions environnementales = biais négatif spécifique d'observation dans ces conditions.

Ex2 : Une espèce n'est plus rapportée au bout d'un certain temps alors qu'elle est là (cf Citron chez les papillons)

Si la probabilité de détection est forte -> Prendre en compte dans les modèle de détection/non-détection (Mackenzie et al., 2002).

Problématique en échantillonnage opportuniste...



## 4. Biais taxonomiques

- ▶ **Préférences spécifiques** courantes en échantillonnage opportuniste, mais dépend du profil de l'observateur (espèces remarquable? rares?).
- ▶ **Confusion taxonomiques** Erreurs systématiques, ignorance/doute sur certaines espèces, censure volontaire.  
Ex : Dans PI@ntNet, la confusion d'une espèce rare avec une espèce courante similaire.



## 4. Biases taxonomiques

Species ranked  
per confidence value:



- 1) *Acer monspessulanum* L. – 99.6
- ...



- 1) *Acer monspessulanum* L. – 21.7
- 2) *Acer pseudoplatanus* L. – 20.8
- 3) *Acer saccharinum* L. – 8.4
- ...

Species ranked per  
confidence value:



- 1) *Acer campestre* L. – 7.3
- 2) *Acer monspessulanum* L. – 5.9
- ...



- 1) *Acer monspessulanum* L. – 44.3
- 2) *Acer campestre* L. – 31.7
- 3) *Hedera helix* L. – 1.6
- ...

## 4. Biais de confusion taxonomique

- ▶ Les confusions taxonomiques peuvent induire des biais. Exemple : Distribution de plantes envahissantes à partir d'occurrences à l'identification incertaine (automatique) -> Le filtrage par le score de confiance améliore les performances (Botella et al., 2018).
- ▶ Méthode plus propre : Tenir directement compte des possibles confusions spécifiques dans l'ajustement du modèle statistique (Guilbault et al., 2019)

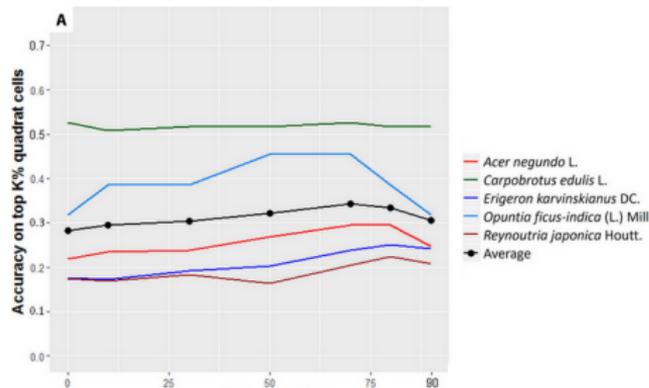
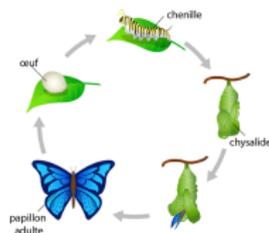


Figure – Performance SDM vs score minimum de confiance d'identification des observations

# Exemple : biais spatio-temporel d'échantillonnage

Modélisation spatio-temporelle de dynamiques de papillons à partir de données opportunistes dans le Parc National des Écrins (PNE)

Travail collaboration entre BioSP (+ Christophe Botella) et le PNE



- ▶ Modèles de distribution d'espèces (SDM)
- ▶ Détecter précocité de certaines espèces et voir l'évolution de cette précocité au cours des années. Effet climatique (changement des températures) sur l'émergence de lépidoptères.

Choix de 5 espèces d'intérêt : papillons univoltins diurnes



Aurore



Citron



Gazé



Demi-deuil



Apollon

# Données

- ▶ données de présence seule
- ▶ données de 1990 à 2020
- ▶ Grille régulière de pas 2000 m
- ▶ Pas de temps pour les comptages d'observation : 12 jours
- ▶ Covariables environnementales : altitude, couvert forestier

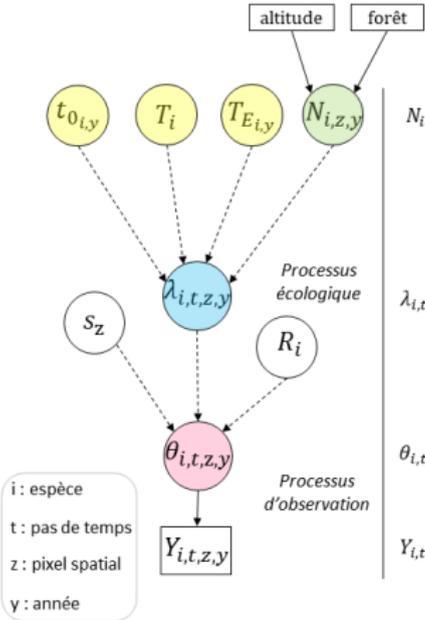


**Objectif (biologie)** : Modéliser la dynamique spatio-temporelle de l'émergence de papillons dans un environnement alpin

**Objectif (méthodo)** : concevoir une **estimation conjointe de la dynamique** spatio-temporelle de l'émergence de papillons et de l'**effort d'échantillonnage** (cadre : biais spatio-temporel d'échantillonnage).

Approche **mécanistico-statistique** : équation mécaniste (Soulsby & Thomas (2012)) pour la dynamique des papillons couplée à un processus d'observation statistique (Botella *et al.*, (2021, MEE))

# Modèle

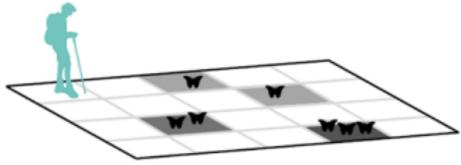
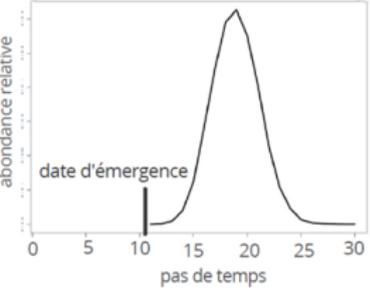
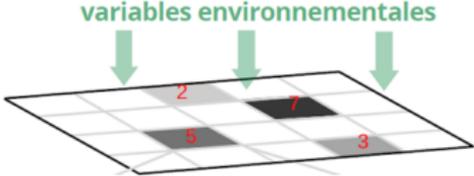


$$N_{i,z,y} = \exp(\beta_0 + \beta_1 * altitude(z) + \beta_2 * forêt(z))$$

$$\lambda_{i,t,z,y} = f(t_{0,i,y}, T_i, T_{E_{i,y}}, N_{i,z,y})$$

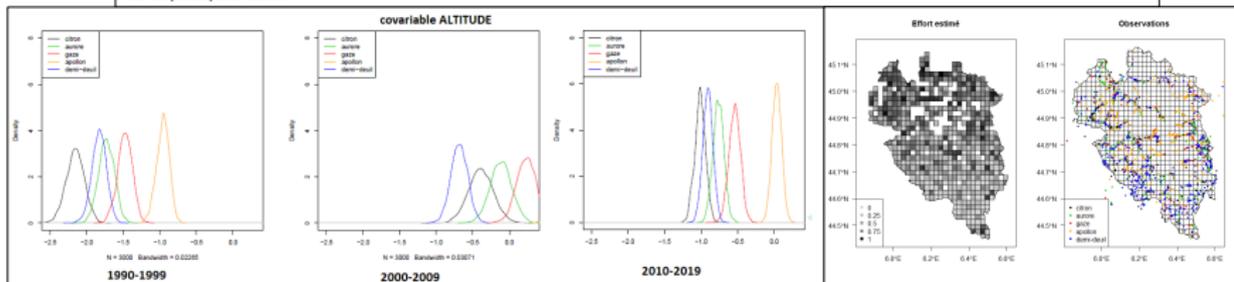
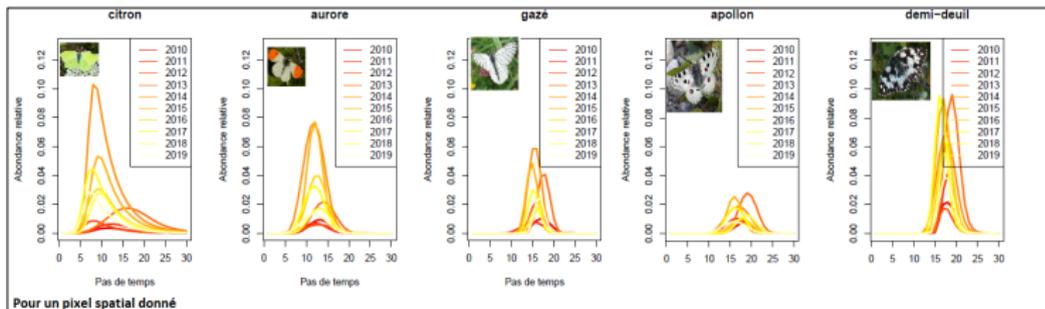
$$\theta_{i,t,z,y} = \lambda_{i,t,z,y} R_i S_z$$

$$Y_{i,t,z,y} \sim Poisson(\theta_{i,t,z,y})$$



# Résultats

Estimation de l'abondance de 5 espèces de lépidoptères, des effets de covariables, et de l'effort d'échantillonnage



# Conclusions

**Les données de SP apportent des grands volumes de données, une grande richesse spécifique, et une couverture spatio-temporelle inégalable !**

MAIS

- ▶ fort investissement nécessaire à la gestion du programme
- ▶ veiller à la pertinence des données SP par rapport aux questions
- ▶ tenir compte des biais potentiels liés au type de données SP

Donc intérêt des méthodes statistiques adaptées à cette complexité (modèles hiérarchiques bayésiens par exemple).

Observateurs deviennent des vecteurs d'images (collecte d'images, sons, vidéos, sans identification par l'observateur). Exemple : PI@ntNet



## Questions ouvertes ou en cours

- ▶ Beaucoup d'inconnues sur le **comportement sociologique de l'observateur**. Peu d'études quantifient la lassitude ou le désengagement des observateurs, la censure dans le reporting.
- ▶ Assemblage de données (Baptiste Alglave, soutenance 09/12/2022) : données protocolées + données professionnelles (échant. préférentiel)  
espèces plus rapportées que d'autres
- ▶ Aller vers la modélisation **multi-espèces** + **spatial** en tenant compte des biais potentiels (community SDM, Joint Species Distribution Model)
- ▶ Suivi/reconstruction de la dynamique spatiale (Botella et al., 2022), vers des prédictions de tendances futures ?



# Merci pour votre attention !





Kelly R, Fleming A, Pecl GT, von Gönner J, Bonn A. (2020)  
Citizen science and marine conservation : a global review.  
*Philosophical Transactions of the Royal Society B*, 375(1814), 20190461.



Feldman MJ, Imbeau L, Marchand P, Mazerolle MJ, Darveau M, Fenton NJ (2021)  
Trends and gaps in the use of citizen science derived data as input for species distribution models : A quantitative review.  
*PloS One*, 16(3)



Botella, C., Joly, A., Monestiez, P., Bonnet, P., & Munoz, F. (2020).  
Bias in presence-only niche models related to sampling effort and species niches : Lessons for background point selection.  
*Plos one*, 15(5), e0232078.



Botella, C., Joly, A., Bonnet, P., Munoz, F., & Monestiez, P. (2021).  
Jointly estimating spatial sampling effort and habitat suitability for multiple species from opportunistic presence-only data.  
*Methods in Ecology and Evolution*, 12 :933–945



Guilbault, E., Renner, I., Mahony, M., & Beh, E. (2021).  
How to make use of unlabeled observations in species distribution modeling using point process models.  
*Ecology and evolution*, 11(10), 5220-5243.



Botella, C., Bonnet, P., Hui, C., Joly, A., & Richardson, D. M. (2022).  
Dynamic Species Distribution Modeling Reveals the Pivotal Role of Human-Mediated Long-Distance Dispersal in Plant Invasion.  
*Biology*, 11(9), 1293.