

## Modélisation mécanistico-statistique de la chararose

Coralie Fritsch



Collaboration avec

Anne Gégout-Petit



Benoît Marçais

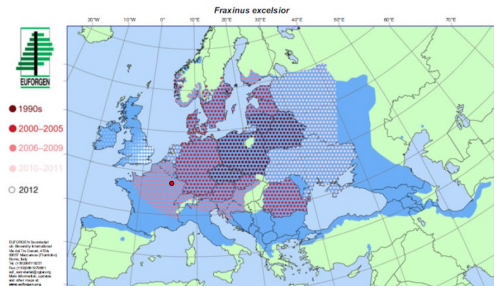


Marie Grosdidier



# La charlarose, une maladie fongique qui affecte les frênes

- ▶ émerge en Pologne en 1990
- ▶ 2008 en France
- ▶ invasion rapide et massive



This distribution map, showing the natural distribution area of *Fraxinus excelsior*, was compiled by members of the EUFORGEN Networks

Citation: Distribution map of Common ash (*Fraxinus excelsior*) EUFORGEN 2009, [www.euforgen.org](http://www.euforgen.org)

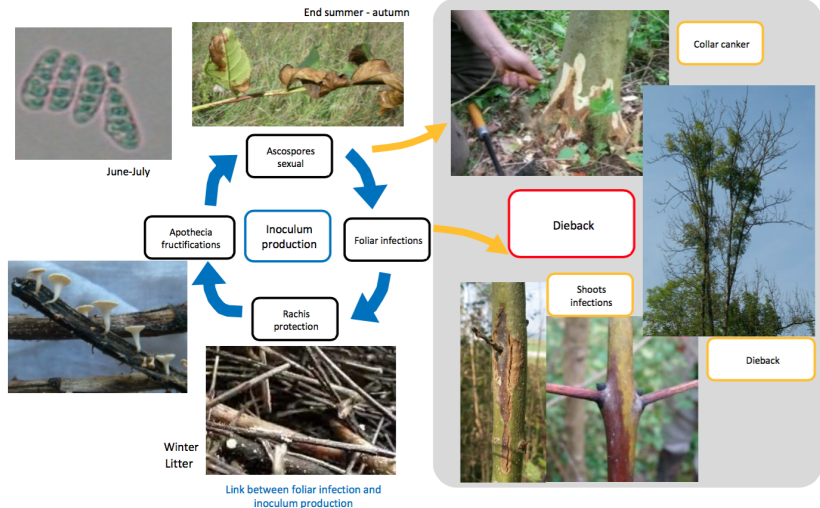
First published online in 2003 - Updated on 30 July 2008

0 250 500 1000 km

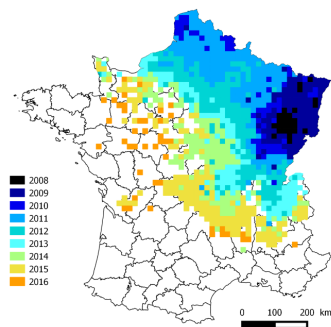


McKinney et al., 2014

# Cycle de la maladie

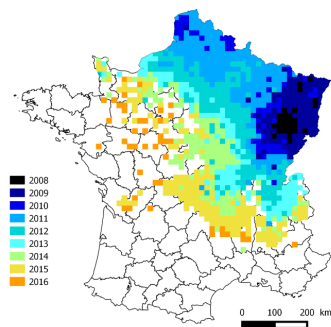


# Observations



1 signalement = proportion de Frênes infectés parmi 30 Frênes observés dans un quadrat  
~ 500 signalements / an

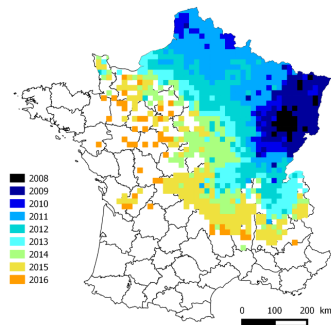
# Observations



- ▶ modèle de réaction-diffusion pour l'infection des rachis
- ▶ modèle pour le développement des symptômes
- ▶ loi des observations

1 signalement = proportion de Frênes infectés parmi 30 Frênes observés dans un quadrat  
~ 500 signalements / an

# Observations

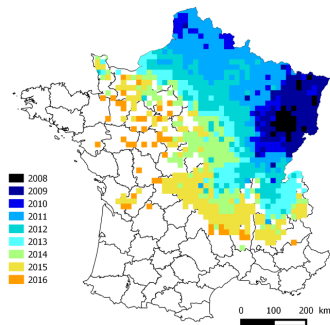


- ▶ modèle de réaction-diffusion pour l'infection des rachis
- ▶ modèle pour le développement des symptômes
- ▶ loi des observations



1 signalement = proportion de Frênes infectés parmi 30 Frênes observés dans un quadrat  
~ 500 signalements / an

# Observations



1 signalement = proportion de Frênes infectés parmi 30 Frênes observés dans un quadrat  
~ 500 signalements / an

- ▶ modèle de réaction-diffusion pour l'infection des rachis
- ▶ modèle pour le développement des symptômes
- ▶ loi des observations



⇒ Estimation des paramètres

Méthode inspirée de Roques, Soubeyrand, Rousselet (2011) et Abboud, Bonnefon, Parent, Soubeyrand (2018)

# Plan

## Modèles

- Modèle de réaction-diffusion pour l'infection des rachis

- Modèle pour le développement des symptômes

- Modélisation des observations

## Inférence statistique et résultats numériques

- Vraisemblance

- Identification des "bonnes" zones de paramètres

- Lois a posteriori

- Évolution de la chararose pour le paramètre qui maximise la vraisemblance

- Comparaison de plusieurs indices de température



## Modèle de réaction-diffusion pour l'infection des rachis

$R_a(x)$  = quantité de rachis infectés actifs au printemps de l'année  $a$  en  $x$ .

**Printemps de l'année  $a$  : production de la quantité de spores  $\nu_a$**

**Été de l'année  $a$  : dispersion des spores**

**Automne de l'année  $a$  : infection des rachis**

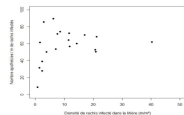
## Modèle de réaction-diffusion pour l'infection des rachis

$R_a(x)$  = quantité de rachis infectés actifs au printemps de l'année  $a$  en  $x$ .

**Printemps de l'année  $a$  : production de la quantité de spores  $\nu_a$**

$$\nu_a(x) = \begin{cases} H_a(x) R_a(x) & \text{if } R_a(x) \geq r \\ H_a(x) \frac{R_a(x)^2}{r} & \text{if } R_a(x) < r \end{cases}$$

avec  $H_a(x) = \beta_0 + \beta_1 h_a(x)$  où  $h_a$  est la pluviométrie du printemps de l'année  $a$ .



**Été de l'année  $a$  : dispersion des spores**

**Automne de l'année  $a$  : infection des rachis**

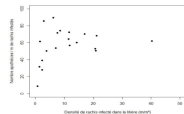
## Modèle de réaction-diffusion pour l'infection des rachis

$R_a(x)$  = quantité de rachis infectés actifs au printemps de l'année  $a$  en  $x$ .

**Printemps de l'année  $a$  : production de la quantité de spores  $\nu_a$**

$$\nu_a(x) = \begin{cases} H_a(x) R_a(x) & \text{if } R_a(x) \geq r \\ H_a(x) \frac{R_a(x)^2}{r} & \text{if } R_a(x) < r \end{cases}$$

avec  $H_a(x) = \beta_0 + \beta_1 h_a(x)$  où  $h_a$  est la pluviométrie du printemps de l'année  $a$ .



**Été de l'année  $a$  : dispersion des spores**

$$\frac{\partial w_a(t, x)}{\partial t} = D \Delta w_a(t, x) + \frac{\nu_a(x)}{\tau}$$

avec  $w_a(0, x) = 0$ ,  $\tau = 60$  j. À la fin de l'été, la quantité de spores est  $w_a(\tau, \cdot)$ .

**Automne de l'année  $a$  : infection des rachis**

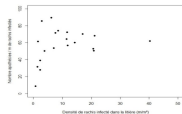
## Modèle de réaction-diffusion pour l'infection des rachis

$R_a(x)$  = quantité de rachis infectés actifs au printemps de l'année  $a$  en  $x$ .

**Printemps de l'année  $a$  : production de la quantité de spores  $\nu_a$**

$$\nu_a(x) = \begin{cases} H_a(x) R_a(x) & \text{if } R_a(x) \geq r \\ H_a(x) \frac{R_a(x)^2}{r} & \text{if } R_a(x) < r \end{cases}$$

avec  $H_a(x) = \beta_0 + \beta_1 h_a(x)$  où  $h_a$  est la pluviométrie du printemps de l'année  $a$ .



**Été de l'année  $a$  : dispersion des spores**

$$\frac{\partial w_a(t, x)}{\partial t} = D \Delta w_a(t, x) + \frac{\nu_a(x)}{\tau}$$

avec  $w_a(0, x) = 0$ ,  $\tau = 60$  j. À la fin de l'été, la quantité de spores est  $w_a(\tau, \cdot)$ .

**Automne de l'année  $a$  : infection des rachis**

Les rachis sont actifs durant 2 années

$$\chi_a(x) = (w_a(\tau, x) \wedge S) d(x)$$

$$R_{a+1}(x) = \chi_a(x) + \chi_{a-1}(x)$$

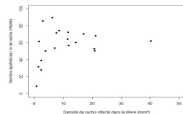
# Modèle de réaction-diffusion pour l'infection des rachis

$R_a(x)$  = quantité de rachis infectés actifs au printemps de l'année  $a$  en  $x$ .

**Printemps de l'année  $a$  : production de la quantité de spores  $\nu_a$**

$$\nu_a(x) = \begin{cases} H_a(x) R_a(x) & \text{if } R_a(x) \geq r \\ H_a(x) \frac{R_a(x)^2}{r} & \text{if } R_a(x) < r \end{cases}$$

avec  $H_a(x) = \beta_0 + \beta_1 h_a(x)$  où  $h_a$  est la pluviométrie du printemps de l'année  $a$ .



**Été de l'année  $a$  : dispersion des spores**

$$\frac{\partial w_a(t, x)}{\partial t} = D \Delta w_a(t, x) + \frac{\nu_a(x)}{\tau}$$

avec  $w_a(0, x) = 0$ ,  $\tau = 60$  j. À la fin de l'été, la quantité de spores est  $w_a(\tau, \cdot)$ .

**Automne de l'année  $a$  : infection des rachis**

Les rachis sont actifs durant 2 années

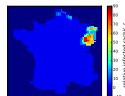
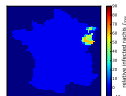
$$\chi_a(x) = (w_a(\tau, x) \wedge S) d(x)$$

$$R_{a+1}(x) = \chi_a(x) + \chi_{a-1}(x)$$

5 paramètres à estimer

+ conditions initiales  $\rightarrow C_{init}$

$$R_{2006} = C_{init} \Psi_{2006}, \quad R_{2007} = C_{init} \Psi_{2007}$$

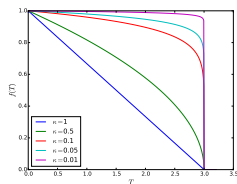


## Modèle pour le développement des symptômes

Un frêne sans symptôme, localisé en  $x$ , développe des symptômes l'année  $a$  si  $Y_a(x) = 1$  où

$$Y_a(x) \sim \text{Ber} \left( \frac{\chi_{a-1}(x) f(T_{a-1}(x))}{r_S} \wedge 1 \right)$$

avec  $f(T) = [1 - T/\gamma]_+^{\kappa}$  et  $T_a(x) = \text{nb de jours dont la temp. max est } > 28^\circ\text{C}$  et  $\gamma$  le nombre de jours  $> 28^\circ\text{C}$  à partir duquel l'infection n'est plus possible.

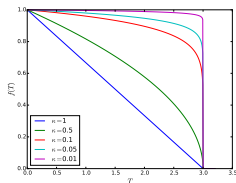


## Modèle pour le développement des symptômes

Un frêne sans symptôme, localisé en  $x$ , développe des symptômes l'année  $a$  si  $Y_a(x) = 1$  où

$$Y_a(x) \sim \text{Ber} \left( \frac{\chi_{a-1}(x) f(T_{a-1}(x))}{r_S} \wedge 1 \right)$$

avec  $f(T) = [1 - T/\gamma]_+^{\kappa}$  et  $T_a(x) = \text{nb de jours dont la temp. max est } > 28^\circ\text{C}$  et  $\gamma$  le nombre de jours  $> 28^\circ\text{C}$  à partir duquel l'infection n'est plus possible.



Le paramètre moyen de la Bernouilli sur le quadrat  $\omega_i$  est

$$\tilde{q}_a^i = \frac{\chi_{a-1}^i f(T_a(x))}{r_S d(i)} \wedge 1$$

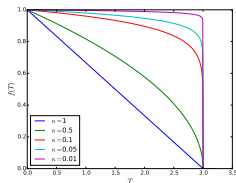
où  $\chi_{a-1}^i = \int_{\omega_i} \chi_{a-1}(x) dx$  est la quantité rachis nouvellement infectés dans  $\omega_i$ .

## Modèle pour le développement des symptômes

Un frêne sans symptôme, localisé en  $x$ , développe des symptômes l'année  $a$  si  $Y_a(x) = 1$  où

$$Y_a(x) \sim \text{Ber} \left( \frac{\chi_{a-1}(x) f(T_{a-1}(x))}{r_S} \wedge 1 \right)$$

avec  $f(T) = [1 - T/\gamma]_+^\alpha$  et  $T_a(x) = \text{nb de jours dont la temp. max est } > 28^\circ\text{C}$  et  $\gamma$  le nombre de jours  $> 28^\circ\text{C}$  à partir duquel l'infection n'est plus possible.



Le paramètre moyen de la Bernouilli sur le quadrat  $\omega_i$  est

$$\tilde{q}_a^i = \frac{\chi_{a-1}^i f(T_a(x))}{r_S d(i)} \wedge 1$$

où  $\chi_{a-1}^i = \int_{\omega_i} \chi_{a-1}(x) dx$  est la quantité rachis nouvellement infectés dans  $\omega_i$ .

Les symptômes persistent d'une année à l'autre avec probabilité  $C_{\text{pers}}$ . La probabilité qu'un arbre du quadrat  $i$  ait des symptômes l'année  $a$  est donc

$$q_a^i = C_{\text{pers}} q_{a-1}^i + \tilde{q}_a^i (1 - C_{\text{pers}} q_{a-1}^i)$$

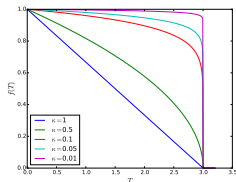


## Modèle pour le développement des symptômes

Un frêne sans symptôme, localisé en  $x$ , développe des symptômes l'année  $a$  si  $Y_a(x) = 1$  où

$$Y_a(x) \sim \text{Ber} \left( \frac{\chi_{a-1}(x) f(T_{a-1}(x))}{r_S} \wedge 1 \right).$$

avec  $f(T) = [1 - T/\gamma]^{\kappa}_+$  et  $T_a(x)$  = nb de jours dont la temp. max est  $> 28^\circ\text{C}$  et  $\gamma$  le nombre de jours  $> 28^\circ\text{C}$  à partir duquel l'infection n'est plus possible.



Le paramètre moyen de la Bernouilli sur le quadrat  $\omega_i$  est

$$\tilde{q}_a^i = \frac{\chi_{a-1}^i f(T_a(x))}{r_S d(i)} \wedge 1$$

où  $\chi_{a-1}^i = \int_{\omega_i} \chi_{a-1}(x) dx$  est la quantité rachis nouvellement infectés dans  $\omega_i$ .

Les symptômes persistent d'une année à l'autre avec probabilité  $C_{\text{pers}}$ . La probabilité qu'un arbre du quadrat  $i$  ait des symptômes l'année  $a$  est donc

$$q_a^i = C_{\text{pers}} q_{a-1}^i + \tilde{q}_a^i (1 - C_{\text{pers}} q_{a-1}^i)$$

4 paramètres à estimer

## Modélisation des observations

À la  $k$ -ième observation de l'année  $a$  :

- ▶  $m = 30$  frênes sont observés dans le quadrat  $i = i_{a,k}$
- ▶ l'observateur "retourne" la proportion  $p_a^k(i)$  d'arbres infectés parmi les  $m$  arbres

**Hypothèse** : les arbres observés sont tous différents  $\rightarrow$  indépendance (sachant  $(\chi_a)_a$ ) des observations

$$\underbrace{m \text{ obs}_a(i) p_a(i)}_{\text{nb arbres infectés observés l'année } a} \sim \text{Bin} \left( \underbrace{m \text{ obs}_a(i)}_{\text{nb arbres observés l'année } a}, q_a^i \right)$$

où  $\text{obs}_a(i)$  est le nombre d'observations du quadrat  $i$  l'année  $a$  et

$p_a(i) = \frac{1}{\text{obs}_a(i)} \sum_{k=1}^{\text{obs}_a(i)} p_a^k(i)$  la proportion d'arbres infectés observés sur le quadrat  $i$  l'année  $a$ .

# Plan

## Modèles

- Modèle de réaction-diffusion pour l'infection des rachis

- Modèle pour le développement des symptômes

- Modélisation des observations

## Inférence statistique et résultats numériques

- Vraisemblance

- Identification des “bonnes” zones de paramètres

- Lois a posteriori

- Évolution de la chararose pour le paramètre qui maximise la vraisemblance

- Comparaison de plusieurs indices de température

La vraisemblance pour  $\theta = (D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{\text{init}}, r_s, C_{\text{pers}})$  est

$$\begin{aligned}\mathcal{L}(\theta) &= \prod_a \prod_i \mathbb{P}(N_a(i) = m_{\text{obs}_a(i)} p_a(i) \mid (q_a^i(\theta))_{a,i}) \\ &= \prod_a \prod_i \binom{m_{\text{obs}_a(i)}}{m_{\text{obs}_a(i)} p_a(i)} (q_a^i(\theta))^{m_{\text{obs}_a(i)} p_a(i)} (1 - q_a^i(\theta))^{m_{\text{obs}_a(i)} (1 - p_a(i))}\end{aligned}$$

La vraisemblance pour  $\theta = (D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{\text{init}}, r_S, C_{\text{pers}})$  est

$$\begin{aligned}\mathcal{L}(\theta) &= \prod_a \prod_i \mathbb{P}(N_a(i) = m_{\text{obs}_a(i)} p_a(i) \mid (q_a^i(\theta))_{a,i}) \\ &= \prod_a \prod_i \binom{m_{\text{obs}_a(i)}}{m_{\text{obs}_a(i)} p_a(i)} (q_a^i(\theta))^{m_{\text{obs}_a(i)} p_a(i)} (1 - q_a^i(\theta))^{m_{\text{obs}_a(i)} (1 - p_a(i))}\end{aligned}$$

**Problème non identifiable** :  $\mathcal{L}(\theta) = \mathcal{L}(\tilde{\theta})$

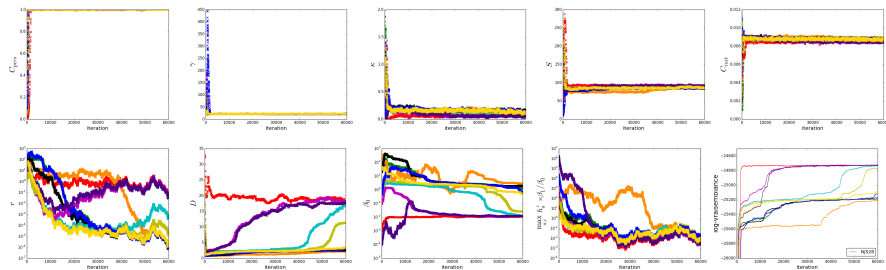
pour  $\theta = (D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{\text{init}}, r_S, C_{\text{pers}})$

et  $\tilde{\theta} = (D, \beta_0, \beta_1, \alpha r, \gamma, \kappa, \alpha S, \alpha C_{\text{init}}, \alpha r_S, C_{\text{pers}})$

$\implies$  on fixe  $r_S = 1000$

# Première estimation grossière des paramètres et réduction de modèle

## Étape 1 : convergence des paramètres pour un algorithme de Metropolis-Hasting



$$r_S \rightarrow 1000$$

$$C_{pers} \rightarrow 1$$

$$\gamma \approx 23$$

$$\kappa \in [0, 0.2]$$

$$S \in [80, 95]$$

$$C_{init} \approx 0.0087$$

$$r \rightarrow 0?$$

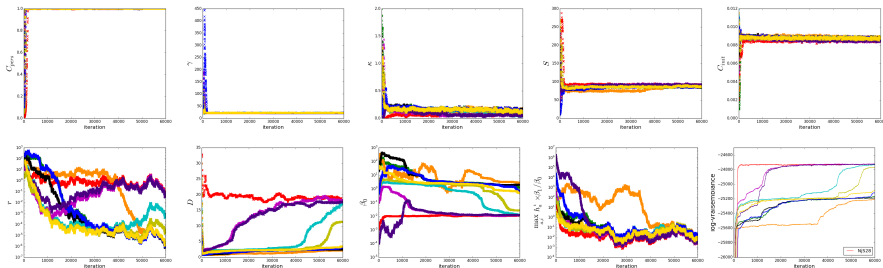
$$\beta_1 \rightarrow 0$$

$$D?$$

$$\beta_0?$$

# Première estimation grossière des paramètres et réduction de modèle

## Étape 1 : convergence des paramètres pour un algorithme de Metropolis-Hasting



$r_S \rightarrow 1000$   
 $C_{pers} \rightarrow 1$

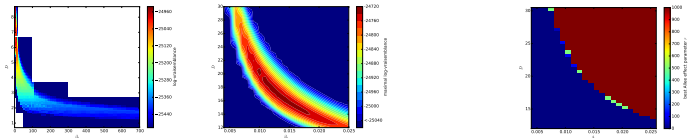
$\gamma \approx 23$   
 $\kappa \in [0, 0.2]$

$S \in [80, 95]$   
 $C_{init} \approx 0.0087$

$r \rightarrow 0?$   
 $\beta_1 \rightarrow 0$

$D?$   
 $\beta_0?$

## Étape 2 : calcul de la vraisemblance des paramètres sur une grille grossière

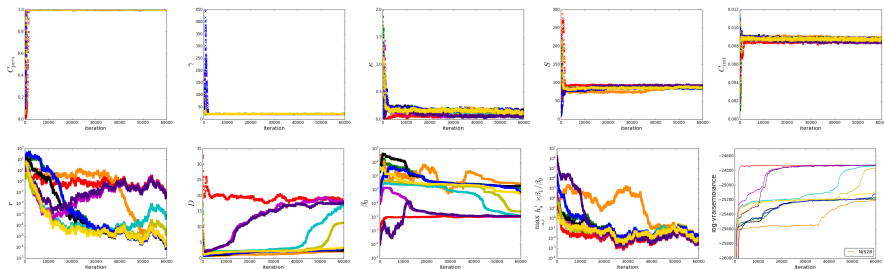


$$(\beta_0, D) \mapsto \max_{r; \gamma; \kappa; S; C_{init}} \mathcal{L}(D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{init}, r_S, C_{pers})$$

$$\operatorname{argmax}_r \rightarrow r \approx 0$$

# Première estimation grossière des paramètres et réduction de modèle

## Étape 1 : convergence des paramètres pour un algorithme de Metropolis-Hasting



$r_S \rightarrow 1000$   
 $C_{pers} \rightarrow 1$

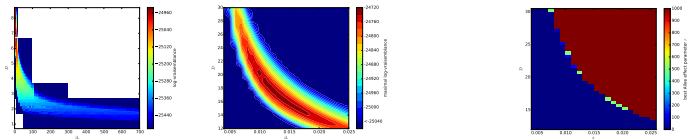
$\gamma \approx 23$   
 $\kappa \in [0, 0.2]$

$S \in [80, 95]$   
 $C_{init} \approx 0.0087$

$r \rightarrow 0?$   
 $\beta_1 \rightarrow 0$

$D?$   
 $\beta_0?$

## Étape 2 : calcul de la vraisemblance des paramètres sur une grille grossière



$$(\beta_0, D) \mapsto \max_{r; \gamma; \kappa; S; C_{init}} \mathcal{L}(D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{init}, r_S, C_{pers})$$

$$\operatorname{argmax}_r \rightarrow r \approx 0$$

À estimer :  $\beta_0, D, \gamma, \kappa, S, C_{init}$



# Adaptive Multiple Importance Sampling Algorithm (AMIS)

Cornuet, Marin, Mira, Robert (2012)

initialization of  $m_1$  and  $\Sigma_1$

**for**  $k = 1 \dots N_L$  **do**

$\theta_{li} \sim \mathcal{N}(m_k, \sigma_k), i = 1 \dots N$  {new parameter sample}

$\tilde{w}_{li} \leftarrow \frac{\text{vrai*prior}(\theta_{li})}{\frac{1}{k} \sum_{\ell=1}^k g_{m_\ell, \Sigma_\ell}(\theta_{li})} \quad l = 1 \dots k, i = 1 \dots N$  {weights computation}

$w_{li} \leftarrow \tilde{w}_{li} / \sum_{\ell=1}^k \sum_{i=1}^N \tilde{w}_{\ell i}$  {weights normalization}

$m_{k+1} \leftarrow \sum_{l=1}^k \sum_{i=1}^N w_{li} \theta_{li}$

$\Sigma_{k+1} \leftarrow \sum_{l=1}^k \sum_{i=1}^N w_{li} (\theta_{li} - m_l) (\theta_{li} - m_l)^t$

**end for**

Idées de l'algorithme :

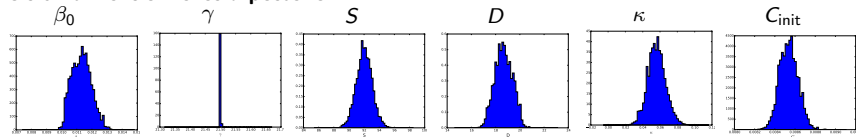
1. On se donne une loi de proposition initiale gaussienne
2. On tire  $N$  paramètres suivant cette loi
3. On calcule les poids de ces paramètres
4. On modifie la loi de proposition en fonction de ces poids
5. On recommence (en mettant à jour les poids des anciens paramètres)

Avantages :

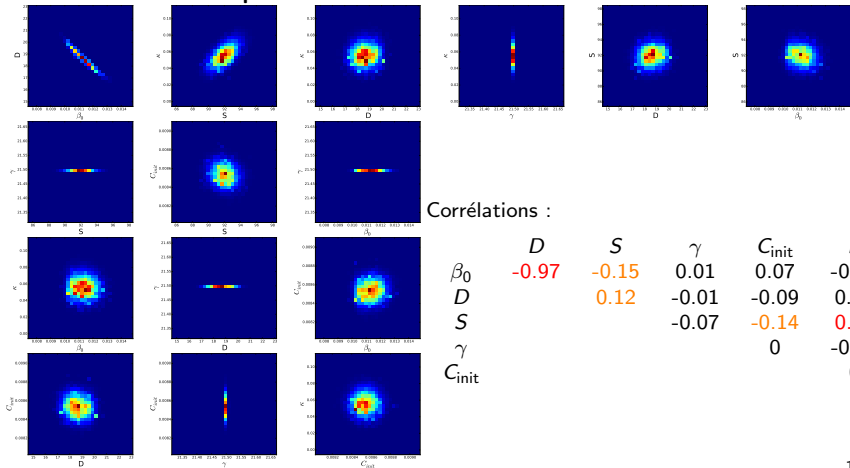
- ▶ Permet de calculer la vraisemblance de  $N = 10000$  paramètres en parallèle
- ▶ Converge relativement vite ( $N_L$  entre 50 et 150)

# Lois a posteriori

## Lois unidimensionnelles a posteriori



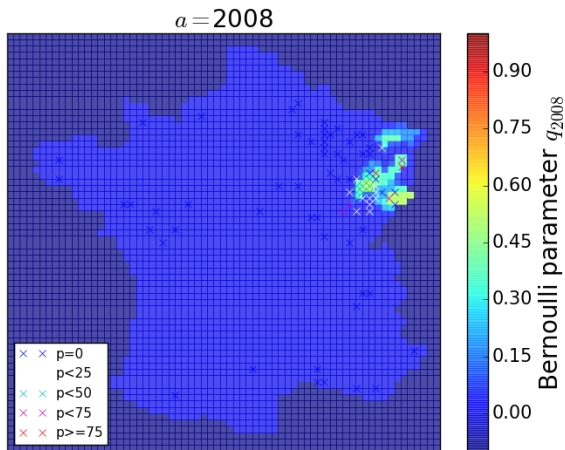
## Lois bidimensionnelles a posteriori



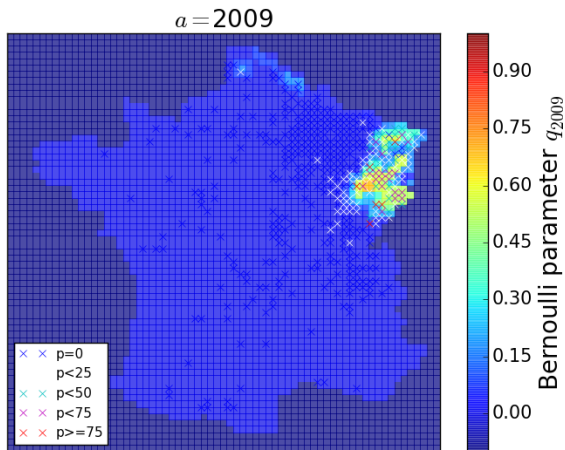
Corrélations :

	$D$	$S$	$\gamma$	$C_{init}$	$\kappa$
$\beta_0$	-0.97	-0.15	0.01	0.07	-0.03
$D$		0.12	-0.01	-0.09	0.06
$S$			-0.07	-0.14	0.58
$\gamma$				0	-0.03
$C_{init}$					0

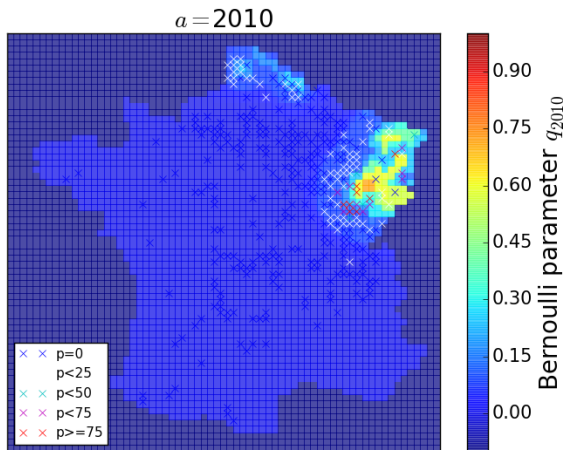
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



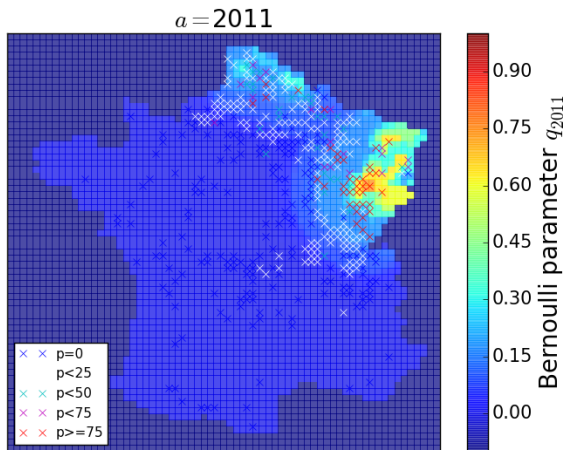
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



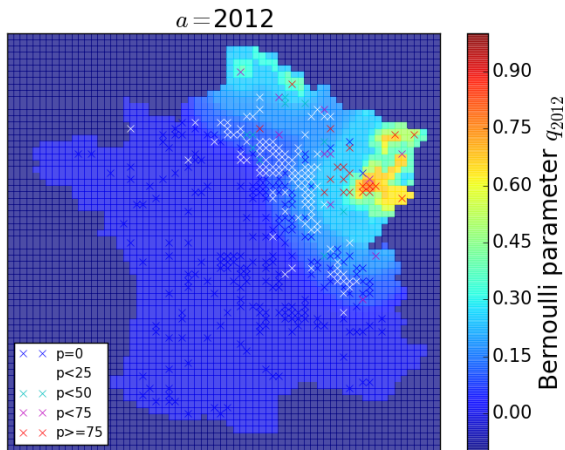
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



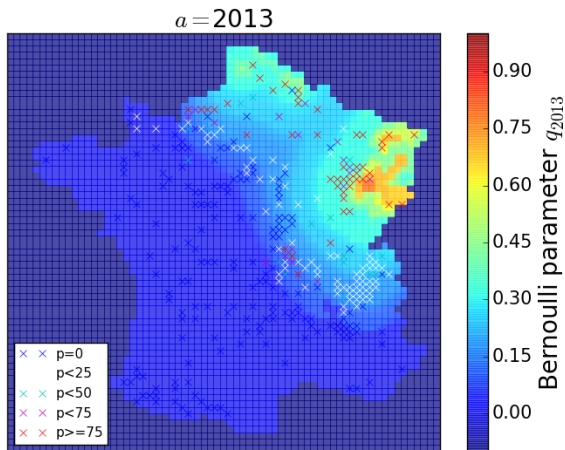
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance

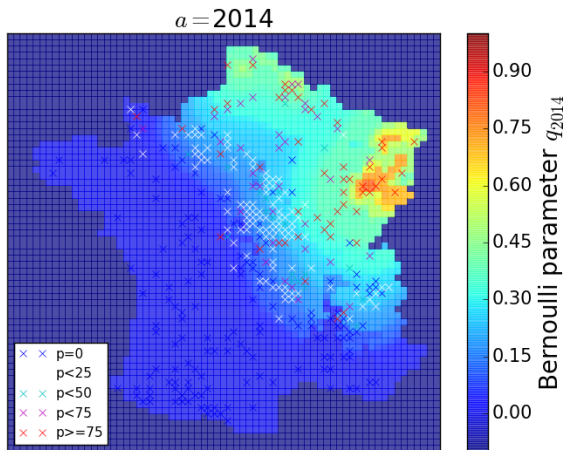


# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance

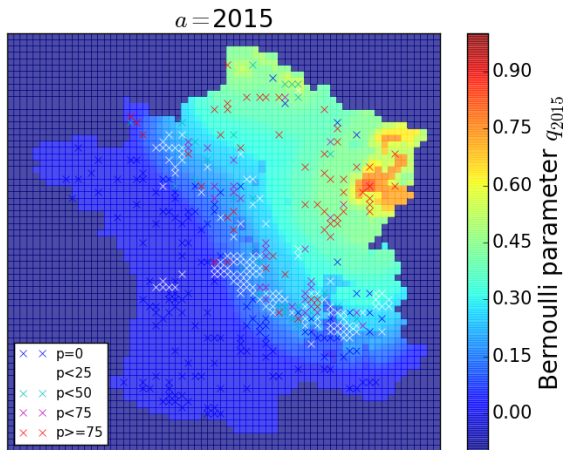




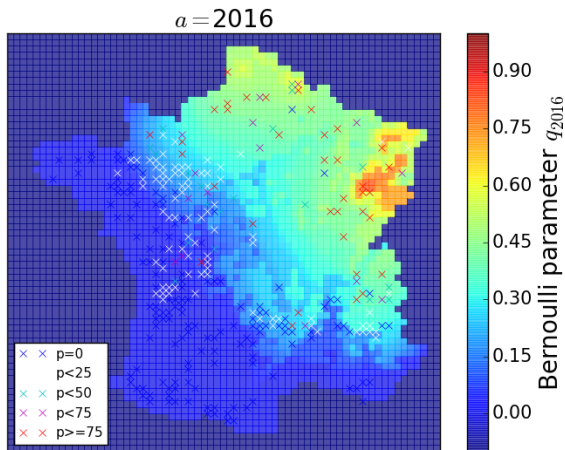
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



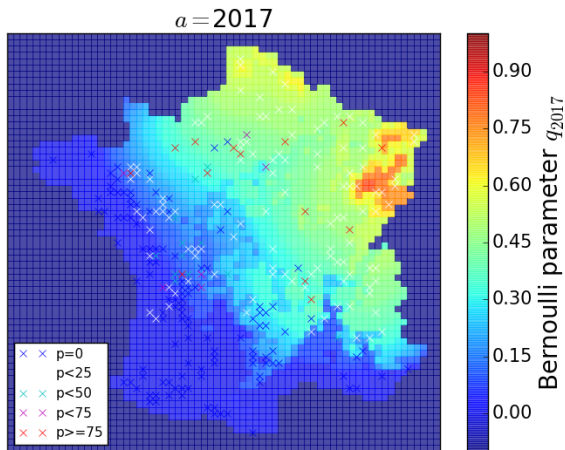
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



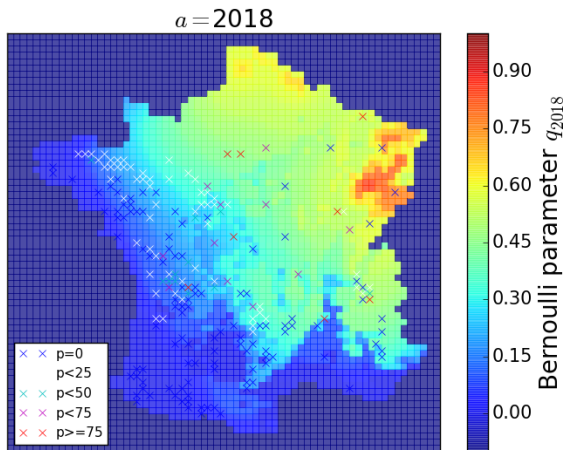
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



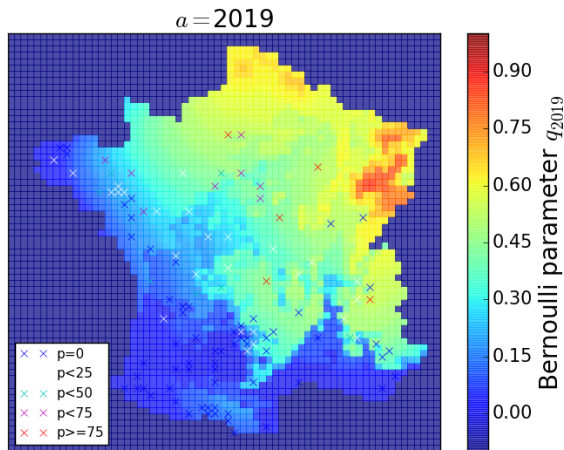
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



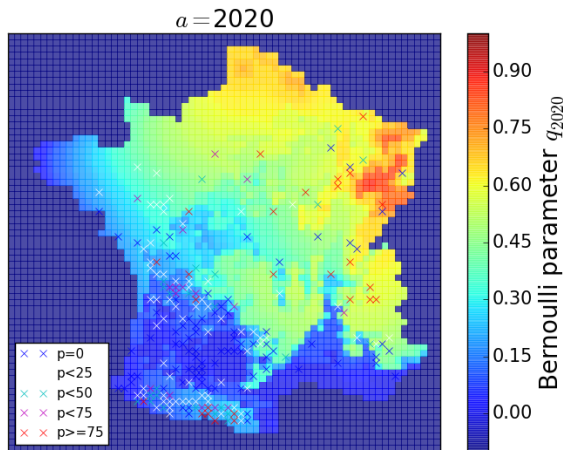
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance

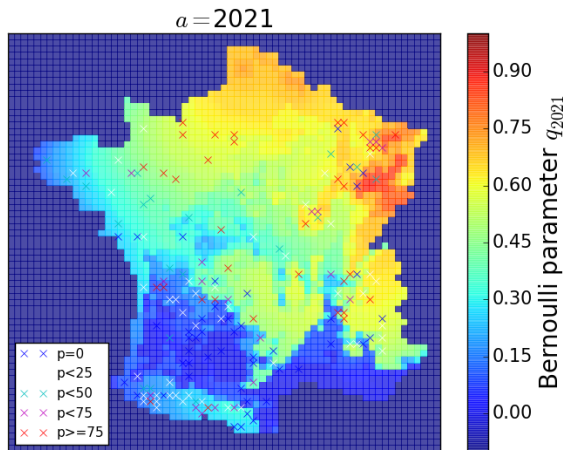


# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



Données 2020 non utilisées pour l'estimation des paramètres

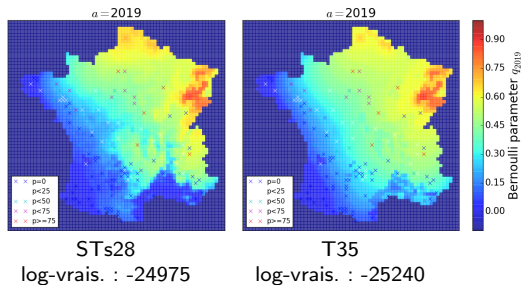
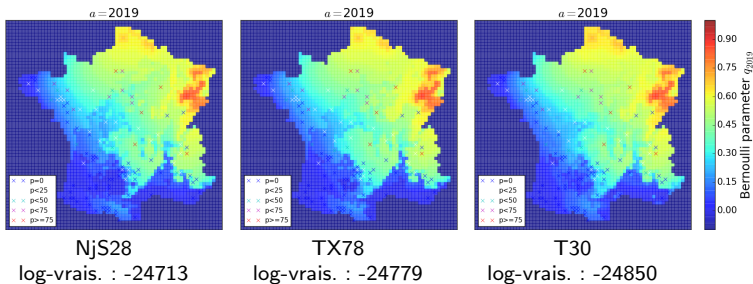
# Évolution de la chalarose pour le paramètre qui maximise la vraisemblance



Données 2021 non utilisées pour l'estimation des paramètres



# Comparaison des dynamiques de plusieurs indices de température



# Conclusion

Un modèle qui reproduit

- ▶ le front de propagation
- ▶ le “blocage” de l'épidémie dans le sud
- ▶ le “saut” de la Garonne

Réponses à certaines questions initiales de Benoît et Marie :

- ▶ passage dans les Pyrénées
- ▶ indice de température : nombres de jour au dessus de  $28^{\circ}\text{C}$
- ▶ hétérogénéité de la pluviométrie négligeable
- ▶ effet Allee négligeable

Pour la suite :

- ▶ vitesse de propagation du front → collaboration avec Thomas Giletti