



**HAL**  
open science

# Mechanistic-statistical model for the expansion of ash dieback

Coralie Fritsch, Marie Grosdidier, Anne Gégout-Petit, Benoit Marçais

► **To cite this version:**

Coralie Fritsch, Marie Grosdidier, Anne Gégout-Petit, Benoit Marçais. Mechanistic-statistical model for the expansion of ash dieback. 2024. hal-04690647

**HAL Id: hal-04690647**

**<https://hal.science/hal-04690647>**

Preprint submitted on 6 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Mechanistic-statistical model for the expansion of ash dieback

Coralie Fritsch<sup>1</sup>, Marie Grosdidier<sup>2</sup>, Anne Gégout-Petit<sup>1</sup>, Benoît Marçais<sup>3</sup>

6th September 2024

## Abstract

*Hymenoscyphus fraxineus* is an invasive forest fungal pathogen that induces severe dieback in European ash populations. The spread of the disease has been closely monitored in France by the forest health survey system. We have developed a mechanistic-statistical model that describes the spread of the disease. It takes into account climate (summer temperature and spring rainfall), pathogen population dynamics (foliar infection, Allee effect induced by limited sexual partner encounters) and host density. We fitted this model using available disease reports. We estimated the parameters of our model, first identifying the appropriate ranges for the parameters, which led to a model reduction, and then using an adaptive multiple importance sampling algorithm for fitting. The model reproduces well the propagation observed in France over the last 20 years. In particular, it predicts the absence of disease impact in the south-east of the country and its weak development in the Garonne valley in south-west France. Summer temperature is the factor with the highest overall effect on disease spread, and explains the limited impact in southern France. Among the different temperature indices tested, the number of summer days with temperatures above 28°C gave the best qualitative behavior and the best fit. In contrast, the Allee effect and the heterogeneity of spring precipitation did not strongly affect the overall expansion of *H. fraxineus* in France and could be neglected in the modeling process. The model can be used to infer the average annual dispersal of *H. fraxineus* in France.

**Keywords:** *Hymenoscyphus fraxineus*; Ash dieback; forest disease; invasive pathogen, reaction-diffusion equation; AMIS algorithm, Bayesian inference

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data description and assumptions</b>	<b>4</b>
2.1	Disease prevalence . . . . .	4
2.2	Ash density and meteorological data . . . . .	5

---

<sup>1</sup>Université de Lorraine, CNRS, Inria, IECL, F-54000 Nancy, France

<sup>2</sup>INRAE, BioSP, 84000 Avignon, France

<sup>3</sup>Université de Lorraine, INRAE, UMR1136 Interactions Arbres/Microorganismes, F-54000 Nancy, France

<b>3</b>	<b>Model description</b>	<b>5</b>
3.1	Reaction-diffusion model for the leaf infection . . . . .	5
3.2	Model for the symptoms development . . . . .	8
3.3	Statistical model for the observation data . . . . .	9
<b>4</b>	<b>Statistical Inference</b>	<b>10</b>
4.1	Parameters . . . . .	10
4.2	Likelihood . . . . .	11
4.3	Bayesian method for the parameters estimation . . . . .	11
4.3.1	First estimation of the parameters: model reduction and parameter initialization . . . . .	11
4.3.2	Adaptive Multiple Importance Sampling Algorithm . . . . .	12
<b>5</b>	<b>Numerical Results for the best temperature index (T28)</b>	<b>13</b>
5.1	Posterior distributions of parameters . . . . .	14
5.2	Predicted expansion of ash dieback for the best set of parameters . . . . .	15
5.3	Model validation . . . . .	17
<b>6</b>	<b>Discussion</b>	<b>20</b>
<b>A</b>	<b>First rough estimation and model reduction</b>	<b>22</b>
A.1	A Metropolis-Hasting algorithm with iteration dependent proposal distribution	23
A.2	Log-likelihood on a coarse grid . . . . .	25
A.3	An almost degenerated posterior distribution for $\gamma$ . . . . .	27
<b>B</b>	<b>Convergence of the AMIS algorithm</b>	<b>28</b>
<b>C</b>	<b>Model behavior for different temperature indices</b>	<b>28</b>
<b>D</b>	<b>Data visualisation</b>	<b>31</b>

# 1 Introduction

Biological invasions significantly impact ecosystems, with both major ecological and economic impact (Mollot et al., 2017; Pyšek et al., 2020). In particular, invasive pathogenic microorganisms are important threats for forest ecosystems and the problem has increased over the last decades (Santini et al., 2013; Desprez-Loustau et al., 2016; Eschen et al., 2023; Ghelardini et al., 2016). This has been associated with the globalization trends, and in particular intensified international trade and increased movement of ornamental plants worldwide (Santini et al., 2013; Sikes et al., 2018). Evaluating the potential distribution of an invasive forest pathogen is necessary to assess the threat it poses to a particular area (Paap et al., 2022). This represents a challenge for organisms that are still spreading and therefore do not occupy all the areas in which they could develop (Bebber, 2015). Consequently, it is difficult to understand the factors that shape the niche for invasive organisms, as only the realized

niche, which is constrained by dispersal, is available, while the potential niche remains unknown. A good knowledge of the dispersal of invasive forest pathogens is therefore necessary to properly assess their impact and also to design adequate management strategies (Hudgins et al., 2017). Dispersal is strongly affected by landscape heterogeneity, both in terms of host availability and climate. Indeed, host density is a key feature in explaining disease emergence and spread (Ward et al., 2022; Keesing et al., 2010). To properly manage invasive pathogens, it is therefore essential to assess how dispersion, host distribution and climate jointly shape their impact (Meentemeyer et al., 2012). Ash dieback is a typical example of a severe epidemic induced by an invasive fungal pathogen. The causal agent, *Hymenoscyphus fraxineus* (anamorph *Chalara fraxinea*), was first reported in the Baltic region at the end of the 20th century (Kowalski, 2006; Agan et al., 2022). The disease has spread to western and southern Europe over the last 30 years, causing severe dieback in affected ash stands (Coker et al., 2019; Marçais et al., 2022).

The life cycle of *H. fraxineus* takes place mainly on ash leaves, which are infected by ascospores in late spring and summer (Gross et al., 2014). Leaves remain symptomless for most of the summer (Cross et al., 2017) until late August, when necrotic lesions develop. After leaf fall, the pathogen survives in forest litter on the rachis, which encompasses the petiole and the midrib of the compound leaf (Gross et al., 2014). It produces pseudosclerotial plates on colonized rachises, which are melanized tissues offering protection against desiccation and microbial competitors. *H. fraxineus* is able to survive and produce apothecia on pseudosclerotial rachises for several years (Gross and Holdenrieder, 2013; Kirisits, 2015; Laubray et al., 2024). The majority of inoculum is produced by fruiting bodies, i.e. apothecia, which develop on pseudosclerotial rachis, although apothecia can occasionally be observed on other tissues infected by *H. fraxineus* such as shoots or the collar of young trees (Kowalski and Holdenrieder, 2009; Kirisits and Freinschlag, 2012; Baxter et al., 2023). Asexual spores, i.e. conidia, are considered to have a limited role in disease dispersal (Gross et al., 2014; Marçais et al., 2022). Their main role is that of spermatia, *H. fraxineus* being a heterothallic fungus (Gross et al., 2012). The encounter of sexual partners on the forest floor can be a limiting factor at low ash densities or low infection levels, leading to a significant Allee effect (Laubray et al., 2023). Apothecia formation takes place from spring to midsummer, depending on location in Europe. Dispersal is by airborne ascospores, with an average dispersal distance of around 1 to 2 km (Timmermann et al., 2011; Grosdidier et al., 2018a). Apothecia and ascospore production, as well as the infection process, require humid conditions and are favored by rainy weather during the apothecia production period (Dvořák et al., 2016, 2023; Havrdová et al., 2017; Chumanová et al., 2019). Ash dieback is caused by colonization of twigs from infected leaves in late summer and autumn (Gross et al., 2014). At this stage of the disease, the limiting environmental condition is temperature. *H. fraxineus* survives poorly at temperatures of 36°C and above (Hauptman et al., 2013). Consequently, transfer of the pathogen from leaves to shoots may be limited after hot summers, and ash dieback is of limited severity in the Mediterranean climate of southeastern France (Grosdidier et al., 2018b). Nevertheless, *H. fraxineus* can infect and reproduce on ash leaves without inducing significant shoot mortality (Marçais et al., 2023). Infected ash trees then behave as healthy carriers. This occurs under specific micro/meso climatic conditions, notably after summer heat waves, on isolated trees or in hedgerows (Grosdidier et al., 2020).

In this article, we aim to develop a spatio-temporal model describing the expansion of *H. fraxineus*, taking into account climate data (in time and position) and ash tree density. By estimating the model parameters, we can measure the impact of local conditions on epidemic expansion. In particular, we try to identify the temperature index that gives the best qualitative and quantitative results among a set of temperature indices related to the intensity of heat during the summer.

We are developing a mechanistic-statistical model comprising three nested sub-models: (1) a reaction-diffusion model describing the expansion of leaf infection by *H. fraxineus* and the subsequent colonization of leaf debris in the litter (rachis); (2) a stochastic model describing the development of dieback symptoms as a function of leaf spore infection; (3) a stochastic model describing observation data as a function of symptoms and quadrats visited. We use a Bayesian framework to estimate the model parameters. The model and statistical inference methods we use are consistent with previous work by Roques et al. (2011) on pine processionary moth expansion processes, and by Abboud et al. (2019, 2023) on the invasion of a phytopathogenic bacterium.

In Section 2, we describe the data used to fit the model while in Section 3, we describe the three sub-models of our mechanistic-statistical model. Section 4 is devoted to the statistical inference method used for parameter estimation. We first proceed to a rough estimation of the parameters leading to a model reduction (more technical details about this step are given in Appendix). Then we use an adaptive multiple importance sampling algorithm to estimate the posterior distribution of the parameters of the reduced model. This method is repeated for different temperature indices. Section 5 presents the numerical results (parameters estimation and model dynamics) only for the best temperature index (the comparison and discussion of the different temperature indices is postponed in Appendix). The model validation is also discussed in this section. Finally, we discuss in Section 6 the biological information provided by the parameter estimation, the temperature index selection as well as the dynamics of this model.

## 2 Data description and assumptions

### 2.1 Disease prevalence

Since the first report of ash dieback in France in 2008 (Ioos et al., 2009), the disease progress has been tightly monitored by the forest health survey system (Département de la Santé des Forêts, DSF). The DSF implements a database which stores records of health problems observed in France by a network of foresters trained for the diagnosis of damages induced by abiotic, entomological or microbial causes. Altogether, from 2008 to 2023, about 6000 health problems were recorded on ash, with numbers ranging from 100 to 750 per year. Each record include a date, a geographic location, a host, a list of observed problems and an associated prevalence as well as a brief description of the affected ecosystem. A specific effort was done by the DSF to monitor ash dieback (so called observation strategy), with, in particular, reports of ash dieback absence. We collapsed reports made the same year on the same location and removed reports made in 2021-23 on emerald ash borer (*Agrilus*

*planipennis*) absence as it wasn't possible to conclude for them on *H. fraxineus* presence / absence. Altogether we have 2840 reports of ash dieback presence, 2428 reports of ash dieback absence and 743 reports of a problem different without any reference to ash dieback. We assumed that ash dieback was absent in these last reports, as is not reported and the reports are usually far from any previous report of ash dieback (median distance of 50 km). The symptoms induced by *H. fraxineus* are no very specific and a reliable diagnostic needs confirmation in the laboratory. Thus, new reports of ash dieback in area previously disease-free were usually confirmed by a qPCR test in laboratory, using the protocol of Ioos et al. (2009) (69% of the case in a previously disease-free  $16 \times 16$  km<sup>2</sup> quadrat). Figure 16 show the infection data from 2008 to 2023. Data are grouped by  $16 \times 16$  km<sup>2</sup> quadrat.

## 2.2 Ash density and meteorological data

The model's dynamics depends on ashes density, rainfall during the spring and heat during the summer. Even though in the long term ash dieback leads to tree mortality, we neglected it in our model, assuming that the host density is constant in time and uniform within each quadrat. This is an acceptable assumption as data shows that the volume of ash did not yet significantly decrease in the parts of France with the oldest presence of ash dieback (Gomez-Gallago and Marçais, 2022). Figure 17 shows the basal area of ashes in France on each quadrat. The data is computed from the database of the french forest inventory from 2006 to 2018 (<https://inventaire-forestier.ign.fr/dataifn/>). Informations on meteorological data collection are also given in Appendix D. The main ascospore production period in north-east France is from the 15 of June to the 15 of July (Grosdidier et al., 2018a). However, in the Pyrénées, apothecia may be observed as early as beginning of May (Marçais et al., 2023). For our estimation, precipitation during the period of apothecia production was computed as June rainfall that is represented in Figure 18. Concerning the temperature impact on the colonization by *H. fraxineus*, we tested several temperature indices. Figure 19 shows the evolution of the index which gives the best qualitative and quantitative results: the number of days during July and August with the maximal temperature above 28°C. Both meteorological and ash density data are grouped by the same  $16 \times 16$  km<sup>2</sup> quadrat grid used for ash dieback reports by the DSF.

## 3 Model description

The full model is composed of three nested sub-models. The first one describes the dynamics of leaf infection (production and diffusion of spores, leaf infection). The second model represents the development of crown symptoms on ashes, i.e. shoot mortality and dieback, depending on leaf infection given by the first model. Finally, the third model describes the statistical model for data.

### 3.1 Reaction-diffusion model for the leaf infection

We detail here the reaction-diffusion model describing the dynamics of leaf infection by *H. fraxineus*. The cycle of rachis colonization follows three steps, represented in Figure 1:

production of spores on colonized rachises on the forest floor during spring, spores diffusion and infection of leaves during the summer, and colonization of the newly produced rachises by *H. fraxineus* during fall.

We denote by  $R_a(x)$  the quantity of rachises colonized by *H. fraxineus* at the spring of the year  $a$  in location  $x$ .

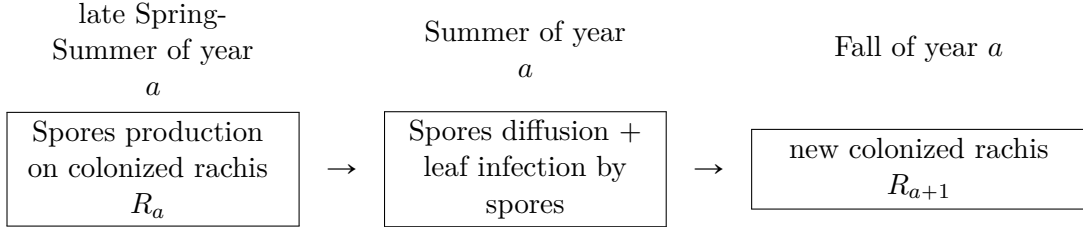


Figure 1: Reproduction cycle of *Hymenoscyphus fraxineus*.

### Late spring and summer of the year $a$ : spores production

The pathogen produces apothecia on colonized rachises, which then produces spores (ascospores). We assume that the production of apothecia depends on humidity, and is subject to an Allee effect under a quantity of colonized rachises  $r$  as suggested by Laubray et al. (2023). The Allee effect is induced by a limited encounter of sexual partners at low rachis density on the forest floor. According to Laubray et al. (2023), we propose to encode it in the model by reducing the ascospores production rate when the quantity of colonized rachises  $R_a(x)$  is below a threshold  $r$ . Here, as it is usual to model Allee effect, the ascospores production is quadratic in  $R_a(x)$  under  $r$ . The spores quantity  $\nu_a(x)$  produced in location  $x$  during late spring and summer of the year  $a$  is then given by

$$\nu_a(x) = \begin{cases} H_a(x) R_a(x) & \text{if } R_a(x) \geq r \\ H_a(x) \frac{R_a(x)^2}{r} & \text{if } R_a(x) < r \end{cases} \quad (1)$$

with

$$H_a(x) = \beta_0 + \beta_1 h_a(x) \quad (2)$$

which depends linearly on the rainfall  $h_a(x)$  in June of year  $a$  at location  $x$  (see Figure 18) to account for the effect of the humidity condition on spore production. We therefore expect a positive parameter  $\beta_1$ .

### Late spring / summer of the year $a$ : spores diffusion

The spores quantity  $\nu_a(x)$  produced by rachises of the forest floor throughout late spring / summer diffuses at long range (Grosdidier et al., 2018a) and infects the leaves. We assume that *H. fraxineus* release spores uniformly during the  $\tau = 60$  days diffusion period (June and July). The dynamics of the spores quantity  $w_a(t, x)$  located in  $x$  at time  $t$  is then given by

the following reaction-diffusion equation

$$\frac{\partial w_a(t, x)}{\partial t} = D \Delta w_a(t, x) + \frac{\nu_a(x)}{\tau} \quad (3)$$

with  $w_a(0, x) = 0$ ,  $\Delta$  is the Laplacian operator and  $D$  is the diffusion coefficient. At the end of the diffusion period of year  $a$ , the spores quantity is then  $w_a(\tau, \cdot)$ .

### Fall of year $a$ : colonization of newly produced rachises

In the fall, infected leaves are shed resulting in the production of newly colonized rachises on the forest floor. The local amount of infected leaves shed depends on both the ashes density  $d$  (represented in Figure 17) and on the level of leaf infection assumed to be linearly dependent on the spores quantity  $w_a(\tau, x)$  at the end of the diffusion period. We denote by  $\chi_a(x)$  the quantity of colonized rachis produced in the location  $x$  at the end of the summer of year  $a$ . It is assumed that, above a saturation threshold  $S$ , the increase in spore load does not increase infection because all leaves are already infected. Then,  $\chi_a(x)$  can be expressed as

$$\chi_a(x) = (w_a(\tau, x) \wedge S) d(x). \quad (4)$$

It is known that both in artificial conditions (laboratory, container in nursery) and forest litter, colonized rachises can produce apothecia during several years (Gross and Holdenrieder, 2013; Kirisits, 2015; Laubray et al., 2024). We assumed that colonized rachises fully persist during at least two years. The total colonized rachis quantity  $R_{a+1}$  on the forest floor thus depends on the quantity produced during two successive falls and is given by

$$R_{a+1}(x) = \chi_a(x) + \chi_{a-1}(x). \quad (5)$$

Note that we do not account for a poor survival of *H. fraxineus* at temperature above 35°C in rachises on the forest floor. In shaded forest conditions, the effect of high temperatures at soil level is supposed to be low and negligible compared to the temperature impact on the pathogen survival in the crown.

The initialization of the model has then to be made for two initial years and is made from observation data. We discuss it at Section 3.2.

**Numerical scheme:** Numerically, we use the same spatial discretization for the model as for the rainfall  $h_a$ , temperature  $T_a$  and ash dieback observation data (see Section 3.3), that is a discretization at the scale of one quadrat of  $16 \times 16$  km<sup>2</sup>. The model is computed by a Crank–Nicolson scheme. We consider the Neumann boundary conditions  $\nabla w_a(t, \cdot) \cdot \mathbf{n} = 0$ , where  $\mathbf{n}$  denotes the normal to the boundary. However, note that, for convenience, we numerically consider a rectangular domain, where the density of ashes is zero outside France (that is, we neglect the countries bordering France). The boundary condition is then on the bounds of this rectangular domain and not directly on the bounds of France. Neglecting spore coming from the borders may be justified as at the simulation onset, all areas close to Germany and the Flemish part of Belgium were already affected by ash dieback, which is taken into account by the initialization. Others borders of France (Wallonia part of Belgium, SW part of Switzerland, Italy, Spain) were affected by ash dieback after the French neighboring part. This condition is of course met when the border is maritime. At the end of the section, Table 1 gives the notations of the principal modeled quantities.



## 3.2 Model for the symptoms development

The development of crown dieback is caused by the transfer of the pathogen from infected leaves to the shoots before leaf shedding. The infection then remains latent and becomes perceptible during the winter following the infection. We assume that an ash tree without symptom, in location  $x$ , develops dieback observable the year  $a$ , with a probability linearly depending on the pathogen quantity infecting its leaves in the year  $a-1$ , that is proportional to  $\chi_{a-1}(x)$ . Reports made in December of the year  $n$  were counted as occurring in the year  $n+1$  as most of them concern young stand where symptoms can be seen quicker. Making the assumption that pathogen quantities by tree are uniform in each quadrat, this probability is proportional to  $\chi_{a-1}^i/d(i)$  where  $\chi_a^i = \int_{\omega_i} \chi_a(x) dx$  is the total pathogen quantity in the quadrat  $\omega_i$  and  $d(i)$  the density of ashes in the quadrat  $\omega_i$ . Moreover, this also depends on the temperature during the summer of the year  $a-1$ . *H. fraxineus* is negatively impacted by high temperatures (Hauptman et al., 2013; Grosdidier et al., 2018b). It was shown that the transfer from infected leaves to shoots, and thus the shoot mortality and dieback symptoms are hampered when the summer temperatures are too high (Marçais et al., 2023). Finally, an ash tree of the quadrat  $\omega_i$  develops observable symptoms the year  $a$  with probability

$$\tilde{q}_a^i = \frac{\chi_{a-1}^i f(T_{a-1}^i)}{r_S d(i)} \wedge 1, \quad (6)$$

where  $r_S$  is the parameter tuning the proportionality and where  $T_a^i$  is a temperature index quantifying heat during the summer of the year  $a$ . We will consider several temperature indices. For instance, one of them is the number of days in July and August of year  $a$  for which the maximal daily temperature is over  $28^\circ\text{C}$  on the quadrat  $\omega_i$  (see Section C for the description of the different indices, and Figure 14 for their corresponding results). We will consider that  $f$  is of the form

$$f(T) = \left[1 - \frac{T}{\gamma}\right]_+^\kappa, \quad (7)$$

where  $\gamma$  is the threshold of the temperature variable from which the symptoms development is no more possible and the power  $\kappa > 0$  describes how is the impact of low (but positive) values of  $T$  (namely the impact of few days with temperature getting over  $28^\circ\text{C}$  for the selected model).

Moreover, if a tree develops symptoms, the symptoms persist from one year to the following one with a probability that we denote by  $C_{\text{pers}}$ . Therefore, the probability that a tree of quadrat  $\omega_i$  has dieback symptoms during the year  $a$  is

$$q_a^i = C_{\text{pers}} q_{a-1}^i + (1 - C_{\text{pers}} q_{a-1}^i) \tilde{q}_a^i. \quad (8)$$

Now, it is natural to consider that conditionally to the pathogen leaves infection rate (or equivalently the colonized rachis quantity), the symptoms development of one tree does not depends on the symptoms development of the other trees.

### Initialisation of the model

The new colonized rachis quantity  $\chi_{2007}$  and  $\chi_{2008}$  for the two first years had to be imputed

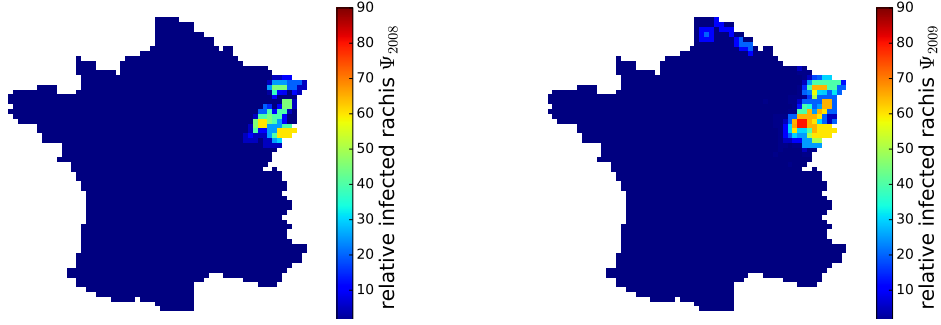


Figure 2: Prevalence intensity of ash dieback symptoms for the years 2008 (left) and 2009 (right) used to compute initial colonize rachis quantities.

from exogenous information. They were derived from the average local prevalence intensity of ash dieback observed by the DSF in each quadrat in 2008 and 2009 and denoted by  $\Psi_{2008}^i$  and  $\Psi_{2009}^i$ , using the INLA package of R to obtain a smoothed map. In 2009, we assumed that ash dieback also appeared in northern France due to pathogen invasion from the Flemish region of Belgium and we attributed low prevalence values to quadrats adjacent to that border (5%). Both quantities  $\Psi_{2008}^i$  and  $\Psi_{2009}^i$  are given in Figure 2. We assume that  $\chi_{2006}$  can be neglected, meaning that the pathogen was almost absent in 2006 and then that ash diebacks were very limited in 2007 ( $q_{2007}^i \approx 0$ ). Then we approximate  $q_{2008}$  and  $q_{2009}$  up to a multiplicative constant  $C_{\text{init}}$

$$q_{2008}^i \approx C_{\text{init}} \Psi_{2008}^i, \quad q_{2009}^i \approx C_{\text{init}} \Psi_{2009}^i, \quad (9)$$

and then from (8), we directly impute the colonized rachis quantities by

$$\chi_{2007}^i \approx C_{\text{init}} \Psi_{2008}^i \times \frac{r_S d(i)}{f(T_{2007}^i)}, \quad \chi_{2008}^i \approx C_{\text{init}} \frac{\Psi_{2009}^i - C_{\text{pers}} \Psi_{2008}^i}{(1 - C_{\text{pers}} C_{\text{init}} \Psi_{2008}^i)} \times \frac{r_S d(i)}{f(T_{2008}^i)}.$$

### 3.3 Statistical model for the observation data

For each  $a \in \llbracket 2008, 2023 \rrbracket$ , we denote by  $I_a$  the set of indices of observed quadrats during the year  $a$  and for each  $i \in I_a$ , we set  $\text{obs}_a(i)$  the number of observed plots in quadrat  $i$  during the year  $a$ . The observed data are given by  $(p_a^k(i))_{a \in \llbracket 2008, 2023 \rrbracket, i \in I_a, k \in \llbracket 1, \text{obs}_a(i) \rrbracket}$  where  $p_a^k(i)$  is the proportion of infected trees for the  $k$ -th observed plot from quadrat  $i$  in year  $a$ . As noticed in Section 3.2, reports made in December are counted as occurring in the following year. At each plot, around  $m = 30$  trees are observed so that  $m \text{obs}_a(i)$  is the number of observed trees in quadrat  $i$ .

We assume that each tree is observed at most once during all the observation period, such that the different observations are independent. As most stands are observed only once and since trees are not marked in the few stands observed more than one time, this assumption is reasonable. Therefore, we can derive the number  $N_a^i$  of infected trees observed for the

Notation	designation	Equation
$a$	year	
$x$	location	
$i$	index of quadrat $\omega_i$	
$R_a(x)$	quantity of rachises colonized by <i>H. fraxineus</i> at the spring	Fig.1,(5)
$\nu_a(x)$	spores quantity produced in $x$ during late spring and summer	(1)
$\omega_a(t, x)$	spores quantity at time $t$ in June and July	(3)
$\chi_a(x)$	colonized rachis produced at the end of the summer	(4)
$\tilde{q}_a^i$	probability that a tree develops observable symptoms	(6)
$q_a^i$	probability that a tree has dieback symptoms	(8)
$p_a(i)$	observed proportion of trees with symptoms	(11)

Table 1: Notations for the modeled quantities

quadrat  $i$  in the year  $a$ , through a binomial mechanism from the quantities  $q_a^i$  given by the model of symptoms (Eq. (8)) by

$$N_a^i \sim \text{Bin}(m \text{ obs}_a(i), q_a^i). \quad (10)$$

Moreover, conditionally on the process  $(q_a^i)_{a,i}$ , it is natural to consider that variables  $(N_a^i)_{a,i}$  are independent.

Letting  $p_a(i)$  the observed proportion of infected trees for the quadrat  $i$  in the year  $a$  defined by

$$p_a(i) = \frac{1}{\text{obs}_a(i)} \sum_{k=1}^{\text{obs}_a(i)} p_a^k(i), \quad (11)$$

then the number of infected trees observed in the quadrat  $\omega_i$  is  $(m \text{ obs}_a(i) p_a(i))_{a,i}$ ; it is a realisation of  $(N_a^i)_{a,i}$ . The data  $(p_a(i))_{a,i}$  are represented in Figure 16.

## 4 Statistical Inference

### 4.1 Parameters

The model is composed of the following ten parameters:

- **Rachis colonization model:** the diffusion coefficient  $D$  (Eq. (3)), the spores production parameter  $\beta_0$ , the humidity impact measurement parameter  $\beta_1$  (Eq. (2)), the Allee effect parameter  $r$  (Eq. (1)), the colonized rachis saturation quantity  $S$  (Eq. (4)), the multiplicative constant  $C_{\text{init}}$  for colonized rachis initialization (Eq. (9));
- **Symptoms development model:** the tuning parameter  $r_S$  of the symptoms development Bernoulli parameter (Eq. (6)), the threshold  $\gamma$  of the temperature variable which control the possibility of transfer of the infection from the leaves to the shoots, the parameter  $\kappa$  of the temperature function  $f$  (Eq. (7)), and the probability of symptoms persistence  $C_{\text{pers}}$  (Eq. (8)).

It is easy to see that, for all  $\alpha > 0$ , the parameters  $\theta = (D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{\text{init}}, r_S, C_{\text{pers}})$  and  $\tilde{\theta} = (D, \beta_0, \beta_1, \alpha r, \gamma, \kappa, \alpha S, C_{\text{init}}, \alpha r_S, C_{\text{pers}})$  give the same model dynamics. In fact, changing the coefficient  $\alpha$  just means a change of the measure unit of the colonized rachis. We arbitrarily set the parameter  $r_S = 1000$  in the following in order to set the rachis measure unit.

## 4.2 Likelihood

We use data from years 2008 to 2019 to estimate our model parameters and we keep data from 2020 to 2023 to measure the model's forecast capabilities. By assumption of independence between observations, conditionally on the process  $(q_a^i)_{a,i}$ , and by (10), the likelihood of a set of parameters  $\theta = (D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{\text{init}}, r_S = 1000, C_{\text{pers}})$  is given by

$$\mathcal{L}(\theta) = \prod_{a=2008}^{2019} \prod_{i \in I_a} \binom{m \text{ obs}_a(i)}{m \text{ obs}_a(i) p_a(i)} (q_a^i(\theta))^{m \text{ obs}_a(i) p_a(i)} (1 - q_a^i(\theta))^{m \text{ obs}_a(i) (1-p_a(i))} \quad (12)$$

where  $p_a(i)$  is the function of the observations  $(p_a^k(i))_{1 \leq k \leq \text{obs}_a(i)}$  defined by (11) and  $(q_a^i(\theta))_{a,i}$  is the process of the dieback symptoms probabilities defined by (8) for the model described in Sections 3.1 and 3.2 with the set of parameters  $\theta$ . At a first look, the likelihood seems very simple but the processes  $(q_a^i(\theta))_{a,i}$  are derived from the reaction diffusion process so that it is impossible to have an explicit formula for the optimal  $\theta$ . It is why we choose a bayesian approach to optimise the likelihood.

## 4.3 Bayesian method for the parameters estimation

In this section, we describe the process that leads to the parameters estimation. We have chosen to compute the posterior distribution of the parameter  $\theta$  by the Adaptive Multiple Importance Sampling (AMIS) algorithm (Cornuet et al., 2012) described in Section 4.3.2. Even if the main aim of AMIS is to overcome the difficulty to choose a proposal distribution as close as possible to the posterior distribution, authors strongly require that a significant part of the computing effort be spent on the initialization stage. That is crucial to design an efficient importance sampling algorithm and we have first proceeded to an initial rough estimation of the parameters. The purpose of this stage is twofold: to see if some parameters can be dropped or fixed to a constant and also to have a proposal not so far from the posterior. The results of this procedure are described in the following sections, while the technical details are postponed in Appendix.

### 4.3.1 First estimation of the parameters: model reduction and parameter initialization

It is well known that the initialization of the algorithms is one of the main difficulties in the parameters estimation by iterative methods. The process used for the first rough estimation of the parameters as well the numerical results are described in Appendix A. This approach leads to a model reduction setting four parameters to specific values. In particular, it stands

out that the probability of the symptom persistence  $C_{\text{pers}}$  is equal or very close to 1. This is consistent with field observation. Ash dieback usually result in severe mortality of at least small and medium shoots and those can be observe on the tree during several years with thus persistent symptoms. Moreover, the parameters  $\beta_1$  seems to be negligible with respect to  $\beta_0$  in (2), meaning that the rainfall variations have a minor impact on the long range dispersal of *H. fraxineus*. Furthermore, the Allee effect parameter  $r$  is negligible in the more likely zone of parameters. Finally, on our first estimates, it appears that the posterior distribution of the parameter  $\gamma$  is, for the temperature index considered in Section 5, almost degenerated in a value close to, but strictly larger than 21.5 (see Section A.3).

After the first numerical tests, we then set the probability of symptoms persistence ( $C_{\text{pers}}$ ) to one, the Allee effect parameter ( $r$ ) and impact of the rainfall ( $\beta_1$ ) to zero and the threshold  $\gamma$  of the temperature index to 21.50001:

$$C_{\text{pers}} = 1, \quad \beta_1 = 0, \quad r = 0, \quad \gamma = 21.50001. \quad (13)$$

In addition, this first step gives a rough estimation of the parameters  $D$ ,  $\beta_0$ ,  $\kappa$ ,  $S$  and  $C_{\text{init}}$  for the initialization of the AMIS algorithm, described in the next section.

### 4.3.2 Adaptive Multiple Importance Sampling Algorithm

Following the model reduction (13), the posterior distribution of the set of parameters  $\theta = (D, \beta_0, \kappa, S, C_{\text{init}})$  is computed by the AMIS algorithm described by Algorithm 1 (Cornuet et al., 2012). The principle of the algorithm is the following: at each step 1) we sample  $N$  sets of parameters according to a normal proposal distribution  $\mathcal{N}(m_1, \Sigma_1)$ ; 2) we compute the weight associated to each parameter (included parameters samples of previous steps) with respect to the likelihood function, the prior density and the proposal density; 3) we update the proposal distribution according to the parameters weights. This algorithm then provides weighted parameters sets describing the posterior distribution of  $\theta$ .

This algorithm has the advantage to allow the parallelization of the likelihood computations of the sampled parameters at each iteration step and to recycle all simulated parameters (which is particularly important when the likelihood is costly to compute). Moreover, the efficiency of the AMIS algorithm for banana shape target (as suggested by our first estimation step for the 2-dimensional posterior distribution of  $(D, \beta_0)$ , see Figure 10a) has been illustrated by Cornuet et al. (2012). The main difficulty is to determine the initial proposal distribution (namely the mean  $m_1$  and the covariance matrix  $\Sigma_1$ ).

Even though the first estimation of the parameters reveals a strong correlation between some parameters (in particular between  $\beta_0$  and  $D$ , see Figure 10a), our prior information remains imprecise. We then choose five independent uniform distributions for our prior as the dependency will be corrected by the algorithm:

$$\pi(\theta) = \mathcal{U}_{[0, \bar{D}] \times [0, \bar{\beta}_0] \times [0, \bar{\kappa}] \times [0, \bar{S}] \times [0, \bar{C}_{\text{init}}]}(D, \beta_0, \kappa, S, C_{\text{init}})$$

with  $\bar{D}$ ,  $\bar{\beta}_0$ ,  $\bar{\kappa}$ ,  $\bar{S}$  and  $\bar{C}_{\text{init}}$  sufficiently large such that they don't impact the posterior distribution.

We initialize the proposal distribution by independent normal distribution (i.e. the covariance matrix  $\Sigma_1$  is diagonal) whose the means and the variances are roughly estimated

initialization of the mean  $m_1$  and the covariance matrix  $\Sigma_1$   
**for**  $k = 1 \dots N_L$  **do**  
 $\theta_{ki} \sim \mathcal{N}(m_k, \Sigma_k), i = 1 \dots N$  {new parameters sample}  
 $\tilde{w}_{li} \leftarrow \frac{L(\theta_{li})\pi(\theta_{li})}{\frac{1}{k} \sum_{\ell=1}^k g_{m_\ell, \Sigma_\ell}(\theta_{li})} l = 1 \dots k, i = 1 \dots N$  {weights}  
 $w_{li} \leftarrow \tilde{w}_{li} / \sum_{\ell=1}^k \sum_{i=1}^N \tilde{w}_{\ell i}$  {normalized weights}  
 $m_{k+1} \leftarrow \sum_{l=1}^k \sum_{i=1}^N w_{li} \theta_{li}$  {proposal distribution update}  
 $\Sigma_{k+1} \leftarrow \sum_{l=1}^k \sum_{i=1}^N w_{li} (\theta_{li} - m_{k+1}) (\theta_{li} - m_{k+1})^t$   
**end for**

Algorithm 1: Adaptive Multiple Importance Sampling Algorithm for parameter estimation of the reduced model. The function  $g_{m, \Sigma}$  is the density of a normal distribution  $\mathcal{N}(m, \Sigma)$  with mean  $m \in \mathbb{R}^5$  and covariance matrix of size  $(5, 5)$ . For  $\theta = (D, \beta_0, \kappa, S, C_{\text{init}})$ ,  $L(\theta) = \mathcal{L}(D, \beta_0, \beta_1 = 0, r = 0, \gamma = 21.50001, \kappa, S, C_{\text{init}}, r_S = 1000, C_{\text{pers}} = 1)$  is the likelihood function defined in Section 4.2 with the model reduction (13).

thanks to the first estimation step (see Section A.2 for more details). The values of the mean  $m_1$  and the covariance matrix  $\Sigma_1$  are given in Table 2 for the temperature index used for the numerical results of the next section.

$$\begin{array}{lll}
m_D = 18 & m_{\beta_0} = 24 & m_\kappa = 0.05 \\
m_S = 90 & m_{C_{\text{init}}} = 0.0085 & \\
\sigma_D^2 = 0.4 & \sigma_{\beta_0}^2 = 3 & \sigma_\kappa^2 = 1.5 \times 10^{-5} \\
\sigma_S^2 = 1 & \sigma_{C_{\text{init}}}^2 = 1 \times 10^{-8} & 
\end{array}$$

Table 2: Initial values of the mean  $m_1 = (m_D, m_{\beta_0}, m_\kappa, m_S, m_{C_{\text{init}}})$  and the covariance matrix  $\Sigma_1 = \text{diag}(\sigma_D^2, \sigma_{\beta_0}^2, \sigma_\kappa^2, \sigma_S^2, \sigma_{C_{\text{init}}}^2)$  of the AMIS Algorithm 1 for the temperature index T28.

The convergence of the algorithm is discussed in Appendix B.

## 5 Numerical Results for the best temperature index (T28)

The method described in Section 4.3 have been repeated for several temperature indices. The qualitative and quantitative behaviors of the model for these different temperature indices are discussed in Section C in Appendix. In this section, we only present the numerical results for the model where the temperature variable  $T_a^i$  is the number of days of July and August of the year  $a$  for which the temperature is over  $28^\circ\text{C}$  on the quadrat  $\omega_i$  (called the *temperature index T28*). Among the temperature indices tested, T28 gives the best qualitative performance (in particular, it correctly explains the lack of expansion of ash dieback in south-east France, see Section 5.2) as well as the best quantitative performance (highest likelihood for the period 2008-2023 among the temperature indices tested, see Section C).

We run<sup>1</sup> Algorithm 1 with  $N_L = 60$  iterations, the samples size  $N = 10000$ , the initial-

<sup>1</sup>Simulations are run on the babycluster of the Institut Élie Cartan de Lorraine.

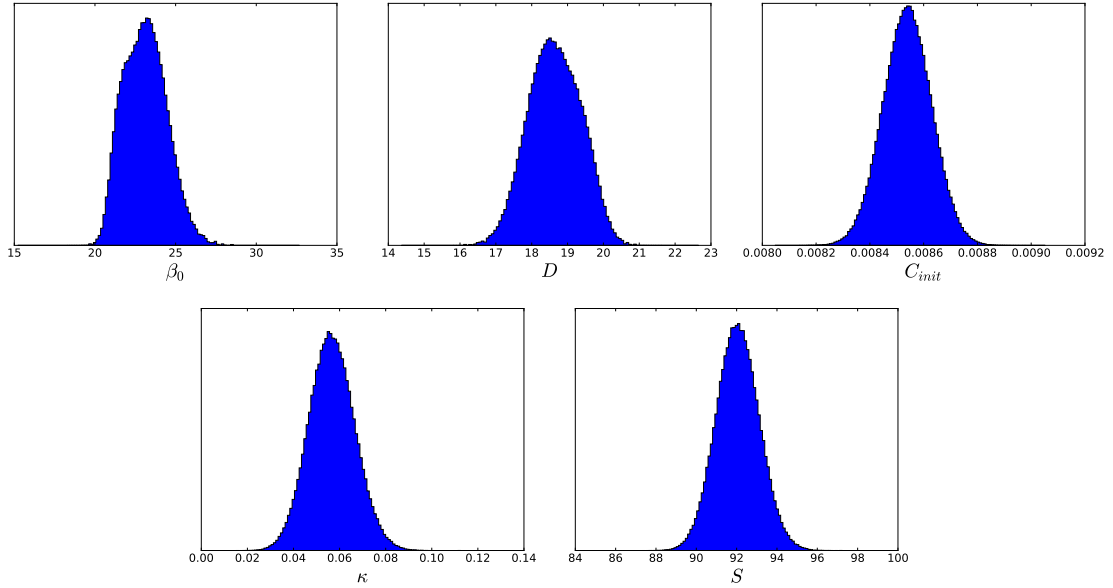


Figure 3: Marginal posterior distributions of the five estimated parameters  $\beta_1$ ,  $D$ ,  $C_{\text{init}}$ ,  $\kappa$  and  $S$ .

ization of the posterior distribution given in Table 2 and the temperature index T28. We first present the computed posterior distributions. Then we plot the dynamics for the best parameter set. Finally we discuss the validation of the model.

## 5.1 Posterior distributions of parameters

The one-dimensional posterior distributions of the five estimated parameters are presented in Figure 3. They give the acceptable ranges for the five estimated parameters  $\beta_1$ ,  $D$ ,  $C_{\text{init}}$ ,  $\kappa$  and  $S$ .

The correlation coefficients between estimated parameters are summarize in Table 3. In addition, the 2-dimensional posterior distributions of the parameters  $(\beta_0, D)$  and  $(S, \kappa)$ , for which the correlations are the highest, are presented in Figure 4. As already observed in the first estimation step (see Appendix A and Figure 10a), there is a strong negative correlation between parameters  $\beta_0$  and  $D$ , which can be interpreted as follows: increasing  $\beta_0$  increases the source term  $\nu_a(x)/\tau$  in (3) (amount of spores produced at  $x$ ) and then speeds up the propagation of *H. fraxineus*, which is counterbalanced by decreasing the diffusion coefficient  $D$ . However, increasing  $\beta_0$  and decreasing  $D$  increases the local infection / colonization.

Both parameters  $S$  and  $\kappa$  are also correlated.  $S$  set the maximal spore load possible ( $\chi_a(x)$  in (4)). This will determine the amount of foliar infection at a given environmental condition.  $\kappa$  determines how high summer temperatures control the rate of leaf infection transfer to shoot infection. When the spore load and thus the leaf infection are higher (high  $S$ ), there is a compensation by a stronger and more gradual control of the transfer of the pathogen from leaves to shoots (higher  $\kappa$ ).

	D	$\beta_0$	$\kappa$	$S$
$\beta_0$	<b>-0.97</b>			
$\kappa$	0.06	-0.04		
$S$	0.09	-0.11	<b>0.58</b>	
$C_{\text{init}}$	-0.09	0.08	0.02	-0.14

Table 3: Correlation coefficients between estimated parameters.

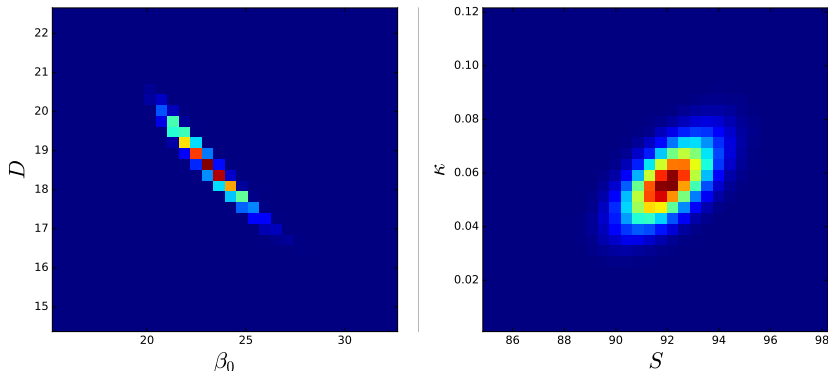


Figure 4: 2-dimensional posterior distributions of the parameters  $(\beta_0, D)$  and  $(S, \kappa)$ .

## 5.2 Predicted expansion of ash dieback for the best set of parameters

In this section, we present the model dynamics for the parameters set, among the parameters sampled by the AMIS algorithm, which maximises the likelihood. It corresponds to the parameter  $\theta^{\max} = (D^{\max}, \beta_0^{\max}, 0, 0, 21.50001, \kappa^{\max}, S^{\max}, C_{\text{init}}^{\max}, 1000, 1)$  with

$$\begin{aligned} D^{\max} &= 18.51; & \beta_0^{\max} &= 23.26; & \kappa^{\max} &= 0.056; \\ S^{\max} &= 92.0; & C_{\text{init}}^{\max} &= 0.00855. \end{aligned} \quad (14)$$

Figure 5 represents the evolution of the predicted probability that an ash tree develops observable symptoms (bernoulli parameter  $q_a^i$  in (8)) from  $a = 2008$  to  $a = 2019$ , i.e. for the period used for the estimation, with the parameter set  $\theta^{\max}$ . The front propagation, represented by the white crosses for the data on Figure 5, is generally well captured by our model. In some years, isolated reports were made well in advance of the predicted propagation front, in areas not contiguous to previously infected areas (2011 in central France, 2012-13 in the Cotentin, 2015 in the Charentes). They are often reports in ash plantation, in particular in the Cotentin, 2012-13. However, they usually remained isolated reports and led to a clear regional outbreak only in 2015 in the Charentes (central western France). After 2015, ash dieback is still predicted to expand in western, and, to a lesser degree, in southwest France, while its expansion is predicted to stop in southeast France. Expansion is limited by high summer temperatures in this region. This fits well the observations as no further ash dieback expansion is observed in southeast France after 2015.



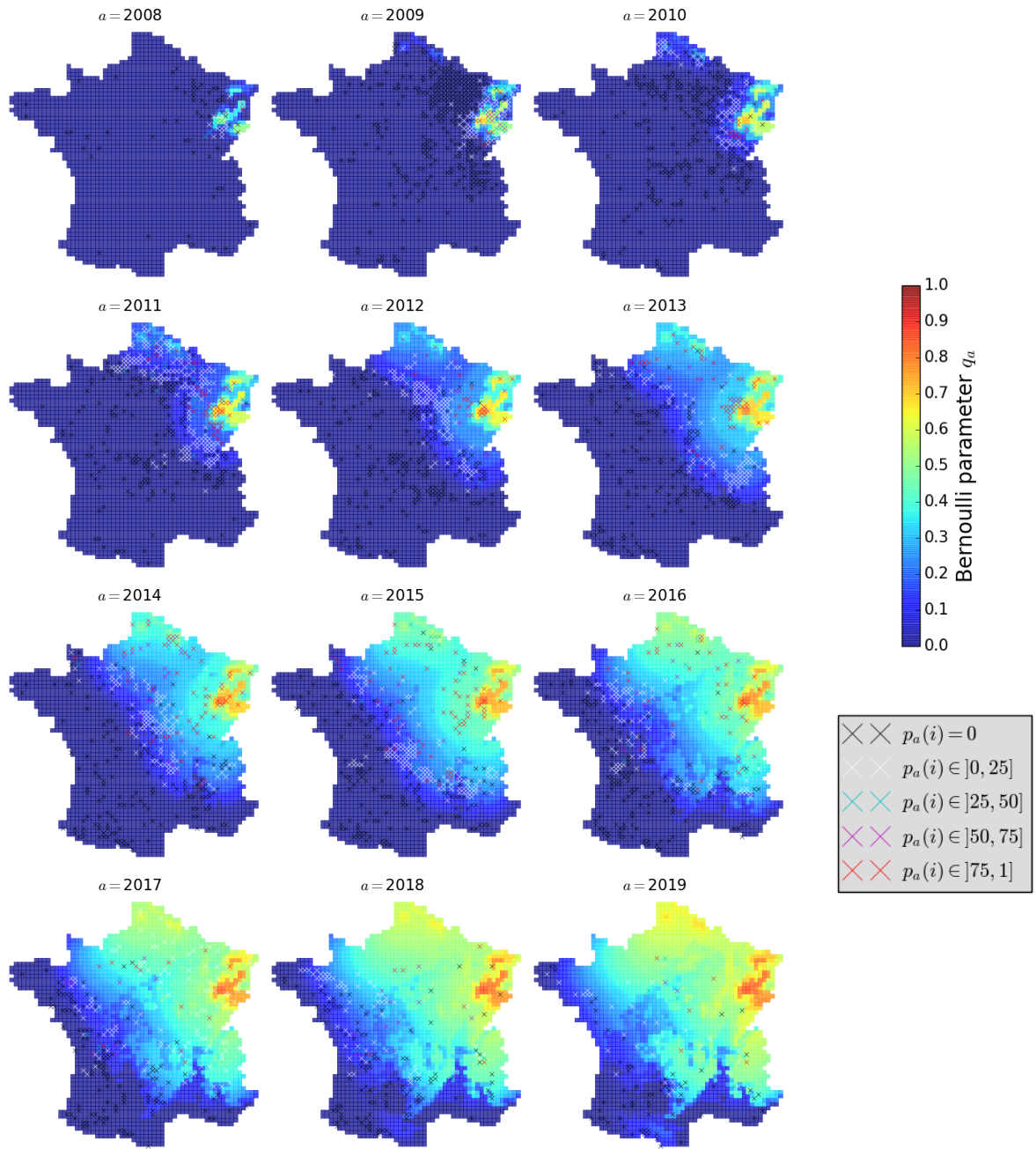


Figure 5: Evolution of the estimated probability that an ash tree of the quadrat  $\omega_i$  develops observable symptoms (Bernoulli parameter in (8)) for the set of parameters  $\theta^{\max}$  given by (14) which maximizes the likelihood for the temperature index T28. Crosses: observations data  $p_a(i)$  (see (11)) from 2008 to 2019. Black crosses if no observed symptoms and white, cyan, magenta and red crosses for proportions of trees with dieback symptoms in  $]0, 0.25]$ ,  $]0.25, 0.5]$ ,  $]0.5, 0.75]$  and  $]0.75, 1]$  respectively.

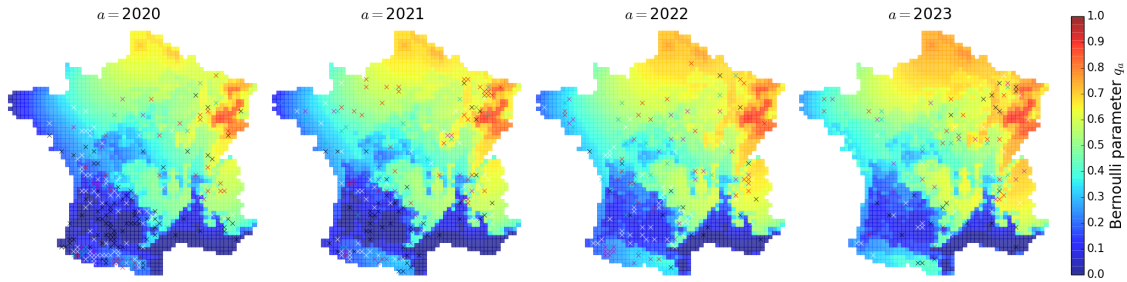


Figure 6: Prediction of the Bernoulli parameter (8) for the set of parameters  $\theta^{\max}$  given by (14) (fitted on data from 2008 to 2019 for the temperature index T28) from 2020 to 2023. Crosses: observation data from 2020 to 2023 (see Figure 5 for the color coding).

Figure 6 represents the prediction of the probability that an ash tree will develop observable symptoms ( $q_a^i$ ) for years 2020 to 2023 for the parameter  $\theta^{\max}$ . Note that this parameter was estimated with data from 2008 to 2019 and that we are in a context of prediction. We can see that the model fits well what is observed in 2020-2023. In particular, it predicts that ash dieback will expand toward Spain in the eastern part of the Pyrénées throughout a relatively narrow corridor. Notice that no observation of ash dieback has been reported by the DSF in this area during the fitted period 2008-2019, ash dieback was observed for the first time in the Pyrénées in 2020. The corridor corresponds to an area of higher elevation where summer temperature remains cooler and ash density is significant (Figure 17). Report of ash dieback remained scarce in the Garonne valley and the ash dieback prevalence of the few reports available in the area in 2020-23 is limited (under 10%).

Figures 7 and 8 represent the dynamics of the spore quantity  $w_{a-1}(\tau, x)$  and of the saturated spore quantity  $w_{a-1}(\tau, x) \wedge S$  respectively, from 2010 to 2019. Note that, due to the delay between infection and the dieback visibility, graphs compare observation data of the year  $a$  and (saturated) spore quantities of the year  $a - 1$ . We observe that, except on the propagation front, the zone is either infected and saturated or non infected ( $w_a(\tau, x) = w_a(\tau, x) \wedge S \in \{0, S\}$ ), meaning that in infected zones, the rachis infection given by (4) is maximal and the spores quantity is non limiting. We also observe that the spore quantity decreases the first two years in east France, probably reflecting that the initial conditions over estimates the quantities of colonized rachis in 2007 and 2008 in order to compensate the neglected symptoms before 2008. In addition, we can notice that, as nothing limits the propagation of spores in the model, they entirely colonize France. High summer temperatures only impact the development of the dieback symptoms, as observed in Figure 5.

### 5.3 Model validation

To complete Figures 5 and 6, we computed the Root Mean Square Error (RMSE) for each year  $a$ , for  $a = 2008$  to  $a = 2023$ , given by

$$\text{RMSE}(a) = \sqrt{\frac{\sum_{i \in I_a} \sum_{k=1}^{\text{obs}_a(i)} (p_a^k(i) - q_a^i)^2}{\sum_{i \in I_a} \text{obs}_a(i)}}$$

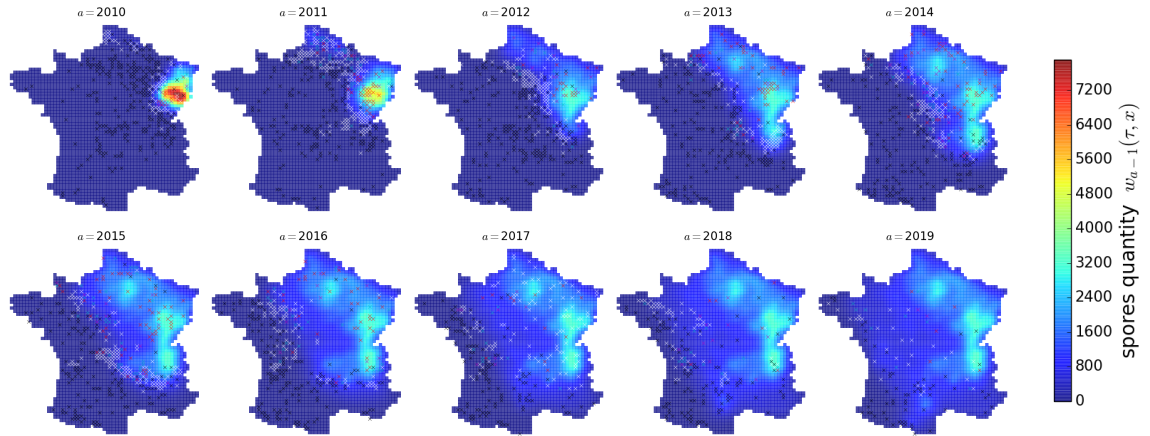


Figure 7: Evolution of the spore quantity  $w_{a-1}(\tau, x)$ , for the set of parameters  $\theta^{\max}$  given by (14), from 2010 to 2019. Crosses: observation data from 2010 to 2019 (see Figure 5 for the color coding).

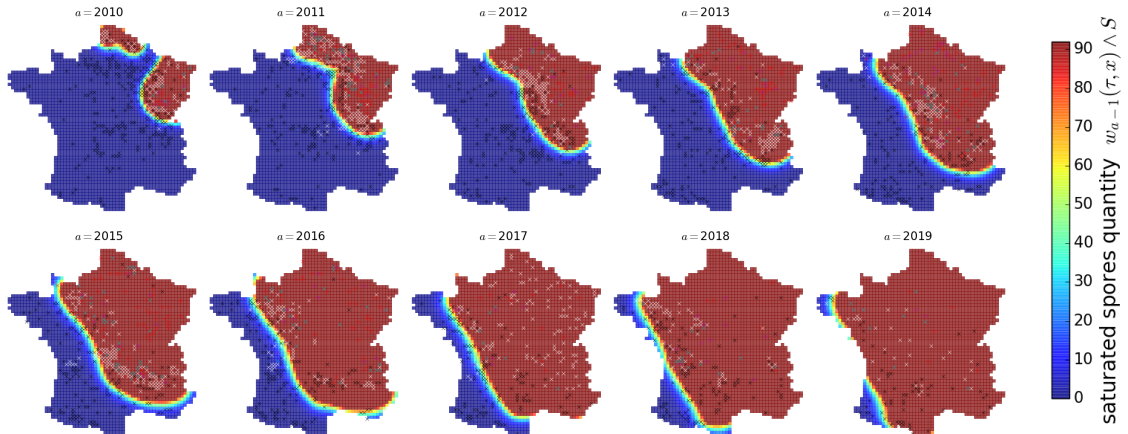


Figure 8: Evolution of the saturated spore quantity  $w_{a-1}(\tau, x) \wedge S$ , for the set of parameters  $\theta^{\max}$  given by (14), from 2010 to 2019. Crosses: observation data from 2010 to 2019 (see Figure 5 for the color coding).

Year $a$	$\sum_{i \in I_a} \text{obs}_a(i)$	RMSE( $a$ )	RMSE(2008 $\rightarrow a$ )
2008	183	0.2222	0.2222
2009	753	0.1465	0.1641
2010	504	0.1898	0.1735
2011	620	0.2387	0.1954
2012	515	0.1940	0.1951
2013	438	0.2621	0.2062
2014	493	0.2739	0.2170
2015	520	0.2720	0.2249
2016	410	0.2437	0.2267
2017	341	0.2956	0.2323
2018	254	0.2484	0.2331
2019	132	0.2207	0.2328
2020	346	0.3071	0.2381
2021	215	0.3286	0.2421
2022	166	0.3413	0.2455
2023	111	0.3819	0.2487

Table 4: Number of observed parcels in the year  $a$ , RMSE of the year  $a$ , and RMSE of the period from 2008 to the year  $a$ , with  $a$  from 2008 to 2023, for the model fitted with data from 2008 to 2019.

where  $p_a^k(i)$  (resp.  $q_a^i$ ) is the observed Bernoulli parameter (resp. the Bernoulli parameter estimated by the model of the previous section (that is for the set of parameter  $\theta^{\max}$  given by (14))). It allows us to measure the accuracy of the model for each year. In addition, we also compute, for each year  $a$ , the RMSE of the period from 2008 to the year  $a$  given by

$$\text{RMSE}(2008 \rightarrow a) = \sqrt{\frac{\sum_{y=2008}^a \sum_{i \in I_y} \sum_{k=1}^{\text{obs}_y(i)} (p_y^k(i) - q_y^i)^2}{\sum_{y=2008}^a \sum_{i \in I_y} \text{obs}_y(i)}}.$$

Contrary to  $\text{RMSE}(a)$ , it gives a lower weight to the year with a low number of measures on the period. They are given in Table 4. Smaller is the RMSE, better is the prediction. We expect a model that fits better, at least, that a model that gives a purely random prevision. To check this, we simulated random imputations of the  $q_a^i$  by a uniform distribution on  $[0, 1]$  and computed the empirical distributions of the  $\text{RMSE}(a)$ , for  $a$  in the period 2008-2023, for this purely random model. Simulations of 100000 replica always give  $\text{RMSE}(a)$  between 0.35 and 0.50. In addition, the first percentiles of the empirical distributions are between 0.42 and 0.52, clearly greater than the ones of our fitted model.

Note that our model is clearly better than the one of purely random prevision. We can also remark that years from 2008 to 2019 with many observations (column  $\sum_{i \in I_a} \text{obs}_a(i)$ ) are well fitted because they have a great contribution in the whole likelihood that as to be maximised. Remember that data from 2020 to 2023 were not used in the fit, which explains a RMSE that gradually increases during the period, although it remains still clearly lower than the one of a random model.

## 6 Discussion

We developed a complex mechanistic-statistical model describing the expansion of ash dieback, taking into account climate factors (temperature and rainfall), Allee effect, and heterogeneity in ash tree density. We estimated the parameters of our model by first identifying appropriate ranges for parameters, which led to a model reduction. We then used an Adaptive Multiple Importance Sampling Algorithm for fitting the model. We tested several temperature indices. The temperature index T28, which corresponds to the number of days in summer with a temperature exceeding 28°C, provided the highest forecast likelihood, in the sense that, for the parameter set maximizing the likelihood on the fitted period 2008-2019, it obtained the best likelihood for the period 2008-2023. In addition the model with the temperature index T28 leads to good qualitative features. Specifically, for this index, the model well depicts the qualitative propagation behavior that has been observed in France over the past 20 years. In particular the model successfully reproduces the absence of disease spread in southeast France and the limited spread of the disease in southwest France, with a more significant development occurring once the disease reaches the Pyrénées mountains. Both the Allee effect and the heterogeneity of the rainfall in spring do not have a significant impact on the overall expansion of *H. fraxineus* in France and can be neglected in the modeling. Similar qualitative behaviors and likelihoods have been obtained with other indices, such as the number of days in summer with temperatures exceeding 24°C, 26°C or 30°C (indices T24, T26 and T30), not allowing selection of one of these indices (index T26 even leads to better likelihood for the fitted period 2008-2019). Nevertheless, these temperature indices lead to the same estimation range for parameters (excluding the parameters of the temperature impact function  $f$ ), then leading to similar conclusions about the model dynamics. However, our results suggest that the index corresponding to temperatures exceeding 35°C (index T35) should be excluded. This index does not accurately reproduce the propagation dynamics of the data, particularly failing to show the absence of disease in southeast France.

For the selected temperature index T28, the developed model adequately reproduces several key features of the ash dieback dynamics observed in France. First, the dynamics of the inoculum, i.e. the amount of infected rachises in the litter, show a quick and sharp transition from low to high levels within 1-2 years. This very abrupt transition, which has a significant impact on most local stands just a few years after the first observation of the disease in an area, corresponds to the observed dynamics in France (Grosdidier et al., 2020). Then the model well depicts the absence of dieback in many places, although it predicts the complete colonisation of France by *H. fraxineus*. This is in agreement with the prediction that the pathogen is able to complete its cycle everywhere in France (Marçais et al., 2023). We do have indications that this prediction has some validity. A specific survey in southeast France (Avignon) during 2023 summer was able to locate ash trees with leaf infections by *H. fraxineus*, although shoot infections were absent (Marçais, unpublished). Ash dieback is still not reported around Avignon in accordance with our model which predicts no presence of disease in the area in 2023. *H. fraxineus* may thus be present much more widely in SE France than previously thought, but does not appear to induce any damage as predicted by our model. The model also adequately predicts that the impact of ash dieback will be very limited in the Garonne valley. In SW France, the apparent stop of ash dieback spread

in 2015 and the colonization of the Pyrénées after 2019 is predicted by the model, even though only data from 2008-19 were used for model calibration. Following the widespread presence reported in the Pyrénées after 2020, the DSF conducted extensive surveys in the Garonne valley and confirmed that although the disease may occasionally be found there, its prevalence and severity remain very limited. Our model outputs are in agreement with the hypothesis of Marçais et al. (2023) that undetected spread through foliar infection with limited induced dieback may explain the arrival in the Pyrénées in 2020, with a jump of about 200 km from the closest previous known disease location.

We thus confirm that summer high temperatures are a major climatic factor that affect the impact and spread of ash dieback, as was shown by Hauptman et al. (2013), Grosdidier et al. (2018b) and Marçais et al. (2023). The parameter  $\gamma$  was estimated at 21.5 for the model with the temperature index T28, which means that ash dieback stops to be reported by the DSF in area where temperature exceeds 28°C for more than 21.5 days during the summer in most years. This occurs in France mainly in the SE, although during the period of 2018-2020, it occurred in a much larger area of France. The function  $f$  decreases very sharply close to 21.5 from 0.8 to 0 and is thus almost a threshold function ( $\kappa$  estimated at 0.056, see Figure 12a). This may be caused by the poor survival of *H. fraxineus* at temperature above 35°C (Hauptman et al., 2013). Leaf temperature in the crown may reach value up to 5°C above the air temperature (Granier et al., 2007). Another cause of this strong threshold is probably the observation process. Ash dieback is reported by the forest health survey system when it reaches a level that impacts stand management. When *H. fraxineus* induced little shoot mortality because summer temperature are unfavorable, the trees are able to compensate by producing new foliage and do not show significant dieback. Ash dieback is then not reported as a forest health problem. Moreover, in southeast France, the species of ash trees *F. excelsior* is replaced by *F. angustifolia* and becomes less economically important in the forests, leading to less thorough observation by the DSF and fewer reports of ash dieback.

A feature taken into account in the model is the local ash density (average local basal area of ash per ha). Host density is recognized as a key factor in explaining the spread of pathogens, particularly invasive parasites that affect forests, either microorganisms or insects (Hudgins et al., 2017; Keesing et al., 2010). In the model, it scales the production of inoculum throughout the production of infected rachises that fall on the forest floor each autumn, and thus disease transmission. It was demonstrated that population of ash with a high basal area are more severely impacted by *H. fraxineus* (Grosdidier et al., 2020; Chumanová et al., 2019). The fast colonization of central France (Massif central) as well as the spread of the disease through a corridor between the massif central and the Pyrénées could be partly driven by high ash density in the area. However, we did not observe a slower spread toward the west of France after 2015, despite a very low ash density in the area. The ash density that we used is based on forest inventory data, which only accounts for ash trees in forests. It could inadequately represent ash density in western France where the acreage of forest is low while hedges are very frequent. Ash trees in the hedges are not taken into account by the forest inventory data, resulting in a likely underestimation of the overall ash density in western France.

By contrast, some factors known to affect ash dieback could be neglected in modeling the

spread of the disease. It was shown that a component Allee effect significantly affects the population dynamics of *H. fraxineus* and is induced by limited encounters of sexual partners (i.e. the mating types) at low density of colonized ash rachises in the forest litter (Laubray et al., 2023). Hamelin et al. (2016) showed that an Allee effect induced by a mating limitation significantly reduces the spread rate of *Mycosphaerella fijiensis*. We did not find any evidence that this is the case for ash dieback. The Allee effect observed for *H. fraxineus* is a weak Allee effect, the production of ascospores remaining positive even at very low density of the colonized rachises. This may explain the limited impact of Allee effect on the spread of ash dieback. Another factor that could be neglected is the impact of rainfall. *H. fraxineus* production of apothecia and ascospores has been shown to be positively related to spring and summer rainfall (Havrdová et al., 2017; Chumanová et al., 2019; Dvořák et al., 2023) and we therefore expected that it would significantly affect the disease spread. This is not the case. A possibility is that rainfall in France during spring seldom hampers the pathogen sporulation completely. In fact, Figures 7 and 8 suggest that, due to the saturation effect, the impact of the rainfall heterogeneity would have played a role only on the propagation front. Marçais et al. (2023) showed that rainfall enables the fulfillment of *H. fraxineus* cycle in all French regions, even in the southeast where the disease is not observed. It remain possible that spread of *H. fraxineus* will be reduced in mediteranean areas drier than southeast France. However, in these areas the period of apothecia production is likely to start earlier, in late winter or early spring, during a rainy enough period for the pathogen’s cycle to be fulfilled. In the Pyrénées, apothecia were produced as early as May 1st (Marçais et al., 2023). Also, the pathogen may be able to produced enough apothecia for efficient dispersal in favorable wet sites even during dry springs.

In conclusion, we have proposed a model that well describes the dynamics of dieback symptoms caused by *H. fraxineus* in France. Our numerical analysis suggests that the density of ash trees and high temperatures, particularly the number of days in summer with temperatures exceeding 26-28°C, play significant roles in the spread of the pathogen and the development of dieback symptoms. Additionally, the statistical estimation of the model parameters has allowed us to quantify the impact of these local environmental parameters, as well as the global propagation parameters such as the diffusion parameter  $D$  and the reproduction parameter  $\beta_0$ . In future research, we plan to further investigate the local speed of propagation in the reaction-diffusion model, which is dependent on the local ash tree density.

## A First rough estimation and model reduction

We present in this appendix the first estimation of the parameters leading to the model reduction (13) as well as the initialization of Algorithm 1. This step have been repeated for several temperature indices, however we show here the numerical results only for the temperature index T28 (i.e.  $T_a^i$  is the number of days of July and August of the year  $a$  for which the temperature exceeds 28°C in the quadrat  $\omega_i$ ), which gives the best qualitative and quantitative behaviors (see Section C for a comparison of the model behavior for the different temperature indices).

## A.1 A Metropolis-Hasting algorithm with iteration dependent proposal distribution

In order to obtain a first rough estimation of the parameters, we realized ten runs of a Metropolis-Hasting algorithm with a proposal distribution depending on the iteration step (see Algorithm 2) initialized with different initial conditions. At each iteration only one of the parameters  $D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{\text{init}}, C_{\text{pers}}$  is changed (we remind that  $r_S$  is set to 1000, see Section 4.1) according to a Gamma distribution with a *not too small* variance. Testing parameters one by one allows to not reject a bloc of new parameters because only one parameter is not adequate and, combined with the large variances of the proposal distribution, allows users to capture a fast convergence of parameters which are not strongly correlated to the others ones towards the ‘main support’ of the posterior distribution. More precisely, for  $i \in \llbracket 1, 9 \rrbracket$ , at the  $k = 9\ell + i$ -th iteration, and setting  $\theta = (D, \beta_0, \beta_1, r, \gamma, \kappa, S, C_{\text{init}}, r_S, C_{\text{pers}}) = (\theta_1, \dots, \theta_8, r_S, \theta_9)$  then

$$\tilde{\theta} = (\tilde{\theta}_1, \dots, \tilde{\theta}_8, r_S, \tilde{\theta}_9) \sim Q^k(\cdot|\theta) \iff \begin{cases} \tilde{\theta}_i \sim \Gamma(\lambda, \frac{\theta_i}{\lambda}) \\ \tilde{\theta}_j = \theta_j, \forall j \neq i \end{cases} \quad (15)$$

and  $q^k(\tilde{\theta}|\theta)$  is the density of the gamma distribution  $\Gamma(\lambda, \frac{\theta_i}{\lambda})$  with scale parameter  $\lambda$  and shape parameter  $\theta_i/\lambda$ , that is with mean  $\theta_i$  and variance  $\theta_i^2/\lambda$ . Moreover, without prior information, we choose for  $\tilde{\pi}$  a uniform distribution on  $[0, \bar{D}] \times [0, \bar{\beta}_0] \times [0, \bar{\beta}_1] \times [0, \bar{r}] \times [0, \bar{\gamma}] \times [0, \bar{\kappa}] \times [0, \bar{S}] \times [0, \bar{C}_{\text{init}}] \times \{1000\} \times [0, 1]$  with  $\bar{D}, \bar{\beta}_0, \bar{\beta}_1, \bar{r}, \bar{\gamma}, \bar{\kappa}, \bar{S}$  and  $\bar{C}_{\text{init}}$  sufficiently large such that parameters don’t reach the upper bounds of there domains.

```

initialization of  $\theta^0$ 
 $k \leftarrow 0$ 
while  $k \leq N$  do
   $\hat{\theta} \sim Q^k(\cdot|\theta^k)$  {new parameter to be tested}
   $\delta \leftarrow \frac{\mathcal{L}(\hat{\theta}) \tilde{\pi}(\hat{\theta}) q^k(\theta^k|\hat{\theta})}{\mathcal{L}(\theta^k) \tilde{\pi}(\theta^k) q^k(\hat{\theta}|\theta^k)}$ 
   $u \sim U[0, 1]$ 
  if  $u \leq \delta$  then
     $\theta^{k+1} \leftarrow \hat{\theta}$  {acceptation of the new parameter}
  else
     $\theta^{k+1} \leftarrow \theta^k$  {rejection of the new parameter}
  end if
   $k \leftarrow k + 1$ 
end while

```

Algorithm 2: Metropolis-Hastings Algorithm with a proposal distribution depending on the iteration step.

Figure 9 shows the evolution of the parameters computed by Algorithm 2, with the iteration dependent proposal distribution (15) with  $\lambda = 200$ , starting from ten different initial parameters. Notice that the initial prevalence shape given by  $\Psi_{2008}$  and  $\Psi_{2009}$  constraints some parameters. In fact, the conditions  $q_{2008} < 1$  and  $q_{2009} < 1$  combined to Eq. (9) imply



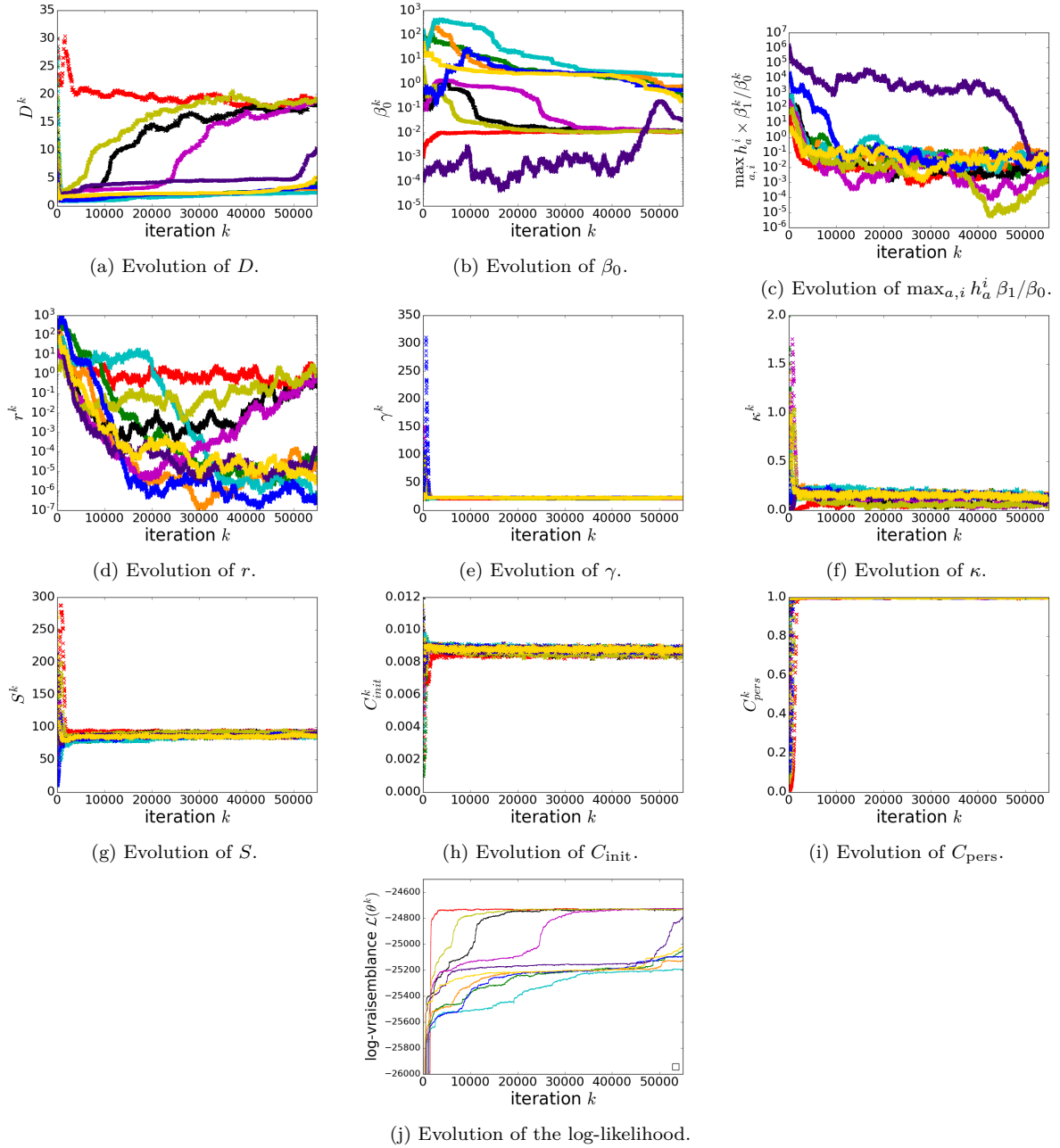


Figure 9: Dynamics of the parameters and the log-likelihood for the Metropolis-Hastings Algorithm 2, with the iteration dependent proposal distribution (15) with  $\lambda = 200$ , starting from 10 different initial parameters  $\theta^0$ . Note that Figures 9b, 9c and 9d are represented on a log-scale.

that  $C_{\text{init}} < 0.0125$ . In addition, the initially non infected quadrats (i.e. such that  $\Psi_{2009}^i = 0$ ) with positive dieback observation in 2010 (i.e.  $p_{2010}^i > 0$ ) have to satisfy  $q_{2010}^i = \tilde{q}_{2010}^i > 0$  (see Eq. (8)) which implies, for the temperature index T28, that  $\gamma > 19.5$ .

For five parameters, the convergence is very quick and seems independent from the four other parameters. In fact, we observe a fast convergence of the probability of the symptom persistence  $C_{\text{pers}}$  towards 1 (Figure 9i). In addition, the parameters  $\gamma$ ,  $\kappa$ ,  $S$  and  $C_{\text{init}}$  quickly converge around fixed values or in restricted ranges of values (Figures 9e, 9f, 9g and 9h).

The parameter  $\beta_1$  reaches values such that the contribution of the term  $\beta_1 h_a$  in  $H_a$  defined by (2) is small with respect to  $\beta_0$ . In fact, Figure 9c represents the evolution of  $\max_{a,i} h_a^i \frac{\beta_1}{\beta_0}$ , where  $\max_{a,i} h_a^i = 322.2$  is the maximum of the local rainfall in June in France between 2008 and 2019 (i.e. the period of data used for the estimation). This extreme value is reached in the South East of France in 2010 (see Figure 18b). In Figure 9c, all runs reach values below 0.06 suggesting that  $\beta_1 h_a$  is neglectable with respect to  $\beta_0$ .

Even if a range of values of parameters  $D$ ,  $\beta_0$  and  $r$  seems to emerge for the runs with the best likelihoods, the convergence of these parameters is not clear. A rough estimation of these parameters is done in the next session.

Finally this step allows to reduce the model setting  $C_{\text{pers}} = 1$  and  $\beta_1 = 0$  and to identify suitable zones for the parameters  $\gamma$ ,  $\kappa$ ,  $S$  and  $C_{\text{init}}$ .

## A.2 Log-likelihood on a coarse grid

Following the model reduction given by the previous section, we then compute the log-likelihood on a coarse grid in order to catch a suitable zone for the parameter initialization of the AMIS algorithm 1. Without idea of the scale of the parameters (particulary for  $\beta_0$  and  $r$ ), we used a large coarse grid that we refined toward the identification of the best zone of parameters.

Figure 10a represents, for each fixed parameters  $\beta_0$  and  $D$  on a (refined) coarse grid, the maximal log-likelihood computed for different values of  $r$ ,  $\gamma$ ,  $\kappa$ ,  $S$  and  $C_{\text{init}}$ , that is

$$\max_{r \in \mathcal{I}_r; \gamma \in \mathcal{I}_\gamma; \kappa \in \mathcal{I}_\kappa; S \in \mathcal{I}_S; C_{\text{init}} \in \mathcal{I}_{C_{\text{init}}}} \{\mathcal{L}(D, \beta_0, \beta_1 = 0, r, \gamma, \kappa, S, C_{\text{init}}, r_S = 1000, C_{\text{pers}} = 1)\} \quad (16)$$

where

$$\begin{aligned} \mathcal{I}_r &= \{0, 10^{-10}, 10^{-9}, \dots, 10^4, 10^5\}, & \mathcal{I}_\gamma &= \{21.001, 21.501, \dots, 23.001\}, \\ \mathcal{I}_\kappa &= \{0.025, 0.05, \dots, 0.25\}, & \mathcal{I}_S &= \{80, 90, 100\}, \\ \mathcal{I}_{C_{\text{init}}} &= \{0.008, 0.0085, 0.009\} \end{aligned} \quad (17)$$

are the discretisation points on the coarse grid for the  $r$ ,  $\gamma$ ,  $\kappa$ ,  $S$  and  $C_{\text{init}}$  components respectively. Note that the chose of the set  $\mathcal{I}_\gamma$  is due to some irregularities of the log-likelihood in the variable  $\gamma$  (see Section A.3).

The maximal log-likelihood on the coarse grid equals -24727.26 and is obtained for the set of parameters

$$\begin{aligned} D &= 18; & \beta_0 &= 24; & \kappa &= 0.05; & \gamma &= 21.501 \\ S &= 90; & C_{\text{init}} &= 0.0085; & r &= 10^{-8}. \end{aligned}$$

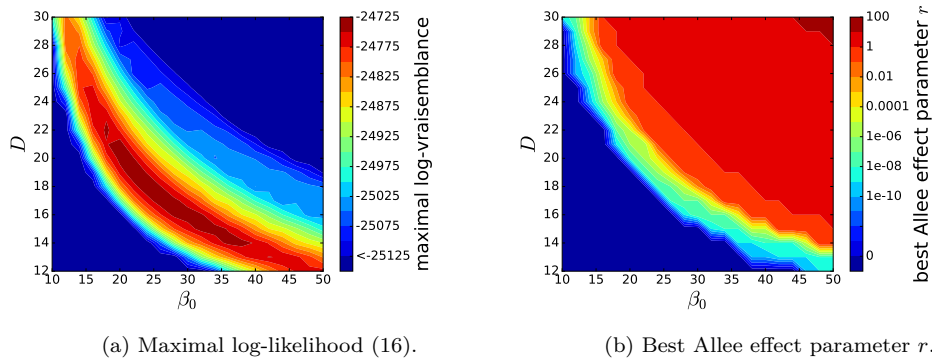


Figure 10: Maximal log-likelihood (16) (left) and value of the Allee effect parameter  $r$  leading to this maximal log-likelihood (right) for the coarse grid (17),  $\beta \in \{10, 12, \dots, 50\}$ ,  $D \in \{12, 13, \dots, 30\}$  and with the model reduction  $\beta_1 = 0$ ,  $C_{\text{pers}} = 1$  and  $r_S = 1000$ .

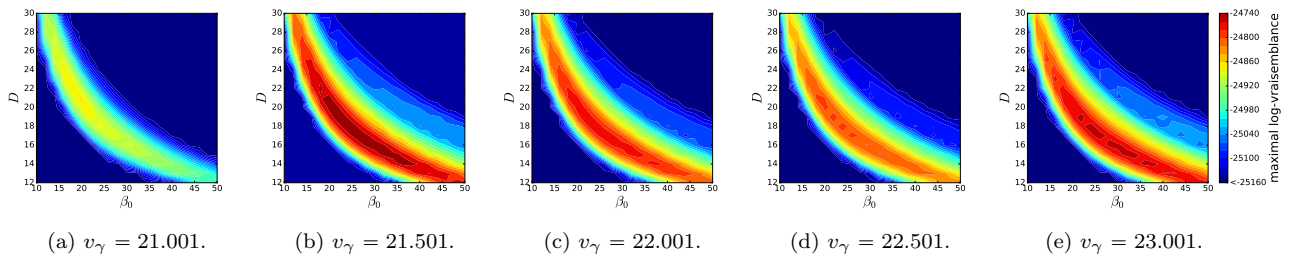


Figure 11: Maximal log-likelihood (16) for the coarse grid (17) except for the parameter  $\gamma$  which is fixed to the value  $v_\gamma$  and with the model reduction  $\beta_1 = 0$ ,  $C_{\text{pers}} = 1$  and  $r_S = 1000$ .

Moreover, we identify a banana shape zone for the choice of the parameters  $\beta_0$  and  $D$  (see Figure 10a).

Figure 10b represents the value of the Allee effect parameter  $r$  leading to this maximal log-likelihood, that is the argument of the maximum in  $\mathcal{T}_r$  of the maximal log-likelihood (16). Note that this figure is represented on a kind of log-scale. We observe that the previous banana shape zone is included in the region where the Allee effect parameter  $r$  is less than 0.0001 and negligible. In fact although the maximal value of the log-likelihood is obtained for  $r = 10^{-8}$ , setting  $r = 0$  instead of  $10^{-8}$  makes the log-likelihood goes from -24727.26054 to -24727.26067.

Figure 11 represents the maximal log-likelihood with respect to  $\beta_0$  and  $D$ , as in Figure 10a, but for different fixed values of  $\gamma$ . We can notice that the maximal likelihood is non-monotonic in  $\gamma$ . However, we identify that the best likelihood is for  $\gamma \approx_{>} 21.5$ .

Finally, this step leads to the model reduction  $r = 0$ . Moreover, we identified an initialisation zone for the parameters  $D$  and  $\beta_0$  as well as refined zone for the initialisation of  $\gamma$ .

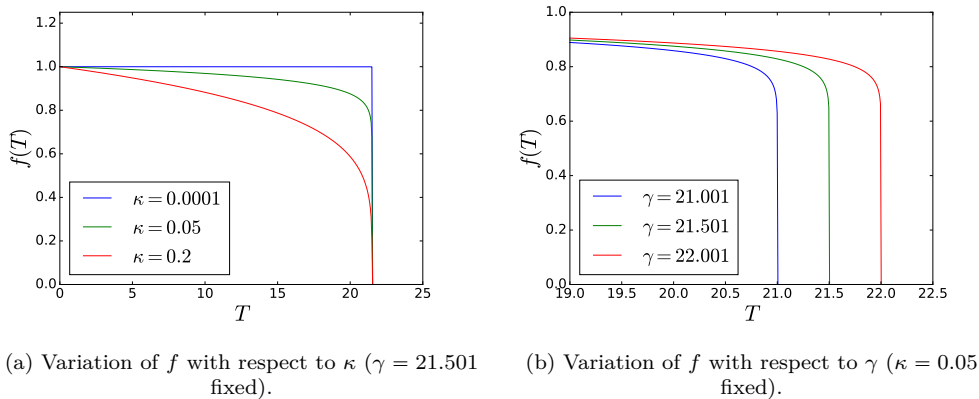


Figure 12: Function of the temperature impact  $f$  defined by (7) for different values of  $\gamma$  and  $\kappa$ .

### A.3 An almost degenerated posterior distribution for $\gamma$

Further to the reduction model and the first rough estimations of parameters induced by the steps of Sections A.1 and A.2, we run an AMIS algorithm for the parameters estimations (see Cornuet et al. (2012)). We do not give details here on the parameters of the algorithm, however we refer to Section 4.3.2 where the algorithm and the parameters are given for the final estimation step (after the reduction model given in this section).

Even if we observe a fast convergence of the support of the one-dimensional posterior distributions as well as their rough shapes, the stabilisation of the distributions was not observed after 150 runs of the AMIS algorithm. The slow convergence is due, in particular, to the almost degenerated and non-symmetric form of the posterior distribution for the parameters  $\gamma$  which weights values very close to (but larger than) 21.5. In fact sampled parameters, generated by normal proposal distributions by the AMIS algorithm, contain a large number of “non appropriated” parameters.

The non regularity of the posterior distribution of the parameter  $\gamma$  in 21.5 is due to the form of the function of the temperature impact  $f$  given by (7). In Figure 9f, we observe that the parameter  $\kappa$  converges toward small values (less than 0.2), and in fact around 0.05 after convergence in the best zone (see Section A.2 and runs of Figure 9f with the best log-likelihoods). Then, for the best zone of parameters, the function of the temperature impact  $T \mapsto f(T)$  is abrupt close to  $T = \gamma$  (see Figure 12).

Moreover, due to the discrete nature of the temperature data, which are a mean of four entire values (see Section D), the log-likelihood contains some irregularities in these discrete values, in particular in  $\gamma = 21.5$ .

By the previous arguments (almost degenerated posterior distribution and best log-likelihood for  $\gamma$  very close to, but strictly larger than 21.5), we reduce the model setting  $\gamma = 21.50001$ .

## B Convergence of the AMIS algorithm

We discuss, in this section, the convergence of the AMIS Algorithm 1 which led to the numerical estimation of  $(D, \beta_0, \kappa, S, C_{\text{init}})$  presented in Section 5. Following Abboud et al. (2019), we compute the deviation measure between two consecutive iterations  $k - 1$  and  $k$  of the algorithm, defined for a partition  $\mathcal{P}$  of the space of parameters  $\mathbb{R}_+^5$  by

$$\mathcal{D}(\mathcal{P}, k - 1, k) = \max_{c \in \mathcal{P}} |\rho_k(c) - \rho_{k-1}(c)| \quad (18)$$

with  $\rho_k(c)$  the estimated probability of the element  $c$  of the partition  $\mathcal{P}$  given by

$$\rho_k(c) = \sum_{i=1}^k \sum_{\ell=1}^N w_{\ell i} 1_{\theta_{\ell i} \in c}$$

where  $\theta_{\ell i}$  and  $w_{\ell i}$  are the parameters sampled by the algorithm and their weights respectively.

Figure 13 represents the deviation measure for five partitions, illustrating the convergence of the five one-dimensional posterior distributions. For example, the first picture on Figure 13 is the deviation measure for a partition  $\mathcal{P}_{\beta_0} = ([n c_\beta, (n + 1)c_\beta]_{n \in \mathbb{N}} \times \mathbb{R}^4)$  with  $c_\beta$  small and illustrates the convergence of the marginal posterior distribution of the parameter  $\beta_0$ . Although there are not presented here, we also checked the convergence of the 2-dimensional posterior distributions. We observe that the convergence of the AMIS algorithm occurs in about 10 iterations.

Importantly, notice that the convergence of the AMIS algorithm toward the posterior distribution is not proved. In fact, although we observe a convergence of our algorithm, it is not excluded that it converges toward parameters for which the log-likelihood are around a local maximum. To restrict this effect, the computation of the log-likelihood on a coarse grid described in Section A.2 have been done first for a large scale of parameters values, before to be refined on the coarse grid presented in Section A.2.

## C Model behavior for different temperature indices

We tested five different temperature indices for the choice of the temperature variable  $(T_a^i)_{a,i}$  impacting the symptoms development (see Eq. (4)): T24, T26, T28, T30 and T35, denoting the mean number of days, between July 1 and August 31, for which the daily maximal temperature exceeds 24°C, 26°C, 28°C, 30°C and 35°C respectively. The threshold of 30°C and 35°C were chosen because Hauptman et al. (2013) showed that *H. fraxineus* survives only few hour in infected ash shoots at temperature above 35°C. The temperature of 28°C is the maximal temperature where the fungus grows *in vitro* (Hauptman et al., 2013). However, leaves temperature often exceeds air temperature by several degrees (Granier et al., 2007) and we thus tried lower temperature thresholds (24°C and 26°C).

Figure 14 shows the results obtained with the different temperature indices tested. Quantitatively, all parameters indices give comparative log-likelihoods. We observed however that the best log-likelihood is obtained for the temperature index T26, for the period 2008-2019 (period of the data used for the estimation) but that T28 gives a better likelihood for the

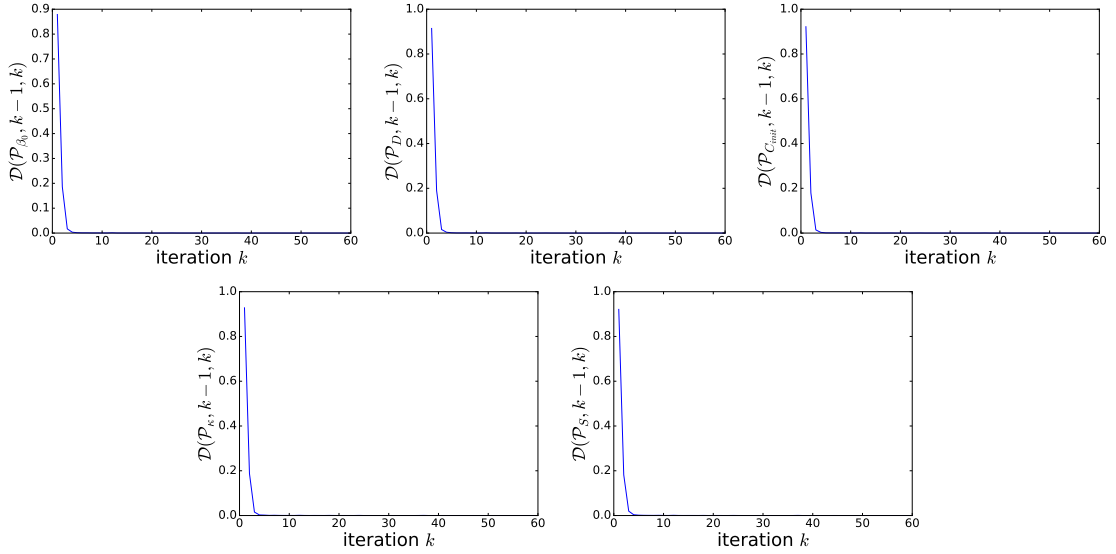


Figure 13: Deviation measure (18) for five partitions  $\mathcal{P}_{\beta_0}$ ,  $\mathcal{P}_D$ ,  $\mathcal{P}_{C_{\text{init}}}$ ,  $\mathcal{P}_\kappa$ ,  $\mathcal{P}_S$  illustrating the convergence of the five one-dimensional distributions.

period 2008-2023 (where 2020-2023 is the predictive period). The temperature index T35 gives the worst log-likelihood. Qualitatively, contrary to other temperature indices, the index T35 allows to describe the limited expansion neither in the southeast of France nor in the Garonne valley. The limited expansion in the Garonne valley seems less strong for the temperature index T30 than T24, T26 and T28.

Figure 15 shows the marginal posterior distributions estimated by Algorithm 1 for the different temperature indexes. We first observe that the posterior distribution of the parameter  $C_{\text{init}}$  is almost the same for the five temperature indices. This parameter, in fact, describes the intensity of the prevalence of the dieback symptoms for the years 2008 and 2009, given by (9), and does not directly depend on the temperature index. The posterior distributions of parameters  $\beta_0$  and  $D$  are with relatively similar supports. Therefore, although the distributions are quite different, all temperature indexes lead to the same order for the intensity of spore production  $\beta_0$  as well as the diffusion coefficient  $D$ . The posterior distributions of the parameter  $\gamma$  can obviously not be compared as they represent different quantities. However, contrary to the temperature index T28 which led to an almost degenerated distribution to  $\gamma$  (see Section A.3), we obtained a posterior distribution with density, although spiky, for other temperature indexes.

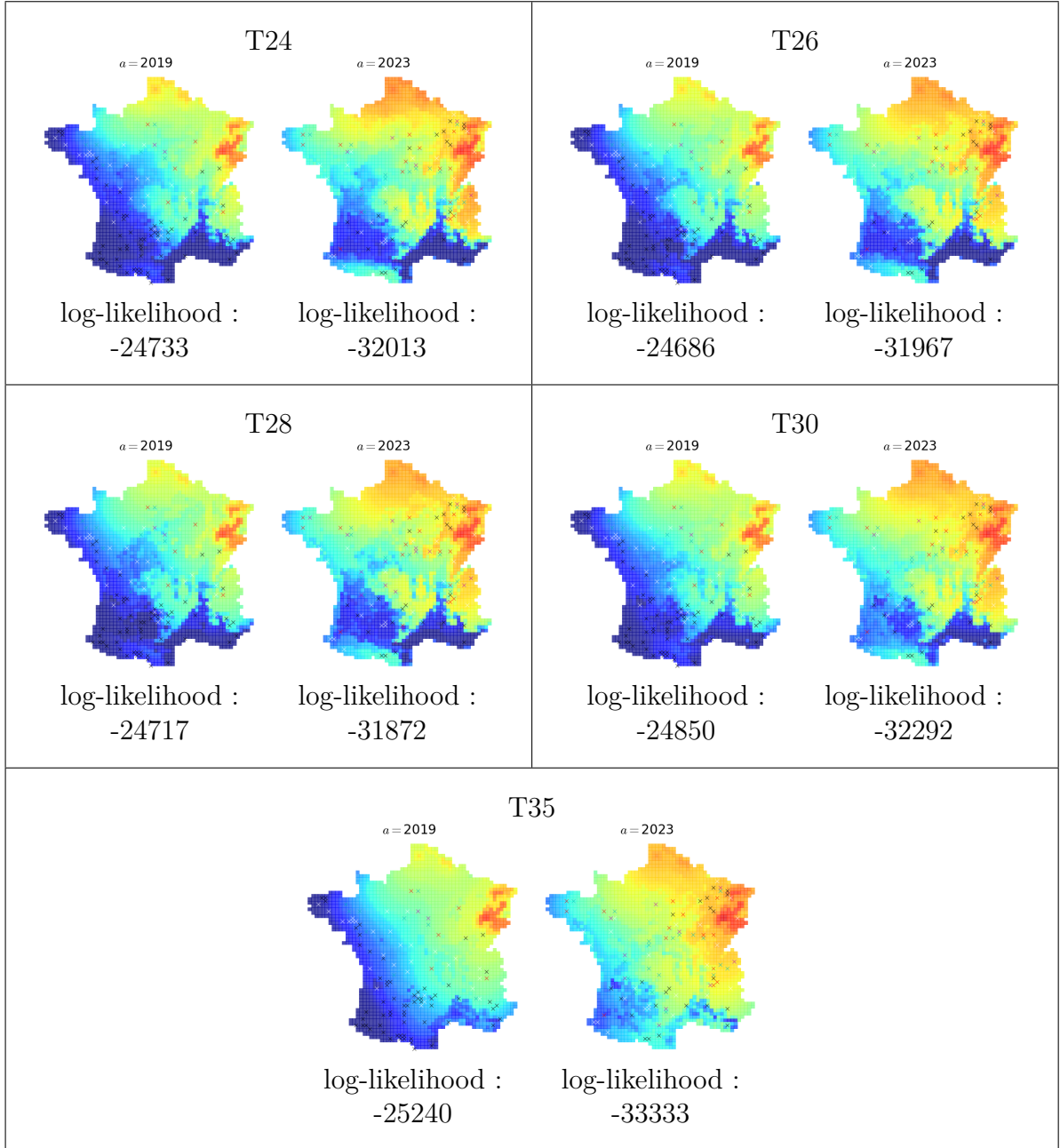


Figure 14: Bernoulli parameters (8) (heat map) and log-likelihood at the years 2019 and 2023 for the set of parameters which maximizes the likelihood and for the temperature indices T24, T26, T28, T30 and T35, as well as the observation data (crosses) of both years (see Figure 5 for the color coding).

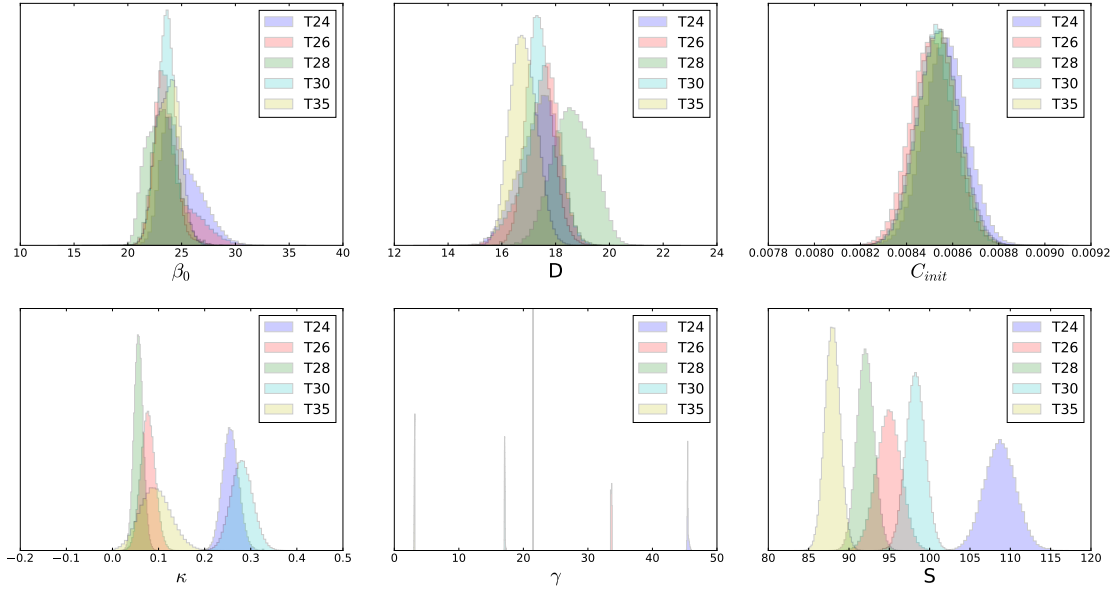


Figure 15: Marginal posterior distributions estimated by Algorithm 1 for the different temperature indices T24, T26, T28, T30 and T35.

## D Data visualisation

This section describes different data used for the inference. We give first the variable of interest that is the proportion of infected trees from 2008 to 2023 (Figure 16). Density of ashes in France is given in Figure 17. The density values are derived from the IGN data from 2006 to 2015 (<https://inventaire-forestier.ign.fr/dataifn/>). The meteorological data used are Safran data from Météo-France. Safran data are computed on a 8 x 8 km grid over France (Quintana-Seguí et al., 2008). We computed relevant meteorological variables and then averaged them over the 16 x 16 km quadrat. Each data is then a mean of four (occasionally one, two or three, on the border of France) meteorological stations located on the quadrat. Figures 18 and 19 describe the June rainfall and the number of days with temperature above 28°C in July and August, respectively.

## Acknowledgments

We acknowledge Matthieu Brachet for our discussions about the numerical scheme of the reaction diffusion model. We also thanks Claude Husson, from the Département de la Santé des Forêts (DSF) for input on factors to take into account in the model and for help to revise the manuscript. This work was supported by grants from the DSF, French Ministry of Agriculture and Forestry and from ANSES. The UMR1136 research unit is supported by a grant managed by the French National Research Agency (ANR) as part of the “Investissements d’Avenir” program (ANR-11-LABX-0002-01, Laboratory of Excellence ARBRE).



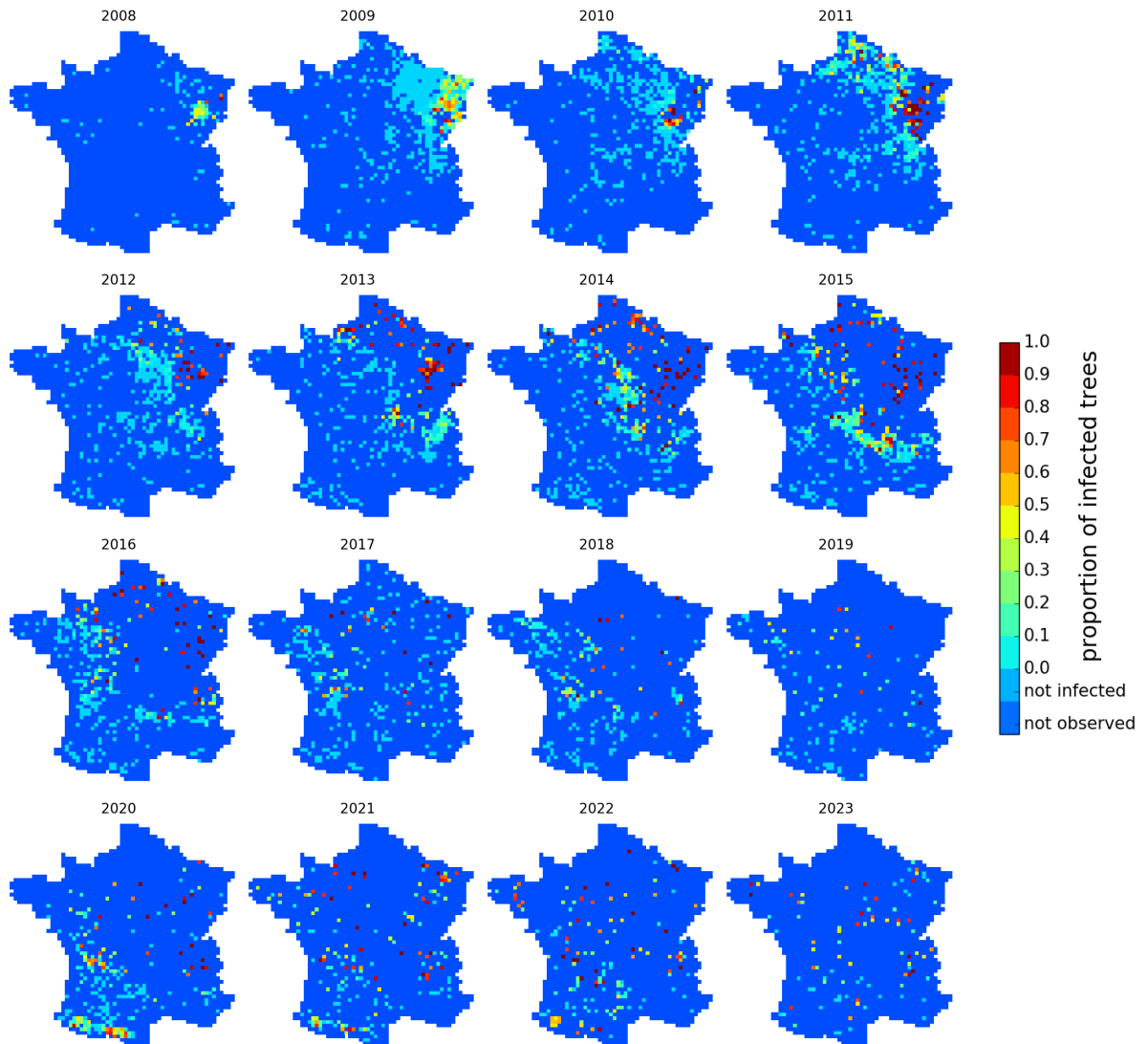


Figure 16: Proportion of infected trees among the observed trees from 2008 to 2023. Dark blue quadrats are non observed quadrats. Medium blue quadrats are non infected observed quadrats. From light blue to red quadrats: few infected to strongly infected observed quadrats.

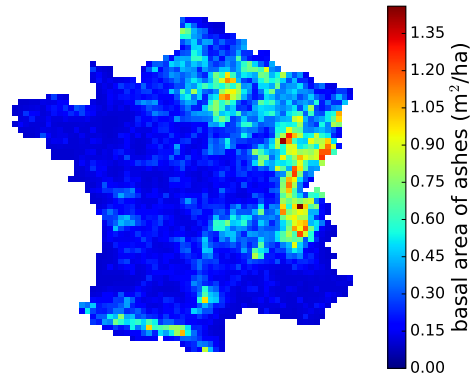


Figure 17: Basal area of ashes (*F. excelsior* and *F. angustifolia*) in France.

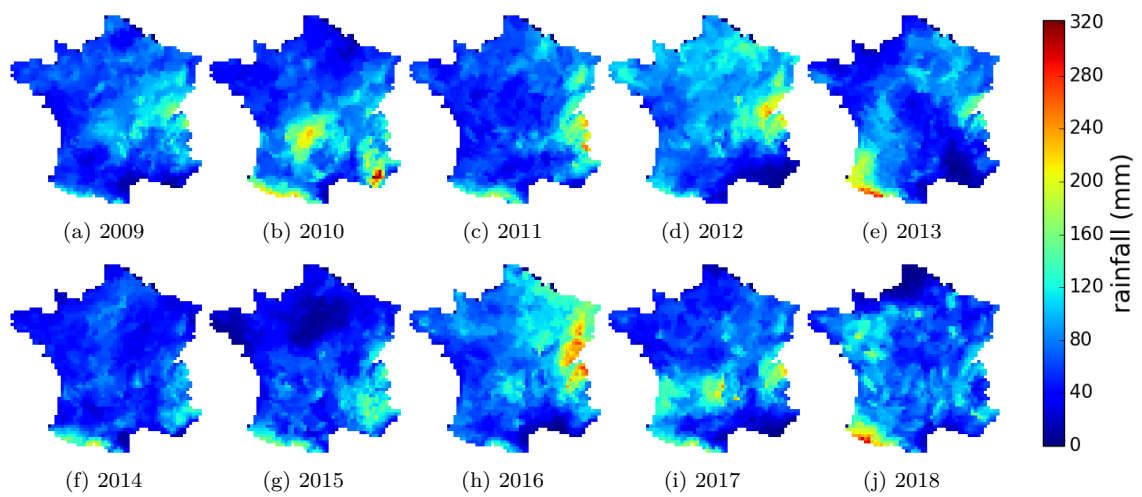


Figure 18: Rainfall of June (sum of the solid and liquid precipitation) from 2009 to 2018. (Source: Safran data from Météo-France).

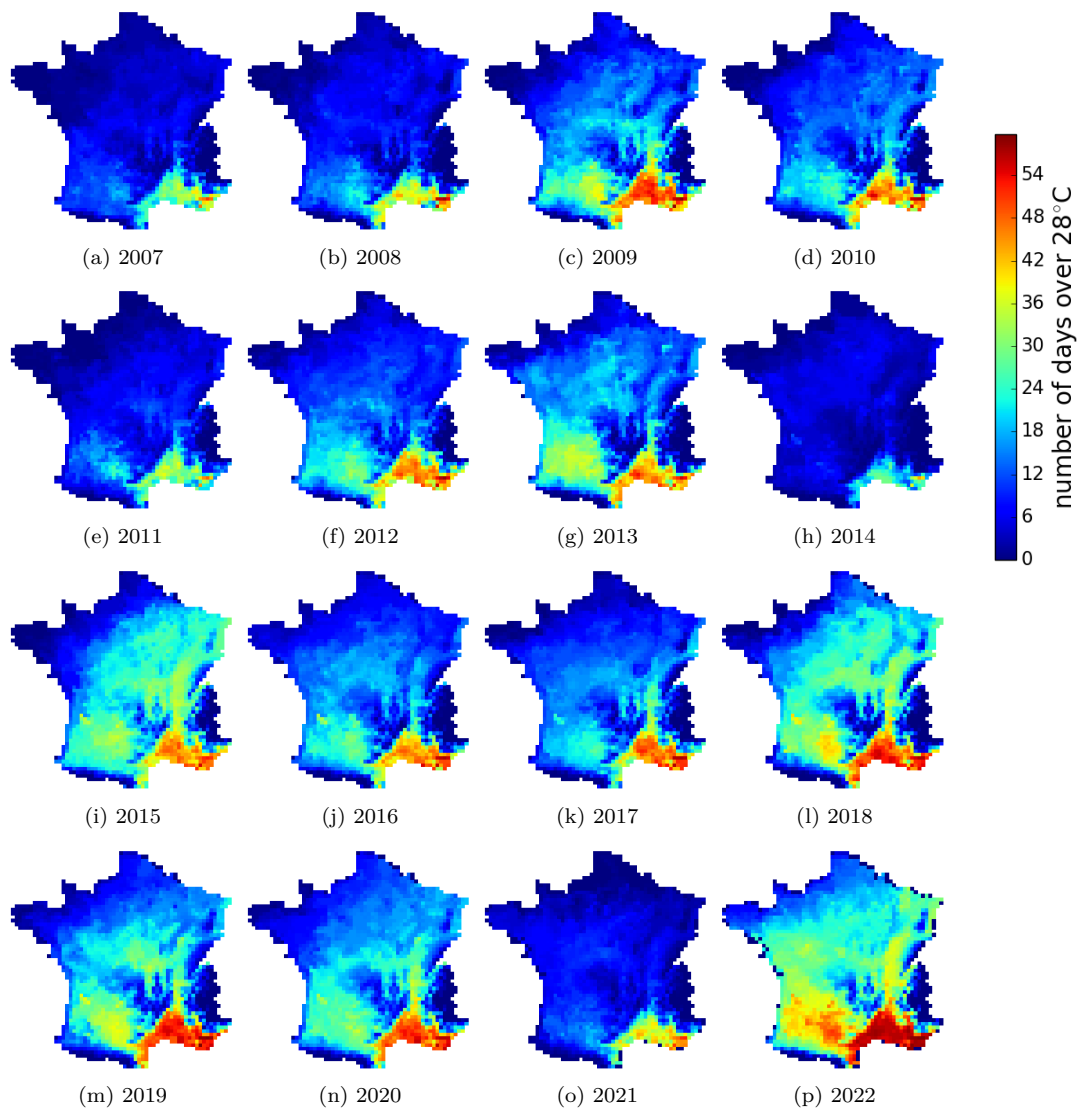


Figure 19: Number of days with maximal temperature over 28°C during the summer (July and August) in France from 2009 to 2022. (Source: Safran data from Météo-France).

## References

- Abboud, C., Bonnefon, O., Parent, E., and Soubeyrand, S. (2019). Dating and localizing an invasion from post-introduction data and a coupled reaction–diffusion–absorption model. Journal of mathematical biology, 79:765–789.
- Abboud, C., Parent, E., Bonnefon, O., and Soubeyrand, S. (2023). Forecasting pathogen dynamics with bayesian model-averaging: Application to xylella fastidiosa. Bulletin of Mathematical Biology, 85(7):67.
- Agan, A., Tedersoo, L., Drenkhan, R., and Hanso, M. (2022). Traces of *Hymenoscyphus fraxineus* in North-Eastern Europe Extend Further Back to History than Expected. Plant Disease, pages PDIS–04–22–0807–RE.
- Baxter, E., Cooke, L. R., Spaans, F., Grant, I. R., and McCracken, A. R. (2023). The introduction of *Hymenoscyphus fraxineus* to Northern Ireland and the subsequent development of ash dieback. Forest Pathology.
- Bebber, D. P. (2015). Range-Expanding Pests and Pathogens in a Warming World. Annual Review of Phytopathology, 53(1):335–356.
- Chumanová, E., Romportl, D., Havrdová, L., Zahradník, D., Pešková, V., and Černý, K. (2019). Predicting ash dieback severity and environmental suitability for the disease in forest stands. Scandinavian Journal of Forest Research, 34:254–266.
- Coker, T. L. R., Rozsypálek, J., Edwards, A., Harwood, T. P., Butfoy, L., and Buggs, R. J. A. (2019). Estimating mortality rates of European ash (*Fraxinus excelsior*) under the ash dieback (*Hymenoscyphus fraxineus*) epidemic. Plants, People, Planet, 1(1):48–58.
- Cornuet, J.-M., Marin, J.-M., Mira, A., and Robert, C. P. (2012). Adaptive multiple importance sampling. Scandinavian Journal of Statistics, 39(4):798–812.
- Cross, H., Sønstebo, J. H., Nagy, N. E., Timmermann, V., Solheim, H., Børja, I., Kauserud, H., Carlsen, T., Rzepka, B., Wasak, K., Vivian-Smith, A., and Hietala, A. M. (2017). Fungal diversity and seasonal succession in ash leaves infected by the invasive ascomycete *Hymenoscyphus fraxineus*. New Phytologist, 213:1405–1417.
- Desprez-Loustau, M.-L., Aguayo, J., Dutech, C., Hayden, K. J., Husson, C., Jakushkin, B., Marçais, B., Piou, D., Robin, C., and Vacher, C. (2016). An evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. Annals of Forest Science, 73(1):45–67.
- Dvořák, M., Rotkova, G., and Botella, L. (2016). Detection of Airborne Inoculum of *Hymenoscyphus fraxineus* and *H. albidus* during Seasonal Fluctuations Associated with Absence of Apothecia. Forests, 7(12).
- Dvořák, M., Štoidl, P., and Rost, M. (2023). Vertical spread of *Hymenoscyphus fraxineus* propagules. NeoBiota, 84:231–246.

- Eschen, R., Kadzamira, M., Stutz, S., Ogunmodede, A., Djeddour, D., Shaw, R., Pratt, C., Varia, S., Constantine, K., and Williams, F. (2023). An updated assessment of the direct costs of invasive non-native species to the United Kingdom. Biological Invasions, 25(10):3265–3276.
- Ghelardini, L., Pepori, A. L., Luchi, N., Capretti, P., and Santini, A. (2016). Drivers of emerging fungal diseases of forest trees. Forest Ecology and Management, 381:235–246.
- Gomez-Gallago, M. and Marçais, B. (2022). Semi-mechanistic modelling of tree mortality induced by an invasive pathogen. IUFRO division 7 conference, Lisbon, Portugal.
- Granier, A., Reichstein, M., Bréda, N., Janssens, I., Falge, E., Ciais, P., Grünwald, T., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Facini, O., Grassi, G., Heinesch, B., Ilvesniemi, H., Keronen, P., Knohl, A., Köstner, B., Lagergren, F., Lindroth, A., Longdoz, B., Loustau, D., Mateus, J., Montagnani, L., Nys, C., Moors, E., Papale, D., Peiffer, M., Pilegaard, K., Pita, G., Pumpanen, J., Rambal, S., Rebmann, C., Rodrigues, A., Seufert, G., Tenhunen, J., Vesala, T., and Wang, Q. (2007). Evidence for soil water control on carbon and water dynamics in european forests during the extremely dry year: 2003. Agricultural and Forest Meteorology, 143(1):123–145.
- Grosdidier, M., Ios, R., Husson, C., Cael, O., Scordia, T., and Marçais, B. (2018a). Tracking the invasion: dispersal of *Hymenoscyphus fraxineus* airborne inoculum at different scales. FEMS Microbiology Ecology, 94:fiy049.
- Grosdidier, M., Ios, R., and Marçais, B. (2018b). Do higher summer temperatures restrict the dissemination of *Hymenoscyphus fraxineus* in France? Forest Pathology, 48:e12426.
- Grosdidier, M., Scordia, T., Ios, R., and Marçais, B. (2020). Landscape epidemiology of ash dieback. Journal of Ecology, 108:1789–1799.
- Gross, A. and Holdenrieder, O. (2013). On the longevity of *Hymenoscyphus pseudoalbidus* in petioles of *Fraxinus excelsior*. Forest Pathology, 43:168–170.
- Gross, A., Holdenrieder, O., Pautasso, M., Queloz, V., and Sieber, T. N. (2014). *Hymenoscyphus pseudoalbidus*, the causal agent of European ash dieback. Molecular Plant Pathology, 15(1):5–21.
- Gross, A., Zaffarano, P., Duo, A., and Grünig, C. (2012). Reproductive mode and life cycle of the ash dieback pathogen *Hymenoscyphus pseudoalbidus*. Fungal Genetics and Biology, 49:977–986.
- Hamelin, F. M., Castella, F., Doli, V., Marçais, B., Ravigné, V., and Lewis, M. A. (2016). Mate Finding, Sexual Spore Production, and the Spread of Fungal Plant Parasites. Bulletin of Mathematical Biology, 78(4):695–712.
- Hauptman, T., Piškur, B., de Groot, M., Ogris, N., Ferlan, M., and Jurc, D. (2013). Temperature effect on *Chalara fraxinea*: heat treatment of saplings as a possible disease control method. Forest Pathology, 43:360–370.

- Havrdová, L., Zahradník, D., Romportl, D., Pešková, V., and Černý, K. (2017). Environmental and Silvicultural Characteristics Influencing the Extent of Ash Dieback in Forest Stands. Baltic Forestry, 23:168–181.
- Hudgins, E. J., Liebhold, A. M., and Leung, B. (2017). Predicting the spread of all invasive forest pests in the United States. Ecology Letters, 20(4):426–435.
- Ioos, R., Kowalski, T., Husson, C., and Holdenrieder, O. (2009). Rapid in planta detection of *Chalara fraxinea* by a real-time PCR assay using a dual-labelled probe. European Journal of Plant Pathology, 125(2):329–335.
- Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., Hudson, P., Jolles, A., Jones, K. E., Mitchell, C. E., Myers, S. S., Bogich, T., and Ostfeld, R. S. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature, 468(7324):647–652.
- Kirisits, T. (2015). Ascocarp formation of *Hymenoscyphus fraxineus* on several-year-old pseudosclerotial leaf rachises of *Fraxinus excelsior*. Forest Pathology, 45(3):254–257.
- Kirisits, T. and Freinschlag, C. (2012). Ash dieback caused by *Hymenoscyphus pseudoalbidus* in a seed plantation of *Fraxinus excelsior* in Austria. Journal of Agricultural Extension and Rural Development, 4:184 – 191.
- Kowalski, T. (2006). *Chalara fraxinea* sp. nov. associated with dieback of ash (*Fraxinus excelsior*) in Poland. Forest Pathology, 36(4):264–270.
- Kowalski, T. and Holdenrieder, O. (2009). The teleomorph of *Chalara fraxinea*, the causal agent of ash dieback. Forest Pathology, 39:304–308.
- Laubray, S., Buée, M., and Marçais, B. (2023). Evidence of a component Allee effect for an invasive pathogen: *Hymenoscyphus fraxineus*, the ash dieback agent. Biological Invasions, 25:2567–2582.
- Laubray, S., Buée, M., and Marçais, B. (2024). *Hymenoscyphus fraxineus* persistence in the ash litter. Plant Pathology, (in press).
- Marçais, B., Giraudel, A., and Husson, C. (2023). Ability of the ash dieback pathogen to reproduce and to induce damage on its host are controlled by different environmental parameters. PLOS Pathogens, 19:e1010558.
- Marçais, B., Kosawang, C., Laubray, S., Kjær, E., and Kirisits, T. (2022). Ash dieback. In Asiegbu, F. O. and Kovalchuk, A., editors, Forest Microbiology, pages 215–237. Elsevier.
- Meentemeyer, R. K., Haas, S. E., and Václavík, T. (2012). Landscape Epidemiology of Emerging Infectious Diseases in Natural and Human-Altered Ecosystems. Annual Review of Phytopathology, 50(1):379–402.
- Mollot, G., Pantel, J., and Romanuk, T. (2017). The Effects of Invasive Species on the Decline in Species Richness. Advances in Ecological Research, 56:61–83.

- Paap, T., Wingfield, M. J., Burgess, T. I., Wilson, J. R. U., Richardson, D. M., and Santini, A. (2022). Invasion Frameworks: a Forest Pathogen Perspective. Current Forestry Reports, 8(1):74–89.
- Pyšek, P., Hulme, P. E., Simberloff, D., Bacher, S., Blackburn, T. M., Carlton, J. T., Dawson, W., Essl, F., Foxcroft, L. C., Genovesi, P., Jeschke, J. M., Kühn, I., Liebhold, A. M., Mandrak, N. E., Meyerson, L. A., Pauchard, A., Pergl, J., Roy, H. E., Seebens, H., Van Kleunen, M., Vilà, M., Wingfield, M. J., and Richardson, D. M. (2020). Scientists’ warning on invasive alien species. Biological Reviews, 95(6):1511–1534.
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S. (2008). Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France. Journal of Applied Meteorology and Climatology, 47(1):92–107.
- 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.
- Santini, A., Ghelardini, L., De Pace, C., Desprez-Loustau, M. L., Capretti, P., Chandelier, A., Cech, T., Chira, D., Diamandis, S., Gaitniekis, T., Hantula, J., Holdenrieder, O., Jankovsky, L., Jung, T., Jurc, D., Kirisits, T., Kunca, A., Lygis, V., Malecka, M., Marcais, B., Schmitz, S., Schumacher, J., Solheim, H., Solla, A., Szabò, I., Tsopelas, P., Vannini, A., Vettraino, A. M., Webber, J., Woodward, S., and Stenlid, J. (2013). Biogeographical patterns and determinants of invasion by forest pathogens in Europe. New Phytologist, 197(1):238–250.
- Sikes, B. A., Bufford, J. L., Hulme, P. E., Cooper, J. A., Johnston, P. R., and Duncan, R. P. (2018). Import volumes and biosecurity interventions shape the arrival rate of fungal pathogens. PLoS biology, 16(5):e2006025.
- Timmermann, V., Børja, I., Hietala, A. M., Kirisits, T., and Solheim, H. (2011). Ash dieback: pathogen spread and diurnal patterns of ascospore dispersal, with special emphasis on Norway. EPPO Bulletin, 41:14–20.
- Ward, S. F., Liebhold, A. M., and Fei, S. (2022). Variable effects of forest diversity on invasions by non-native insects and pathogens. Biodiversity and Conservation, 31(11):2575–2586.