

Introduction

Au départ...



Comment l'environnement peut expliquer -ou non- l'abondance de moustiques du genre *Aedes* en Polynésie Française, par des méthodes de machine learning

Introduction

Et finalement..



Nématode du pin

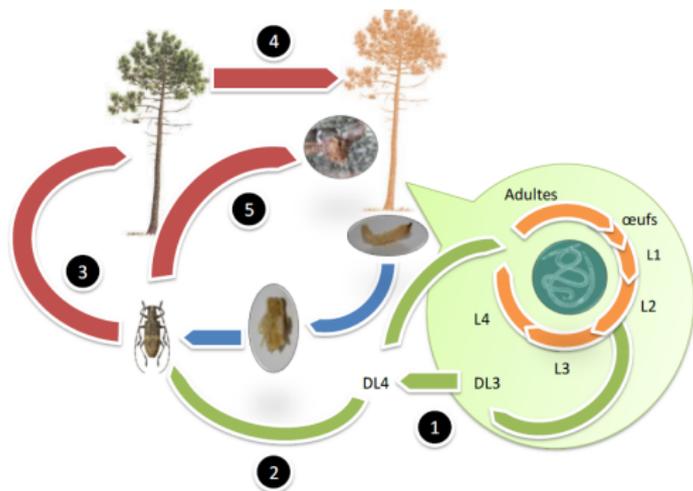
- ▶ *Bursaphelenchus xylophilus*, ver microscopique, parasite xylophage
- ▶ responsable du dépérissement de plusieurs espèces de pins (*Pinus halepensis*, *Pinus nigra* et *Pinus sylvestris*)

Fait des ravages au Portugal depuis 1999 et en Espagne depuis 2008
Risque élevé pour les forêts de pins en France (notamment dans les Landes)

Organisme de Quarantaine Prioritaire → surveillance obligatoire par les pays de l'UE

Dynamique spatio-temporelle du Monochamus

Cycles de vie du Monochamus et du nématode (Thèse G. David, 2014)

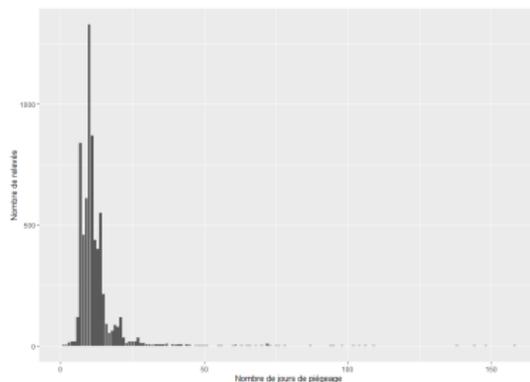
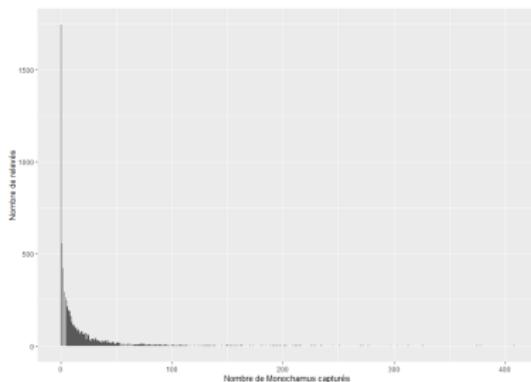


Questions biologiques

1. Identifier quels facteurs environnementaux sont favorables au vecteur Monochamus
2. Evolution des paramètres biologiques au cours des années
3. Perspective : modèle de distribution d'espèce pour pouvoir prédire l'abondance pour les analyses de risque (à partir des paramètres bio de l'insecte + param env)

Données

Monochamus p.



Echantillonnage sur les sites variable dans le temps dans une année y, et variable selon les années

Effort = nb de jours entre pose et relevé du piège, variable selon les sites et les années

Problèmes méthodologiques

- ▶ biais spatio-temporel avec un effort d'échantillonnage hétérogène en espace et en temps
- ▶ la capturabilité de l'espèce peut varier selon la météo, l'environnement forestier



Données

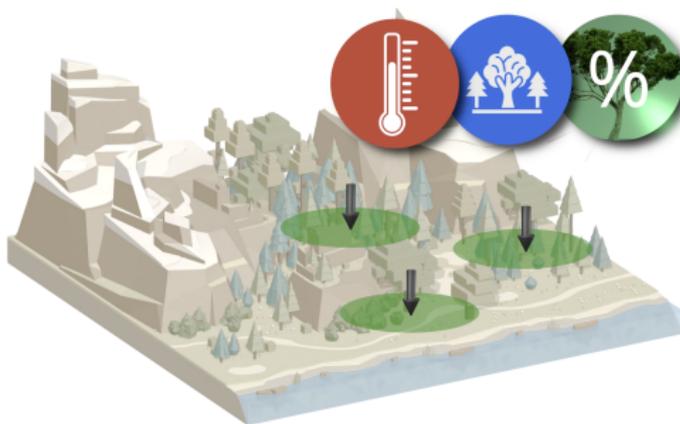
Covariables

Environnement forestier

- ▶ PinAlep : surface en km² de pins d'Alep dans un rayon de 2 km autour du piège (source : BD Foret V2 IGN).
- ▶ Tmixite : surface de forêt de pins / surface de forêt dans un rayon de 2 km autour du piège (source : BD Foret V2 IGN).

Données météo

- ▶ DateEmergence : date à laquelle 50% des Monochamus devraient avoir émergé au site z l'année y. Cette date dépend de la température moyenne journalière (source : Safran de Météo France)



Modèles hiérarchiques bayésiens

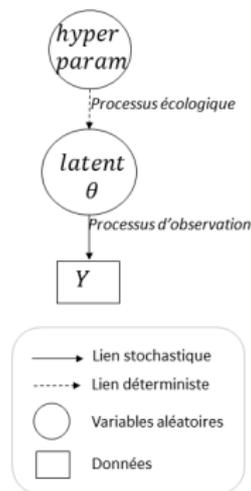
Objectif (méthodo) : concevoir une estimation de la **dynamique** spatio-temporelle de l'émergence de coléoptères tenant compte de l'**effort** d'échantillonnage hétérogène.

Pour tenir compte de ces biais, **modèles hiérarchiques** (dont les modèles à espaces d'état)

- ▶ cadre conceptuel où la couche des données est distincte de celle des variables latentes (couche cachée) et de celle des hyper-paramètres

Approche mécanistico-statistique

- ▶ Couche latente (= dynamique de population) déterministe
- ▶ Couche d'observation (= données) probabiliste/stochastique
- ▶ Méthode d'estimation des paramètres et des variables latentes : **inférence bayésienne**.
Incertitude des variables et des paramètres.

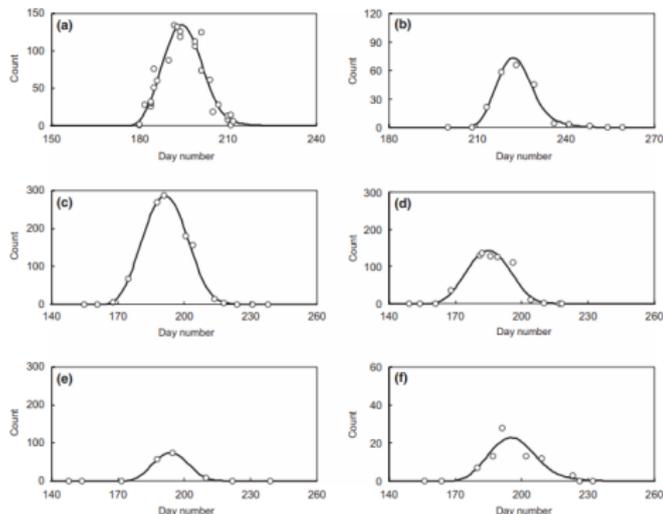


Dynamique spatio-temporelle de *Monochamus*

Approche mécanistico-statistique

Ce que l'on souhaite obtenir pour chaque site et par année :

R. L. Soulsby and J. A. Thomas



Approche **mécanistico-statistique**, avec un modèle inspiré de :

- ▶ Botella *et al.*, (2021, MEE) pour le processus d'observation statistique (avec effort d'échantillonnage hétérogène)
- ▶ Soulsby & Thomas (2012) pour l'équation mécaniste pour la dynamique d'insecte (équation analytique disponible)

Modèle

Processus d'observation (adapté de Botella et al., 2020)

Le nombre d'occurrences Y au temps t , sur le site z , l'année y est modélisé par une loi Binomiale négative :

$$Y_{t,z,y} \sim \text{NegBin}(\text{Prob}_{t,z,y}, r) \quad (1)$$

Ce processus dépend de la dynamique d'abondance $\lambda_{t,z,y}$, de l'effort d'échantillonnage $s_{t,z,y}$:

$$\text{Prob}_{t,z,y} = r / (r + \mu_{t,z,y}) \quad (2)$$

$$\mu_{t,z,y} = \lambda_{t,z,y} s_{t,z,y} \quad (3)$$

Deux hypothèses :

- ▶ Effort hétérogène en espace et en temps
- ▶ r considéré constant pour tous les sites, au cours du temps

Modèle

Modèle mécaniste pour la dynamique d'émergence des insectes (Soulsby et Thomas, 2012)

$$\lambda_{t,z,y} = f(t_{0,y,z}, T, T_{E_{y,z}}, N_{z,y}) \quad (4)$$

$$\frac{d\lambda_{z,y,t}}{dt} = \frac{N_{z,y}}{T_{E_{y,z}}} E\left(\frac{t'}{T_{E_{y,z}}}\right) - \frac{\lambda_{z,y,t}}{T} \quad (5)$$

$N_{z,y}$: la population totale détectable

$t_{0,y,z}$: la date du début de l'éclosion

$T_{E_{y,z}}$: la durée de la période d'éclosion

T : l'espérance de vie moyenne de l'insecte

t' = temps écoulé depuis le début de l'éclosion : $t' = t - t_0$

La fonction d'éclosion $E(\cdot)$ se définit comme :

$$\begin{cases} E(t'/T_E) = \frac{3\pi}{4} \sin^3\left(\frac{\pi t'}{T_E}\right) & \text{pour } 0 \leq t' \leq T_E \\ E(t'/T_E) = 0 & \text{pour } t' < 0 \text{ et } t' > T_E \end{cases}$$

Modèle

Modélisation de l'abondance relative

Régression sur les facteurs environnementaux :

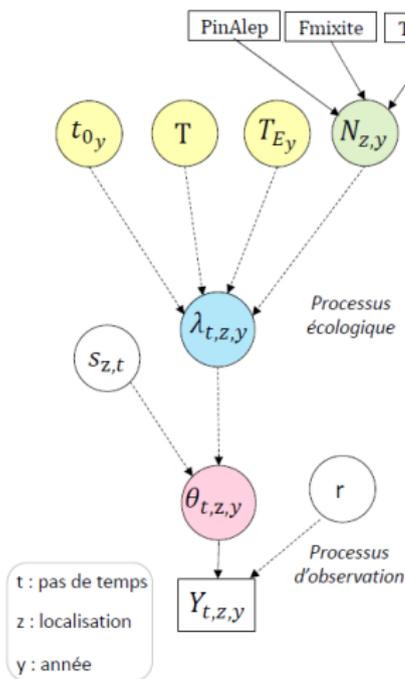
$$N_{z,y} = \exp(\beta_{0z,y} + \beta_1 * Tmixite(z) + \beta_2 * pinAlep(z) + \beta_3 * DateTPemerge(z, y))$$

(6)



Modèle

Processus d'observation, modèle mécaniste, modélisation de l'abondance relative



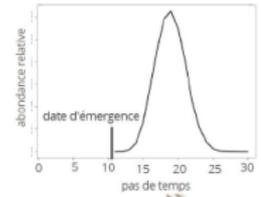
$$N_{z,y} = \exp(\beta_{0,t,y} + \beta_1 * Fmixite(z) + \beta_2 * PinAlep(z) + \beta_3 * T^\circ cumul(z,y))$$

$$\lambda_{t,z,y} = f(t_{0,y}, T, T_{E,y}, N_{z,y})$$

$$\theta_{t,z,y} = \lambda_{t,z,y} S_{z,y}$$

$$prob_{t,z,y} = r + r / \theta_{t,z,y}$$

$$Y_{t,z,y} \sim NegBinomial(prob_{t,z,y}, r)$$



Estimation

Estimation avec Nimble (3 chaînes, 30 000 itérations)

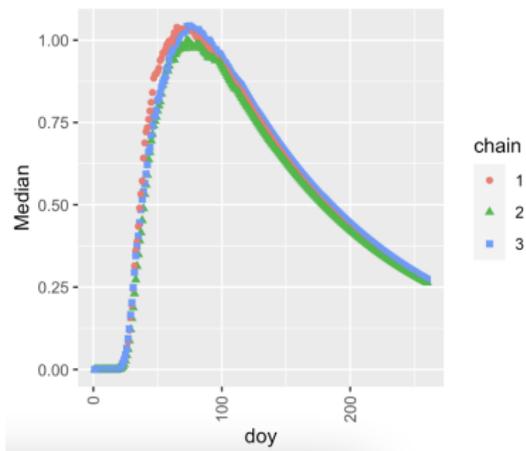
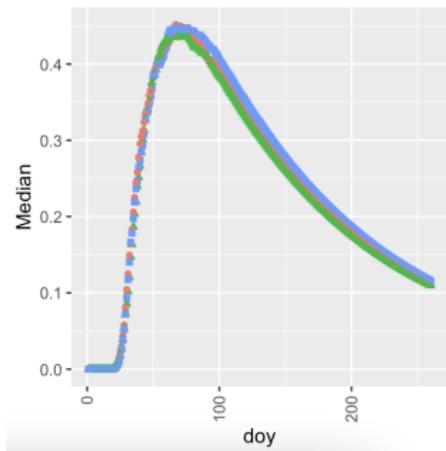
Priors non informatifs, sauf pour l'espérance de vie de l'insecte (126 jours, mesuré en labo selon David 2014)

100 sites sur les 1779 pour l'instant

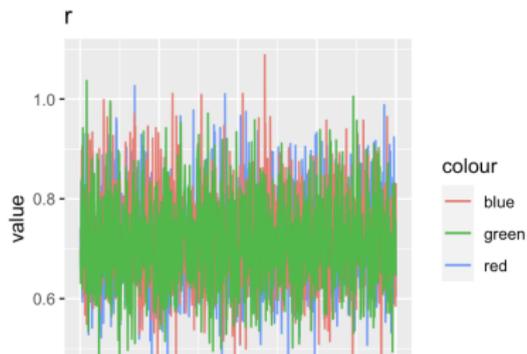


Résultats

Dynamique d'abondance d'insectes λ (1 site donné pour 2017 et 2018)

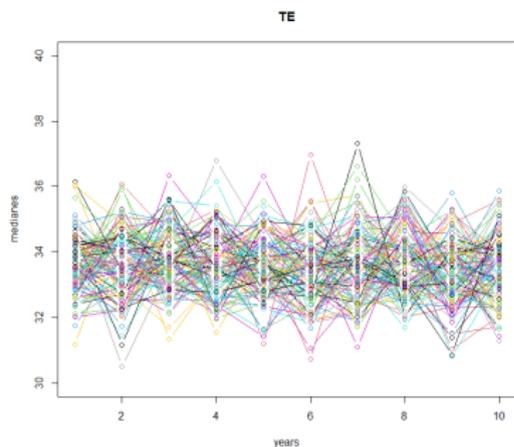
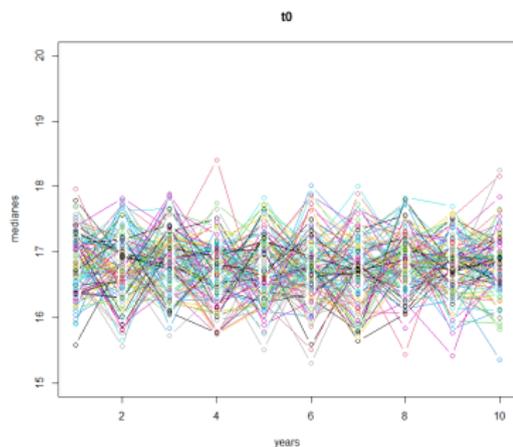


Processus d'observation



Résultats

Modèle mécaniste : paramètres biologiques (date 1ère émergence t_0 , durée d'éclosion TE, espérance de vie TT)



Résultats

Facteurs environnementaux

Problèmes de convergence

$T_{mixite} < 0$

$P_{inAlep} > 0$

$DateEmergence = 0$

Les covariables n'expliquent pas assez l'abondance relative ?



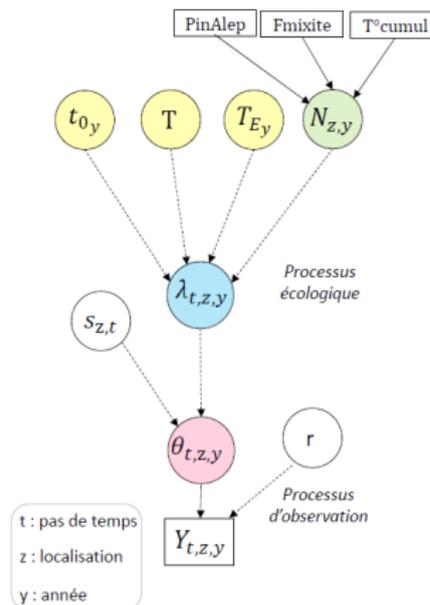
Résultats et discussion

Retour sur les hypothèses et choix de modélisation

- ▶ Estimation journalière car plus petite unité d'échantillonnage, estimation par site
- ▶ grand nombre de sites France entière (1779) : difficile pour l'estimation

Processus d'observation

- ▶ Choix Negative Binomiale ou Poisson ? (et choix de r constant)
- ▶ Où intégrer l'effort d'échantillonnage ?
- ▶ ajout : précip sur processus d'observation (car à partir d'un seuil, moins de captures) ?



Discussion

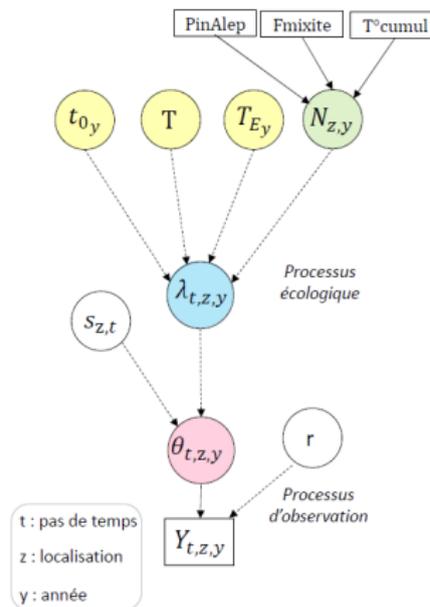
Retour sur les hypothèses et choix de modélisation

Modèle mécaniste

- choix : date de 1ère émergence (t_0) connue en labo et intégrée en inits

Abondance relative en fonction de l'environnement

- Choix des facteurs des covariables : sélectionnés par random forest, mais plus grand choix disponible
- offset comme effet année ou effet site-année ?
- pas d'inter-annuel (N de l'année $y-1$ joue sur N de l'année y)



Conclusion et perspectives

- ▶ Modèle qui prend en compte l'hétérogénéité de l'effort d'échantillonnage, les paramètres biologiques connus en labo et les covariables environnementales
- ▶ Reste à renforcer les résultats, faire l'estimation pour tous les sites
- ▶ Dépendance spatiale entre les pièges pour prendre en compte la dispersion du monochamus
- ▶ Prédiction à faire sur l'ensemble du territoire
- ▶ A terme : combiner de la diffusion courte et à longue distance ? intégrer de l'info génétique dans le modèle ?

Remerciements : Plateforme ESV, membres du GT surveillance nématode du pin, Loïc pour le cluster BioSP

Merci pour votre attention !