



Environnement Méca-Stat (MSE) : présentation de l'environnement et d'un cas d'étude

Webinaire IMOTEP

Olivier Bonnefon, Emily Walker





1. Introduction



Spatio-temporal models of range expansion and distribution

Introduction

- Heterogeneous data
- Multiple sources
- Observation errors
- Multiple temporal and spatial scales...

**Dealing with
realistic data**

**Spatio-temporal models of range
expansion and distribution**

Introduction

- Heterogeneous data
- Multiple sources
- Observation errors
- Multiple temporal and spatial scales...

**Dealing with
realistic data**

- Species interactions
- Dispersal mechanisms
- Attractive spots, behaviour
- Density dependencies
- Env. boundaries...

**Modeling the
effects of the
heterogeneities**

**Spatio-temporal models of range
expansion and distribution**

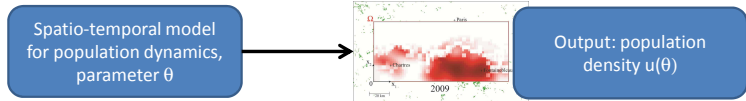
Faire le lien modèle - données : l'approche mécanistico-statistique

Confronter des modèles mécanistes paramétrés à des données observées ou mesurées, ne reflétant qu'une vision imparfaite des processus étudiés (censure, présence-absence, etc.).

Construire un processus d'observation permettant de mesurer la vraisemblance des données pour un jeu de paramètres fixés (problème d'optimisation ou d'inférence bayésienne).

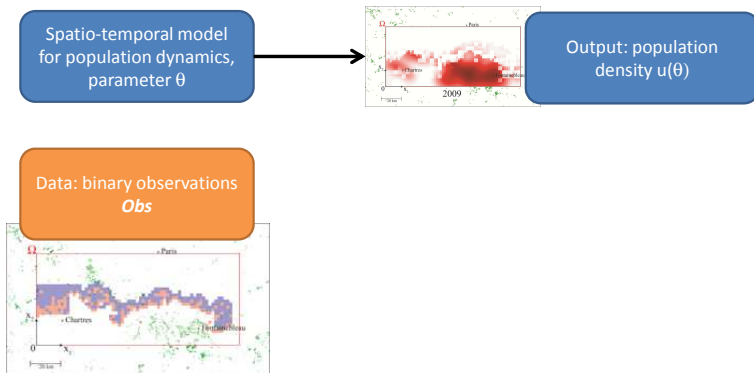
Faire le lien modèle - données : l'approche mécanistico-statistique

Exemple de la processionnaire du pin



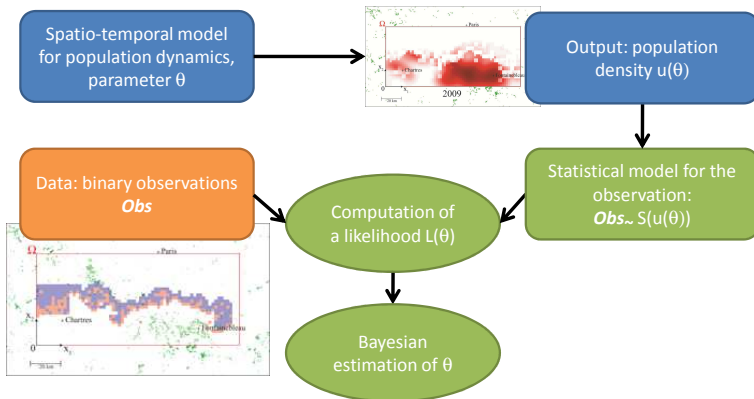
Faire le lien modèle - données : l'approche mécanistico-statistique

Exemple de la processionnaire du pin



Faire le lien modèle - données : l'approche mécanistico-statistique

Exemple de la processionnaire du pin





2. Quelques considérations théoriques

Motivation et difficultés

Soit un modèle mécaniste $u_\theta = \mathcal{M}(\theta)$ paramétré par θ (capacité de dispersion, coefficient de régression, taux de croissance, conditions initiales, ...)

Objectif : estimer les paramètres θ à partir des observations \mathcal{O} .

Méthode classique : on compare les observations \mathcal{O} aux sorties $u_{\tilde{\theta}}$.

Problèmes : liés aux données \mathcal{O} :

- ▶ données bruitées, signal binaire ;
- ▶ forte censure spatiale et temporelle ;
- ▶ observation indirecte de u_θ .

→ Comment comparer \mathcal{O} et $u_{\tilde{\theta}}$?

L'approche mécanistico-statistique

\mathcal{O} : observation d'un processus $J(u_\theta)$ avec $u_\theta = \mathcal{M}(\theta)$.

- ▶ Construire un **modèle probabiliste du processus d'observation** : $\mathcal{O} \sim \mathcal{S}(J(u_\theta))$.
- ▶ Calculer la vraisemblance $\mathcal{L}(\tilde{\theta}) = P(\mathcal{O}|u_{\tilde{\theta}}) = P(\mathcal{O}|\tilde{\theta})$.
- ▶ Différentes méthodes d'estimation peuvent être utilisées : maximum de vraisemblance, méthode bayésienne, ...

Remarque : les modèles mécanistico-statistiques \subset modèles à espace d'états.

Exemples :

- Sciences de l'environnement : Berliner (2003), Wikle (2003a) and Campbell (2004).
- Ecologie : Wikle (2003b), Buckland et al. (2004), Soubeyrand et al. (2009), Abboud et al. (2018), Louvrier et al. (2020), Papaïx et al. (2020), Roques et al. (2020).



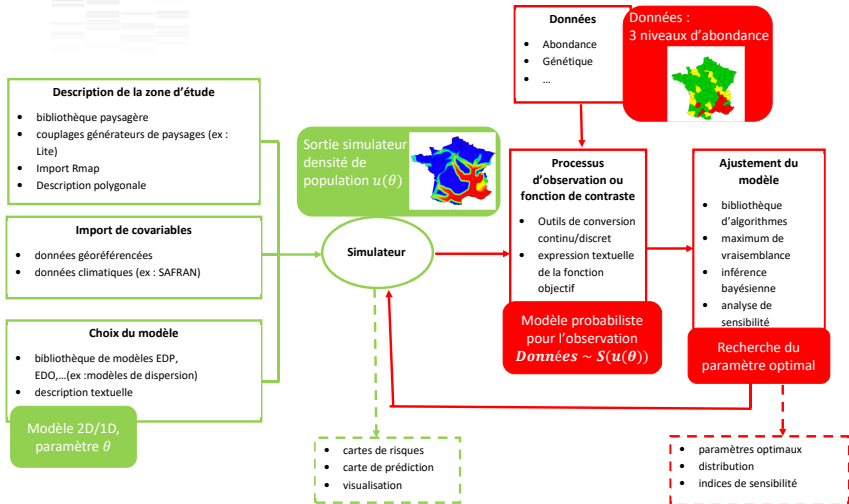
3. Environnement : MSE

Pourquoi un environnement dédié à la mise en œuvre de modèles mécanistico-statistiques ?

Faciliter les études basées sur des **modèles spatiaux en épidémiologie ou en dynamique des populations**

Environnement portable et utilisable par des non spécialistes du calcul

Rendre accessible les modèles d'équations aux dérivées partielles et permettre leur ajustement





- ▶ Code python
- ▶ Interface
- ▶ Connexion avec JAGS, R etc

A partir du site web : [https ://informatique-mia.inrae.fr/mse/](https://informatique-mia.inrae.fr/mse/)

Utilisation d'une image Docker contenant MSE
([https ://informatique-mia.inrae.fr/mse/node/22](https://informatique-mia.inrae.fr/mse/node/22))

Description des modules

Tutoriels

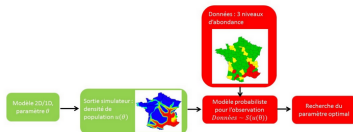
Environnement Mécanistico-Statistique (MSE)

Qu'est-ce que c'est ?

C'est un environnement informatique et numérique facilitant les études basées sur des modèles spatio-temporels en épidémiologie ou en dynamique des populations. Cet environnement se veut portable et utilisable par des non spécialistes du calcul. L'objectif est de rendre accessible les modèles d'équations aux dérivées partielles et de permettre leur ajustement. L'unité de recherche BioSP porte cette thématique de recherche et d'ingénierie, de nombreuses études et publications ont été menées. L'objectif est maintenant de rendre disponible ces travaux au plus grand nombre d'utilisateurs.

L'approche mécanistico-statistique consiste à confronter des modèles mécanistes paramétrés à des données observées ou mesurées. Il s'agit alors d'ajuster au mieux le modèle. Pour cela, on introduit un processus d'observation permettant de mesurer la vraisemblance des données pour un jeu de paramètres fixés. Il peut s'agir d'un problème d'optimisation ou d'inférence bayésienne.

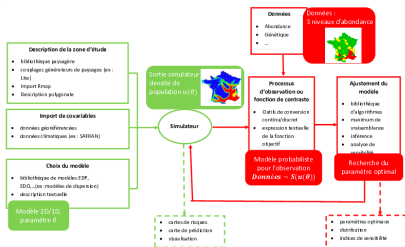
Illustration de l'approche mécanistico-statistique sur l'ajustement d'un modèle de dynamique de la population de moustique



Dans cet exemple, le **modèle mécaniste** est un système d'équations aux dérivées partielles paramétré simulant la propagation du moustique tigre. La **vraisemblance de données d'abondance** au niveau départemental est calculée par un modèle **probabiliste**. Un programme d'**optimisation** permet de calculer le jeu de paramètres maximisant cette vraisemblance.

Architecture modulaire de l'environnement mécanistico-

Figure – Rappels des principes de l'approche mécanistico-statistique



En vert : Un environnement de simulation et d'optimisation de modèles aux équations aux dérivées partielles (ESOMED)

En rouge : Un environnement d'ajustement et d'analyse de modèle

Liens vers la description des modules

[Module géométrique de description de la zone d'étude](#)

[Module d'import de covariables spatio-temporelles](#)

[Module de description du modèle](#)

[Module de processus d'observation ou de fonction de contraste](#)

[Module d'Optimisation](#)

[Module d'inférence bayésienne\(à venir\)](#)

[Module Analyse de sensibilité](#)

Figure – Description informatique des modules

Tutoriel - Estimation de la condition initiale

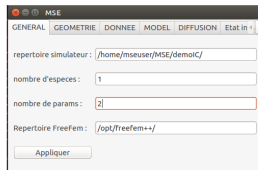
Introduction

Ce tutoriel illustre l'ajustement d'une donnée initiale paramétrée d'un modèle de réaction diffusion. Les deux paramètres ($p1, p2$) représentent le centre de la distribution de la donnée initiale. Il s'agit ici de déterminer les valeurs du couple ($p1, p2$) minimisant l'erreur quadratique entre la simulation et les données.

La suite de cette page décrit comment mener cette étude dans MSE, d'une part via l'interface, d'autre part via un script python.

Description de l'étude

Il s'agit donc d'une étude portant sur 1 espèce avec une donnée initiale paramétrée par 2 paramètres:



MSE

GENERAL GEOMETRIE DONNEE MODEL DIFFUSION Etat in

repertoire simulateur : /home/mseuser/MSE/demoic/

nombre d'espèces : 1

nombre de params : 2

Repertoire FreeFem : /opt/freefem++/

Appliquer

```
import os
import mse
MY_OUTPUT_DIR=os.getenv("MSE_OUTPUT_DIR", "/home/mseuser/MSE/demoic")
mse.setStringParam(mse.MSE_KEY_OUTPUT_DIR_MY_OUTPUT_DIR)
mse.setParam(mse.MSE_KEY_MODEL_NBSPESCES,1)
mse.setParam(mse.MSE_KEY_MODEL_NBPARAMS,2)
mse.initGradEdp()
```

Description de la zone d'étude

Figure – Des tutoriels



4. Application

Exemple de la chalarose du frêne

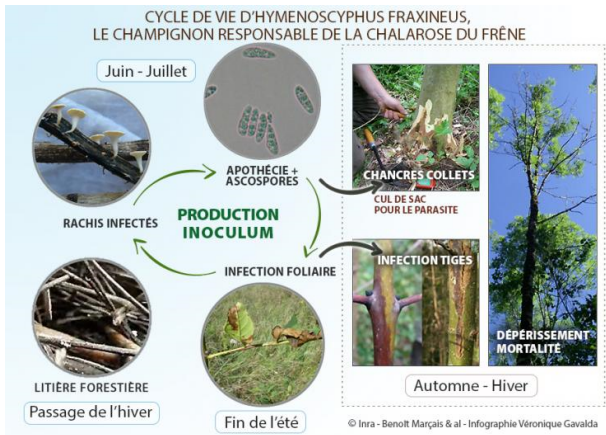
Avec la participation de Jean-Paul Soularue, Marie Grosdidier et Julien Papaïx

La chalarose du frêne

Maladie invasive
due au champignon *H. fraxineus* qui entraîne
le dépérissement des
frênes communs
(*Fraxinus excelsior*)

Premières observations
en Pologne dès le début
des années 90, puis le
reste de l'Europe.

Premier foyer observé
en France en 2008, en
Haute Saône.



La chalarose du frêne

Certaines conditions sont défavorables à l'expression de la maladie :

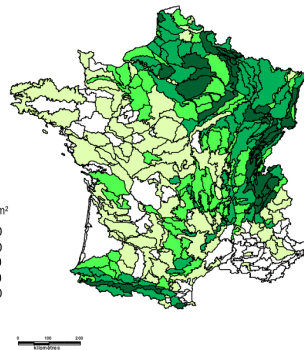
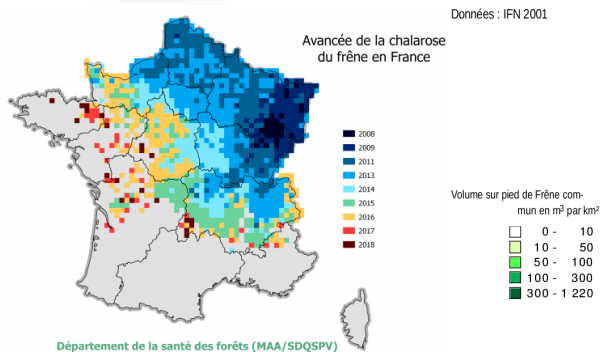
Température : sites trop chauds dans le sud de la France ($>35^{\circ}\text{C}$ en été). La température optimale est de l'ordre de 20°C .

Environnement :

- ▶ arbres isolés dans des microclimats secs et chauds
- ▶ arbres hors forêts, situés dans les petits bosquets, les haies, bords de route ou rivière (faible densité de frênes dans un rayon de 300m)
- ▶ faible densité de frênes dans un peuplement forestier (lorsque mélange avec d'autres essences)
- ▶ arbres en ville avec ramassage des feuilles mortes à l'automne.

Génétique : certains arbres sont résistants.

Les données



En plus de ces données d'observation, des expérimentations en laboratoire sont réalisées pour tester différentes hypothèses (T°, génétique etc.)

Objectifs de l'étude

Reconstruire la **dynamique spatio-temporelle** de la maladie à l'échelle de la France pour :

- ▶ estimer des paramètres clefs de la dynamique (dispersion, taux de croissance, etc.),
- ▶ prédire la dynamique future,
- ▶ comparer les différentes souches de *H. fraxineus*

Hypothèses :

Modélisation des spores sexuées (ascospores) uniquement, infectant le houppier (dispersion par le vent), car elles sont responsables de la propagation à large échelle de la chalarose.

Considération des effets du climat et de la répartition des populations de frênes sur la propagation.

Construction d'un modèle mécaniste paramétré

Modèle simpliste : modèle à 1 compartiment

$$\begin{cases} \partial_t u &= F 10^{p_1} \Delta u + p_2 u(1 - u) \\ u(0, x, y) &= D_{08}(x, y) \\ \partial_{\vec{n}} u &= 0 \end{cases}$$

avec :

- u , la proportion de Frênes infectés,
- F la répartition de frênes,
- p_1 le paramètre multiplicateur de la **diffusion** à estimer,
- p_2 le **taux de reproduction** de Chalarose à estimer.

Pas de prise en compte de la température.

Période de simulation : succession des périodes estivales, du 15 juillet au 31 août de chaque année.

Construction d'un modèle mécaniste paramétré

Modèle à 2 compartiments : spores V et arbres infectés I

$$\begin{cases} \partial_t V &= 10^{p_1} \Delta V + r(t)I - (\gamma_V + \Theta_1(T))V \\ \partial_t I &= p_2 V(F - I) - \Theta_2(T)I \end{cases}$$

avec :

$F(x,y)$: population d'hôtes (frênes), constant dans le temps, hétérogène,

$I(x,y,t)$: hôtes infectés, $F-I$ population saine,

$V(x,y,t)$: spores (vecteurs) qui dispersent (modèle spatial),

p_1 : **diffusion** sur la période d'émission des spores,

p_2 : **taux de croissance** de la Chalarose sur l'hôte,

$r(t) = 0.01$ taux de production de spores sur la période d'émission,

$\gamma_V = 0.01$ taux de mortalité intrinsèque des spores

Construction d'un modèle mécaniste paramétré

Modèle à 2 compartiments : spores V et arbres infectés I

$$\begin{cases} \partial_t V &= 10^{p_1} \Delta V + r(t)I - (\gamma_V + \Theta_1(T))V \\ \partial_t I &= p_2 V(F - I) - \Theta_2(T)I \end{cases}$$

avec :

$F(x,y)$: population d'hôtes (frênes)

$I(x,y,t)$: hôtes infectés, $F-I$ population saine

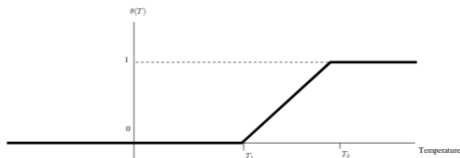
$V(x,y,t)$: spores (vecteurs)

p_1 : diffusion

p_2 : taux de croissance

Θ_1 : effet de la température T sur la mortalité des spores

Θ_2 : effet de la température T sur la mortalité de la Chalarose présente sur l'hôte



Modèle 1

Minimisation de la différence quadratique entre les données et la simulation.

$$c(p_1, p_2) = \sum_t \int (V(t, x, y) - D_t(x, y))^2 dx dy$$

Modèle 2

Processus d'observation :

$$\mathcal{O} \sim \text{Bernoulli}(I/F)$$

Maximum de vraisemblance par optimisation ou inférence bayésienne avec le module JAGS.



Modèle à 1 compartiment (proportion de frênes infectés) :
minimisation quadratique (exécution via docker)

Modèle à 2 compartiments (spores et nombre de frênes infectés) :
maximum de vraisemblance (exécution via docker)

Modèle à 2 compartiments (spores et nombre de frênes infectés) :
analyse bayésienne avec JAGS (exécution sur le cluster)

ça tourne toujours !

```
model {  
  # A PRIORI  
  #10^p1=diffusion  
  #p2=transmission  
  #gamma_v=mortalité  
  #r(t)=production de spores  
  
  thetaB ~ dmnorm(mu,tau)  
  mu <-c(1.5,1.2)  
  tau[1,1] <- 11.11 # (1/0.3*0.3)  
  tau[2,2] <- 1.56 # 1/0.8*0.8  
  tau[1,2] <- 0  
  tau[2,1] <- 0  
  
  theta[1] <- exp(thetaB[1])  
  theta[2] <- exp(thetaB[2])  
  
  e ~ dunif(0,1)  
  
  # Processus dynamique  
  lambda<-dynamic(theta[1],theta[2])  
  
  # vraisemblance  
  for (j in 1:1551){  
    w[j] <- lambda[1,j,1]+0.0001  
    p[j] <- 1-(1-e)^(w[j]*S)  
    Nobs[j] ~ dbern(p[j])  
    Nobs_rep[j] ~ dbern(p[j])  
  }  
}
```



5. Conclusion

Conclusion

Mise en oeuvre de l'approche mécanistico-statistique sur les données de chalarose du frêne grâce à l'environnement (MSE) :

- ▶ reproduire la dynamique de la maladie
- ▶ estimer les paramètres de diffusion et de taux de reproduction
- ▶ permettre de faire des simulations à partir du modèle calibré

Perspectives

- ▶ Amélioration de la définition du processus d'observation (pour mieux tenir compte des caractéristiques du protocole d'observation de la maladie)
- ▶ Prise en compte des résultats d'expériences en labo
- ▶ Effet de l'adaptation du pathogène (différentes souches), adaptation des frênes à la chalarose



Merci !