

# MODÉLISATION SPATIO-TEMPORELLE DE LA CHALAROSE (MALADIE FONGIQUE DU FRÊNE) EN FRANCE

Coralie Fritsch <sup>1</sup> & Anne Gégout-Petit <sup>2</sup> & Benoit Marçais <sup>3</sup> & Maris Grosdidier <sup>4</sup>

<sup>1</sup> *IECL, Université de Lorraine, CNRS, Inria EPI Tosca, coralie.fritsch@inria.fr*

<sup>2</sup> *IECL, Université de Lorraine, CNRS, Inria EPI Bigs,  
anne.gegout-petit@univ-lorraine.fr*

<sup>3</sup> *INRA Nancy-Lorraine, IaM, benoit.marçais@inra.fr*

<sup>4</sup> *INRA Avignon, BioSP, marie.grosdidier@inra.fr*

**Résumé.** La chalarose est une maladie du frêne apparue en Pologne en 1992 et observée pour la première fois en France en 2008. La maladie est causée par un champignon pathogène qui se développe sur les rachis infectés tombés aux pieds des arbres durant l'automne. Les champignons libèrent des spores qui se dispersent durant l'été et infectent ainsi les arbres voisins provoquant la propagation de la maladie. Depuis 2008, environ 500 sites de forêt sont visités chaque année et nous disposons de la proportion d'arbres infectés parmi ceux inspectés pour chaque visite. Basé sur nos connaissances sur le cycle de la maladie et afin de mieux comprendre la propagation de la maladie en France dans les prochaines années, nous avons développé un modèle mécaniste spatio-temporel décrivant la propagation de la maladie. Ce modèle est basé sur un modèle paramétrique latent représentant l'infection des rachis et tenant compte des effets de l'humidité et de la température ainsi que d'un modèle de réaction-diffusion décrivant la propagation des spores. Pour l'inférence des paramètres, uniquement basée sur la proportion d'arbres infectés, nous avons utilisé une approche Bayésienne et des simulations MCMC.

**Mots-clés.** Modèle mécaniste spatio-temporel, méthodes bayésiennes

**Abstract.** Chalara fraxinea causes an illness of the tree ash that appears in Europe by Poland in 1992 and was observed for the first time in France in 2008. The mechanism of the illness is due to a fungus that grows on ash rachis in the previous year's fallen leaves. The spores are produced and are transmitted by wind. From 2008 to now, around 500 places of forest are visited every year in the purpose to notice the proportion of infected ashes. According to our knowledge about chalara fraxinea and in the purpose to a better understanding of the effect of covariates on the spread of the illness, we have built a spatio-temporal mechanistic model of propagation of the illness. It is based on a latent parametric model of spores production accounting on the effect of humidity and temperature and a reaction-diffusion model for the spores diffusion. For the inference, only based on the proportion of infected trees, we have used a Bayesian framework and MCMC simulations.

**Keywords.** spatio-temporal mechanistic model, bayesian approaches

# 1 Modèle mécaniste spatio-temporel pour l'infection des rachis

Le cycle de l'infection des rachis se fait en trois étapes : la production de spores par les rachis infectés durant le printemps, la dispersion des spores durant l'été et l'infection des rachis par les spores durant l'hiver.

Nous notons  $R_a(x)$  la quantité de rachis infectés produit durant l'hiver de l'année  $a - 1$  à l'année  $a$  à la position  $x$ .

## Printemps de l'année $a$ : production de spores

Les rachis infectés sont actifs durant deux années. La quantité de rachis infectés produisant des spores durant le printemps de l'année  $a$  est donc  $R_a(x) + R_{a-1}(x)$ . Le développement des champignons dépend de l'humidité. La production de spores dépend donc elle aussi de l'humidité et est soumise à un effet Allee sous une quantité  $r$  de rachis infectés. La quantité de spores  $\nu_a(x)$  produite en  $x$  durant le printemps de l'année  $a$  est donc donnée par

$$\nu_a(x) = \begin{cases} H_a(x) (R_a(x) + R_{a-1}(x)) & \text{si } R_a(x) + R_{a-1}(x) \geq r \\ H_a(x) \frac{(R_a(x) + R_{a-1}(x))^2}{r} & \text{si } R_a(x) + R_{a-1}(x) < r \end{cases}$$

où  $H_a(x) = \beta_0 + \beta_1 h_a(x)$  est une fonction linéaire de la pluviométrie  $h_a(x)$  durant le printemps de l'année  $a$  en  $x$ .

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

La quantité de spores  $\nu_a(x)$  produite aux pieds des arbres durant le printemps est dispersée durant l'été. Nous admettons que la dispersion peut être modélisée par une diffusion. La quantité de spores  $w(t, x)$  tombée sur les arbres en  $x$  au temps  $t$  est ainsi solution de

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

où  $w(0, x) = 0$ , la période de diffusion (de juillet à août)  $\tau$  vaut 60 jours et  $D$  est le coefficient de diffusion. À la fin de la période de diffusion, la quantité de spores est donc  $w(\tau, \cdot)$ .

## Hiver de l'année $a$ à l'année $a + 1$ : infection des rachis

Les spores dispersés sur les arbres durant l'été infectent ensuite de nouveaux rachis durant l'hiver. La production de ces rachis infectés dépend de la densité  $d$  de frênes mais également de la température durant l'été. En effet, les températures au dessus de  $35^\circ\text{C}$  réduisent l'activité des spores. On suppose que la dépendance de la quantité de rachis infectés produite  $R_{a+1}$ , durant l'hiver de l'année  $a$  à l'année  $a + 1$ , est linéaire en la quantité

de spores avec un seuil de saturation  $S$  :

$$R_{a+1} = (w(\tau, x) \wedge S) d(x) f(T_a(x))$$

avec  $f(T) = [1 - T/\gamma]_+$ , où  $T_a(x)$  est le nombre de jours de l'été pour lesquels la température est dessus de  $35^\circ\text{C}$  et  $\gamma$  est le nombre de jours au dessus de  $35^\circ\text{C}$  à partir duquel l'infection n'est plus possible (mort des spores).

## 2 Modèle binomial pour le développement des symptômes

Le développement des symptômes de flétrissement des arbres est dû aux rachis infectés aux pieds des arbres. Ces symptômes sont observables seulement deux ans après l'infection des rachis dans le sens suivant:

- hiver de l'année  $a$  à l'année  $a + 1$  : infection des rachis (quantité  $R_{a+1}$ )
- année  $a + 1$  : infection des arbres par les rachis infectés  $R_a + R_{a+1}$
- année  $a + 2$  : observation des symptômes de flétrissement

Nous supposons qu'un arbre en  $x$  développe des symptômes, observables l'année  $a$ , avec une probabilité dépendant linéairement de la quantité de rachis infectés  $R_{a-1}(x) + R_{a-2}(x)$  au pied de l'arbre, avec un seuil à 0.95. Le développement des symptômes par un arbre en  $x$  l'année  $a$  est donc modélisé par la loi de Bernoulli suivante

$$Y_a(x) \sim Ber \left( \frac{R_{a-1}(x) + R_{a-2}(x)}{r_S} \wedge 0.95 \right).$$

Les données nous sont fournies sur la France divisée en quadrats de dimensions  $\rho = 16 \times 16 \text{ km}^2$ . La densité d'arbres est supposée uniforme sur chaque quadrat, la probabilité de développement des symptômes par un arbre dans un quadrat  $\omega_i$  est ainsi

$$q_a^i = \frac{R_{a-1}^i + R_{a-2}^i}{r_S d(i)} \wedge 0.95$$

où  $R_a^i$  est la quantité totale de rachis infectés dans le quadrat  $\omega_i$  et la densité de frênes  $d(i)$  dans le quadrat  $\omega_i$  intervenant dans la renormalisation est due au fait que les rachis infectés sont produits par les arbres. La quantité moyenne de rachis infectés au pied des arbres est donc la quantité totale de rachis infectés dans le quadrat renormalisée par le nombre d'arbres dans le quadrat.

Le nombre d'arbres infectés  $\delta_a(i) d(i) \rho$  dans le quadrat  $\omega_i$  est ainsi déterminé par une loi binomiale :

$$\delta_a(i) d(i) \rho \sim Bin(d(i) \rho, q_a^i) \tag{1}$$

où  $\delta_a(i)$  est la proportion d'arbres infectés dans le quadrat  $\omega_i$ .

### 3 Modèle statistique pour le processus d'observations

À chaque observation, une trentaine d'arbres sont observés dans un quadrat et l'observateur retourne la proportion d'arbres infectés (frênes avec symptômes de flétrissement) parmi ceux observés.

Durant la  $k$ -ième observation de l'année  $a$ ,  $m = 30$  arbres parmi les  $d(i_k) \rho$  arbres du quadrat  $\omega_{i_k}$  sont observés. La proportion d'arbres infectés dans le quadrat est  $\delta_a(i_k)$ . Le nombre d'arbres infectés  $m p_a^k(i_k)$  parmi les  $m$  arbres observés est donc une variable hypergéométrique:

$$m p_a^k(i) \sim H(m, \delta_a(i_k), d(i_k) \rho)$$

où  $p_a^k(i_k)$  est la proportion d'arbres infectés retournée par l'observateur.

On suppose que si deux observations sont faites la même année dans un quadrat, les arbres de la seconde observation sont tous différents de ceux de la première observation. Si on note  $\text{obs}_a(i)$  le nombre d'observations du  $i$ -ième quadrat durant l'année  $a$ , alors

$$m \text{obs}_a(i) p_a(i) \sim H(m \text{obs}_a(i), \delta_a(i), d(i) \rho)$$

où la proportion d'arbres infectés  $p_a(i)$  parmi les  $m \text{obs}_a(i)$  arbres observés l'année  $a$  est donnée par

$$p_a(i) = \frac{1}{\text{obs}_a(i)} \sum_{k=1}^{\text{obs}_a(i)} p_a^k(i).$$

Le nombre d'arbres observés  $m \text{obs}_a(i)$  étant relativement petit comparé au nombre d'arbres  $d(i) \rho$  dans chaque quadrat, la loi hypergéométrique ci-dessus peut être approchée par une loi binomiale:

$$m \text{obs}_a(i) p_a(i) \sim \text{Bin}(m \text{obs}_a(i), \delta_a(i)). \quad (2)$$

### 4 Inférence statistique

La vraisemblance des observations sachant un ensemble de paramètres  $\theta = (D, \beta_0, \beta_1, r, r_S, \gamma, S)$  est donnée par

$$\mathcal{L}(\theta) = \prod_a \prod_i \mathbb{P}(N_a(i) = m \text{obs}_a(i) p_a(i) \mid (R_a^\theta(x))_{a,x})$$

où  $(R_a^\theta(x))_{a,x}$  est la dynamique des rachis infectés donnée par le modèle mécaniste de la Section 1 avec l'ensemble de paramètres  $\theta$ . D'après (1) et (2), on obtient ainsi

$$\begin{aligned} \mathcal{L}(\theta) &= \prod_a \prod_i \sum_{k=0}^{d(i)\rho} \left\{ \mathbb{P}(N_a(i) = m \text{ obs}_a(i) p_a(i) \mid \delta_a(i) d(i) \rho = k) \right. \\ &\quad \left. \times \mathbb{P}(\delta_a(i) d(i) \rho = k \mid (R_a^\theta(x))_{a,x}) \right\} \\ &= \prod_a \prod_i \sum_{k=0}^{d(i)\rho} \left\{ \binom{m \text{ obs}_a(i)}{m \text{ obs}_a(i) p_a(i)} \left( \frac{k}{d(i)\rho} \right)^{m \text{ obs}_a(i) p_a(i)} \left( 1 - \frac{k}{d(i)\rho} \right)^{m \text{ obs}_a(i) (1-p_a(i))} \right. \\ &\quad \left. \times \binom{d(i)\rho}{k} (q_a^i)^k (1 - q_a^i)^{d(i)\rho - k} \right\} \end{aligned}$$

Nous utilisons des simulations MCMC pour inférer les paramètres. Ces méthodes, inspirées de Roques et al (2011) et Abboud et al. (2018) ainsi que les résultats seront présentés lors de l'exposé.

## Bibliographie

Abboud, C., Bonnefon, O., Parent, E., and Soubeyrand, S. (2018). Dating and localizing an invasion from post-introduction data and a coupled reaction-diffusion- absorption model., *ArXiv Mathematics e-prints*, arXiv/1808.00868 [q-bio.PE].

Roques, L., Soubeyrand, S., and Rousselet, J. (2011). A statistical-reaction-diffusion approach for analyzing expansion processes, *Journal of Theoretical Biology*, 274(1):43 – 51.