

# A compartmental model for *Xylella fastidiosa* diseases with explicit vector seasonal dynamics

Àlex Giménez-Romero<sup>1</sup>, Eduardo Moralejo<sup>2</sup>, and Manuel A. Matías<sup>1</sup>

<sup>1</sup>*Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC, CSIC-UIB), Campus UIB, 07122 Palma de Mallorca, Spain.*

<sup>2</sup>*Tragsa, Passatge Cala Figuera 6, 07009 Palma de Mallorca, Spain.*

## Abstract

The bacterium *Xylella fastidiosa* (Xf) is mainly transmitted by the meadow spittlebug *Philaenus spumarius* in Europe, where it has caused significant economic damage to olive and almond trees. Understanding the factors that determine disease dynamics in pathosystems that share similarities can help to design control strategies focused on minimizing transmission chains. Here we introduce a compartmental model for Xf-caused diseases in Europe that accounts for the main relevant epidemiological processes, including the seasonal dynamics of *P. spumarius*. The model was confronted with epidemiological data from the two major outbreaks of Xf in Europe, the olive quick disease syndrome (OQDS) in Apulia, Italy, caused by the subspecies *pauca*, and the almond leaf scorch disease (ALSD) in Mallorca, Spain, caused by subspecies *multiplex* and *fastidiosa*. Using a Bayesian inference framework, we show how the model successfully reproduces the general field data in both diseases. In a global sensitivity analysis, the vector-to-plant and plant-to-vector transmission rates, together with the vector removal rate, were the most influential parameters in determining the time of the infectious host population peak, the incidence peak and the final number of dead hosts. We also used our model to check different vector-based control strategies, showing that a joint strategy focused on increasing the rate of vector removal while lowering the number of annual newborn vectors is optimal for disease control.

**Keywords:** Epidemiology, Population Biology, Modeling.

Mathematical and computational modeling in Ecology and, in particular, Epidemiology have been recently recognized as powerful approaches to guide empirical work and provide a framework for the synthesis, analysis and development of conservation plans and policy-making (Chew et al. 2014; Levin 1992; Murray 1989; Sarkar et al. 2006). Plant epidemics, mainly plant-virus diseases, have been often described by compartmental models, which deal with the overriding importance of transmission mechanisms in determining epidemic dynamics (Jeger et al. 2004; Jeger et al. 1998; Madden et al. 2000). These models have contributed to providing answers to some questions related to the ecology of plant diseases and have led to direct applications in disease control while guiding research directions (Jeger et al. 2019).

The emergence of vector-borne plant pathogens in new areas causing huge economic impacts, such as *Xylella fastidiosa* and the *Candidatus Liberibacter* spp. (Huanglongbing or citrus greening), has sparked interest in modeling vector-transmitted plant-disease epidemics (Chiyaka et al. 2012; Jeger et al. 2019). The vector-borne bacterium *X. fastidiosa* (Xf) is a multi-host pathogen endemic to the Americas that causes economically important diseases, mostly in woody crops (Hopkins et al. 2002). Xf is a genetically diverse species with three evolutionary well-defined clades forming the *pauca*, *fastidiosa*, and *multiplex* subspecies, native from South, Central, and North America, respectively (Vanhove et al. 2019). Within each subspecies, diverse genetic lineages with different host ranges are found. Xf is transmitted non-specifically by xylem-sap-feeding insects belonging to

the sharpshooter leafhoppers (Hemiptera: Cicadellidae, Cicadellinae) and spittlebugs (Hemiptera: Cercopoidae) (Redak *et al.* 2004).

Recently, Xf has gained renewed interest due to the massive mortality of olive trees in Apulia, Italy (Saponari *et al.* 2019). The first focus of the olive quick decline syndrome (OQDS) was detected in 2013 around Gallipoli (Apulia, Italy) (Saponari *et al.* 2013) and since then has spread throughout the region by the meadow spittlebug, *Philaenus spumarius*. Although this was the first official detection of Xf in Europe, it has recently been demonstrated that the pathogen arrived much earlier in Corsica (Soubeyrand *et al.* 2018) and in the Balearic islands (Moralejo *et al.* 2020). Around 1993, two strains of the subspecies *fastidiosa* (ST1) and *multiplex* (ST81) were introduced from California to Mallorca (Spain) with infected almond plants (Moralejo *et al.* 2020). To date, over 80% of the almond trees in Mallorca show leaf scorch symptoms and the outbreak has changed the iconic rural landscape of this Mediterranean island (Olmo *et al.* 2021).

The meadow spittlebug, *P. spumarius* (Hemiptera: Aphrophoridae), has recently been shown to be the main vector of Xf in Europe both in transmission experiments and in field studies (Cornara *et al.* 2017a; Cornara *et al.* 2018b; López-Mercadal *et al.* 2022; Moralejo *et al.* 2019; Saponari *et al.* 2019). *P. spumarius* is a polyphagous species from the Palearctic region, presenting one generation per year (univoltine) and overwintering as eggs. Foam-forming nymphs emerge at the end of winter, feeding on herbaceous plants. The time required for their development to the adult stage depends mainly on temperature and humidity (Bodino *et al.* 2019; Chmiel *et al.* 1979; Cornara *et al.* 2018a). In Mediterranean climates, *P. spumarius* adults generally move from the herbaceous cover to the crop canopy as evapotranspiration increases in late spring (May–June). In mid-summer, the populations of *P. spumarius* tend to decrease in the crop canopy, while the insects are captured more frequently in trees and shrubs interspersed in crops. Summer dispersal of spittlebugs to wild hosts as refugee seems a common general pattern in Mediterranean crops in Italy (Bodino *et al.* 2019; Cornara *et al.* 2021) and Spain (Morente *et al.* 2018). Because the bacterium has not been detected in spring on insects feeding on the herbaceous cover or in weeds in Europe (Bodino *et al.* 2019; Cornara *et al.* 2018a; Olmo *et al.* 2021), it is assumed that all spittlebug adults acquire the bacteria from the main crop (olive, almond, vine, etc.). Once infected, Xf colonizes the insect foregut in a persistent and non-circulatory manner without transovarial (parent to offspring) or transstadial (inter-stage) transmission (Almeida *et al.* 2003; Freitag 1951; Purcell *et al.* 1979) and without a period latency after vector acquisition (Almeida *et al.* 1987; Freitag 1951).

Several epidemic models have been already developed for Xf-diseases, but they lack a realistic description of some relevant processes (Jeger *et al.* 2019). Some of these models assume a simple general form for infected host dynamics (Abboud *et al.* 2019; Daugherty *et al.* 2019; White *et al.* 2017) or use a simplified S-I compartmental scheme for hosts, disregarding important features such as the latent period or the host mortality rate (Soubeyrand *et al.* 2018). Models that do take these features into account, however, do not explicitly model the population of vectors responsible for disease transmission (White *et al.* 2020). Other more recent models have taken a step further in explicitly modeling the vector population (Brunetti *et al.* 2020; Giménez-Romero *et al.* 2022b), but the characterization of its dynamics is still relatively simple, as it overlooks the known seasonal patterns of vector abundance. Several recent studies have provided new insights into the ecology and temporal dynamics of the transmission of Xf by *P. spumarius* in olive plants (Bodino *et al.* 2019; Bodino *et al.* 2021). However, these experimental data of the pathosystem have not been yet integrated at the population level. Thus, there is a need to continue advancing in the modeling of Xf diseases by developing more realistic models that can elucidate the fundamental processes involved in vector-host-pathogen interactions and help to design effective control strategies.

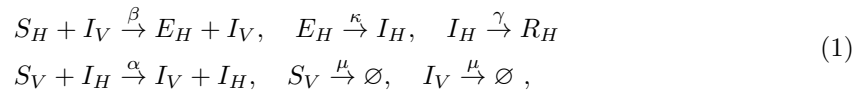
In this work, we develop a deterministic continuous-time compartmental model to describe the general epidemiological dynamics of diseases produced by Xf in Europe. We explicitly account for key biological aspects of the disease, including the seasonal dynamics of its main vector, *P. spumarius*. Our model is able to describe field data from the two major European outbreaks: the olive quick disease syndrome (OQDS) in Apulia, Italy, caused by the subspecies *pauca*, and the almond leaf scorch disease (ALSD) in Mallorca, Spain, caused by subspecies *multiplex* and *fastidiosa*. We aimed to find the most influential parameters in the model with respect to incidence and mortality in both diseases by performing a global sensibility analysis. With this information, the next goal was to explore control strategies acting especially on the vector population.

## MATERIALS AND METHODS

### Epidemic model: the SEIR-V model

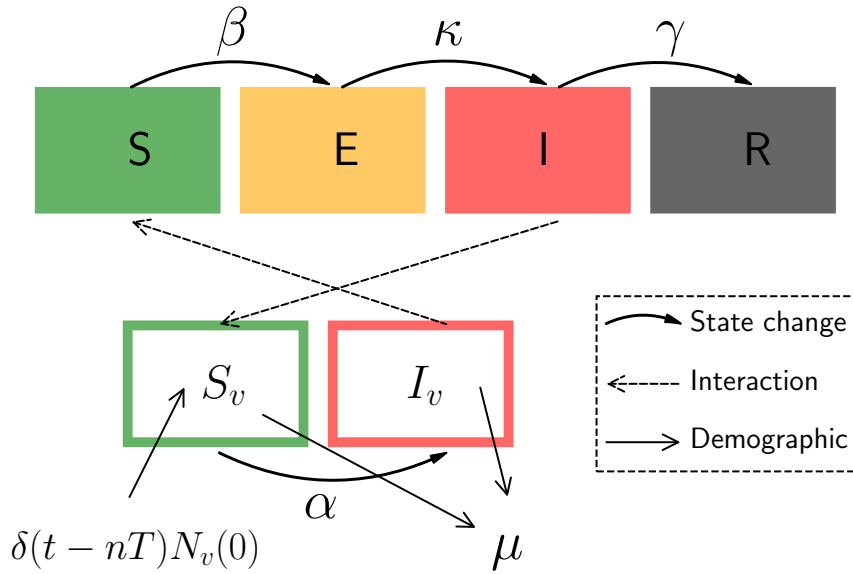
We developed a deterministic continuous-time compartmental model that incorporates the specific biological features of Xf diseases in Europe, including the dynamics of the main relevant vector *P. spumarius* (Cavalieri *et al.* 2019). To build the model we took the following considerations: (i) we assume there is no winter recovery of infected hosts and thus they die sometime after infection; (ii) hosts show an asymptomatic period in which they are non-infectious in practice (exposed compartment) because the bacteria are not yet systemically extended (Stevenson *et al.* 2004; Teviotdale *et al.* 2003), while vectors are infectious immediately after acquiring the bacterium (Fierro *et al.* 2019); (iii) vectors have an annual life cycle without mother-to-offspring disease transmission (Freitag 1951; Purcell *et al.* 1979), so we consider the annual emergence of susceptible newborn vectors and a constant death rate for both susceptible and infected vectors; (iv) infected vectors carry the bacterium during their entire lifespan without affecting their fitness; and finally, (v) we do not consider host recruitment or natural death given that the typical development time of Xf-epidemics is faster than the typical host's life cycle.

Altogether, our deterministic continuous-time compartmental model consists of six compartments, four describing the host population (susceptible,  $S_H$ , exposed,  $E_H$ , infectious,  $I_H$ , and removed,  $R_H$ ), and two describing the vector population (susceptible,  $S_V$ , and infected,  $I_V$ ). The model is defined according to the following processes,



which are illustrated in Fig. 1, being the birth of new susceptible vectors described as a source term. Thus, the host-vector compartmental model is written as,

$$\begin{aligned} \dot{S}_H &= -\beta S_H I_V / N_H \\ \dot{E}_H &= \beta S_H I_V / N_H - \kappa E_H \\ \dot{I}_H &= \kappa E_H - \gamma I_H \\ \dot{R}_H &= \gamma I_H \\ \dot{S}_V &= N_v(0) \sum_{n=1}^{\infty} \delta(t - nT) - \alpha S_V I_H / N_H - \mu S_V \\ \dot{I}_V &= \alpha S_V I_H / N_H - \mu I_V. \end{aligned} \quad (2)$$



**Figure 1:** Schematic representation of the model Eq. (2). Boxes are the compartments in which the population is divided, solid curved arrows represent changes in state, i.e. transitions between compartments, dashed arrows depict the crossed interaction between hosts and vectors and solid straight arrows represent demographic changes in vector population.

The model describes the exposure of susceptible hosts,  $S_H$ , at a rate  $\beta$  through their interaction with infected vectors,  $I_v$ , while susceptible vectors,  $S_v$ , get infected immediately at a rate  $\alpha$  through their interaction with infectious hosts  $I_H$ . Exposed hosts get infectious at rate  $\kappa$ , being the mean latent period  $\tau_E = 1/\kappa$ , while infectious hosts die at rate  $\gamma$ , having a mean infectious period of  $\tau_I = 1/\gamma$ . Infected vectors stay infected and infectious for the rest of their lifetime. Regarding the seasonal dynamics of vectors, we assume that new adults emerge synchronously each year in fields being all susceptible. This is represented by the term  $N_v(0) \sum_{n=1}^{\infty} \delta(t - nT)$  in Eq. (2), where  $T = 1 \text{ yr}$  is the period and  $\delta(t - nT)$  is Dirac delta function, and basically implements a yearly pulse of new vectors at a certain moment in the year. Vectors are removed (die, move to herbaceous vegetation and other non-host trees, exit the field, etc) at a given rate  $\mu$ , which we consider identical for susceptible and infected vectors. For simplicity, we consider that the quantity of annual newborn adults,  $N_v(0)$ , is constant. This outburst of new adults followed by an exponential decay resembles the temporal patterns on the abundance of *P. spumarius* observed in crop fields (Antonatos *et al.* 2021; Beal *et al.* 2021; Cornara *et al.* 2017a; López-Mercadal *et al.* 2021) (see Fig. S1).

In our model (Eq. (2)) the crossed nonlinear terms in  $\dot{S}_H$  and  $\dot{S}_v$ ,  $S_H I_v$  and  $S_v I_H$ , are divided by the total host population,  $N_H$ . Thus, the vector-to-plant infection process is modeled using mass action incidence, which is density dependent, while the plant-to-vector infection process is modeled using standard incidence, which is frequency dependent (Martcheva 2015). This implies that doubling the number of vectors in the crop field would double the number of resulting exposed (or infected) hosts, as this process is population-dependent (mass action incidence), while doubling the number of hosts would not result in more vectors per unit area being infected, as this process only depends on the contact probability, being frequency dependent (standard incidence). We think this is the most reasonable assumption because, for a given plantation framework, increasing the number of hosts is expected to also increase the area of the field, while the number of vectors is an independent quantity.

## Basic reproductive number

The basic reproductive number,  $R_0$ , of the model cannot be trivially computed using standard methods such as the Next Generation Matrix (NGM) (Diekmann *et al.* 2010), as there is no pre-pandemic fixed point in the system of differential equations Eq. (2). For periodically varying vector populations, rigorous methods have been developed (Bacaër 2007), but not for the case of growing or decaying vector populations. Here we use the simple method developed in the work of Giménez-Romero *et al.* 2022a (see Appendix II), which effectively computes the average number of secondary infections produced by an initially infectious individual in one generation. Thus, the effective basic reproductive number is given by

$$R_0 = \frac{\beta\alpha}{\mu\gamma} \frac{S_H(0)}{N_H^2} \frac{N_v(0)}{\mu\tau} (1 - e^{-\mu\tau}) , \quad (3)$$

where  $\tau$  corresponds to the time length of one generation, in our case one year. This  $R_0$  is calculated using the initial susceptible host population,  $S_H(0)$ . Below we will also use a time-dependent  $R_0(t)$  using  $S_H(t)$ .

## Epidemiological data

Epidemiological data from an ALSD outbreak in the island of Mallorca, Balearic Islands, Spain were taken from Moralejo *et al.* 2020. Dated phylogenetic analysis and estimates of disease incidence showed that the introduction of both subspecies occurred around 1993 and  $\sim 79\%$  almond trees were infected in 2017 (Moralejo *et al.* 2020). The annual proportion of infected individuals in the almond tree population between 1993 and 2017 was estimated by analyzing through qPCR the presence of Xf-DNA in the growth rings of 34 sampled trees (cf. Fig. 3 in (Moralejo *et al.* 2020)). The disease progression curve was estimated without distinguishing whether infections were caused by *multiplex* or *fastidiosa* subspecies. In addition, a two-sided bootstrap confidence interval for each data point was set using the SciPy bootstrap function in Python (Virtanen *et al.* 2020). On the other hand, epidemic data for OQDS were retrieved from (White *et al.* 2020). The data consisted of 2 to 3 yearly censuses of symptom prevalence in 17 olive groves infected with Xf subsp. *pauca* in Apulia, Italy, which were aggregated to fit our model as shown in Fig. 4 in (White *et al.* 2020). Because the compartments of our model are not in one-to-one correspondence with those shown in the work of White *et al.* (White *et al.* 2020), we used the sum of the symptomatic and desiccated infected trees in the dataset ( $I_S + I_D$ ) to fit the sum of the infected and dead trees ( $I + R$ ) and the sum of susceptible and asymptomatic hosts ( $S + I_A$ ) to fit the sum of susceptible and exposed hosts ( $S + E$ ). The processed data used to fit the model can be found in (Giménez-Romero 2022), while the raw data can be found in the supplementary data accessible online of the cited articles (Moralejo *et al.* 2020; White *et al.* 2020).

## Model fitting through Bayesian Inference

We employed an informative normal  $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$  prior distribution, with  $\hat{\mu}$  and  $\hat{\sigma}$ , the mean and standard deviation, respectively, for previously measured parameters in the literature, such as the infectious and latent periods for ALSD,  $\tau_I \sim \mathcal{N}(14, 4)$ ,  $\tau_E \sim \mathcal{N}(4, 1)$  (Moralejo *et al.* 2020; Teviotdale *et al.* 2003) and OQDS,  $\tau_I \sim \mathcal{N}(3.5, 1)$ ,  $\tau_E \sim \mathcal{N}(1.75, 0.5)$  (Fierro *et al.* 2019). The corresponding rates are given by  $\gamma = 1/\tau_I$  and  $\kappa = 1/\tau_E$ , respectively. Similarly, a prior normal distribution was used for the removal rate of vectors,  $\mu \sim \mathcal{N}(0.02, 0.0075)$ , as the mean value  $\mu = 0.02$  already captures the vector dynamics observed in field-data (Fig. S1). Regarding the prior distribution for the transmission rates a very wide and uninformative uniform prior distribution,  $\beta \sim \mathcal{U}(0.001, 1)$  and  $\alpha \sim \mathcal{U}(0.001, 1)$ , was used for each parameter. The number of hosts,  $N_H$ , was already provided in the datasets, while, given the lack of information about the vector population, we assumed  $N_v(0) = N_H/2$  for the initial vector population of each year. However, we tested the robustness of our results by changing  $N_v(0)$ .

The posterior distributions of the parameters were approximated using the Markov Chain Monte Carlo algorithm No U-Turn Sampler (NUTS) with the recommended target acceptance rate of 65% (Homan *et al.* 2014). To ensure a proper convergence, we constructed three independent Markov

Chains with  $10^5$  iterations each after a burn-in of  $10^4$  iterations and checked that the results were statistically equivalent. For each chain, we started at the maximum-likelihood parameters yielded by the Nelder-Mead algorithm with 1000 iterations.

The parameters of our compartmental model were determined by fitting the model to data by means of a Bayesian Inference framework using the Turing.jl package (Ge *et al.* 2018) in Julia (Bezanson *et al.* 2017). The scripts used to fit the model can be found in (Giménez-Romero 2022).

## Sensitivity Analysis

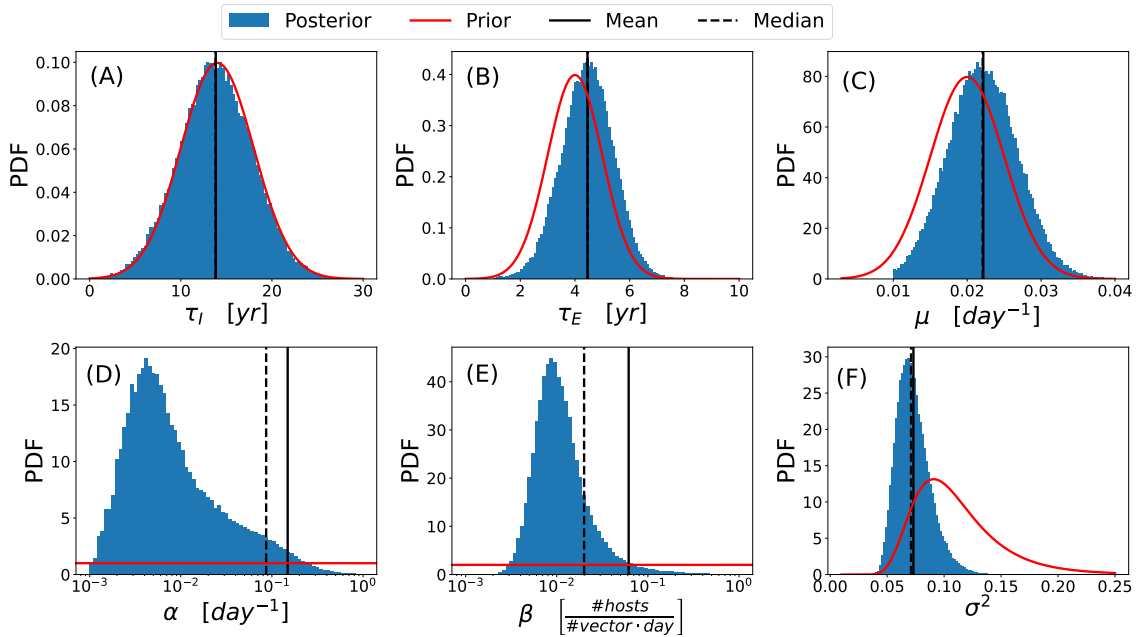
We performed a Global Sensitivity Analysis (GSA) (Saltelli *et al.* 2004) of the model to assess the relative contribution of its parameters and their interactions with different features of the epidemic. In contrast to the Local Sensitivity Analysis (LSA), the GSA assesses the influence of a large domain of the parameter space in the desired outputs of the model. We performed GSA by means of a variance-based analysis, the Sobol method (Sobol 2001). This particular method provides information not only on how a particular parameter alone influences the model outputs (as happens with LSA), but also due to the nonlinear interactions among two or more parameters. Briefly, the method considers the model output,  $Y$ , as a general function of the inputs,  $f(x_1, \dots, x_n)$ , so that the variance of the output,  $Var(Y)$  is decomposed as the sum of the variances given by the variations of the parameters alone and its interactions,  $Var(Y) = \sum_{i=1}^n Var(f(x_i)) + \sum_{i<j}^n Var(f(x_i, x_j)) + \dots$ . This information is organized in what are known as Sobol indices. The total order indices are a measure of the total variance of the output quantity caused by variations of the input parameter and its interactions,  $S_T = Var(f(x_1, \dots, x_n))/Var(Y)$ . First order (or “main effect”) indices are a measure of the contribution to the output variance given by the variation of the parameter alone, but averaged over the variations in other input parameters,  $S_i = Var(f(x_i))/Var(Y)$ . Second-order indices take into account first-order interactions between parameters,  $S_{ij} = Var(f(x_i, x_j))/Var(Y)$ . Further indices can be obtained, describing the influence of higher-order interactions between parameters, but these are not going to be considered.

Following the Sobol method, we analyzed the variation of the time at which the infectious population peaks,  $t_{peak}$ , the magnitude of this peak,  $I_{peak}$  and the final number of dead hosts,  $R_\infty$ , relative to variations of the model parameters. The method was implemented within the Julia high-level programming language (Bezanson *et al.* 2017) using the sub-package DiffEqSensitivity.jl in DifferentialEquations.jl package (Rackauckas *et al.* 2017).

## RESULTS

### Model fit and parameter estimates

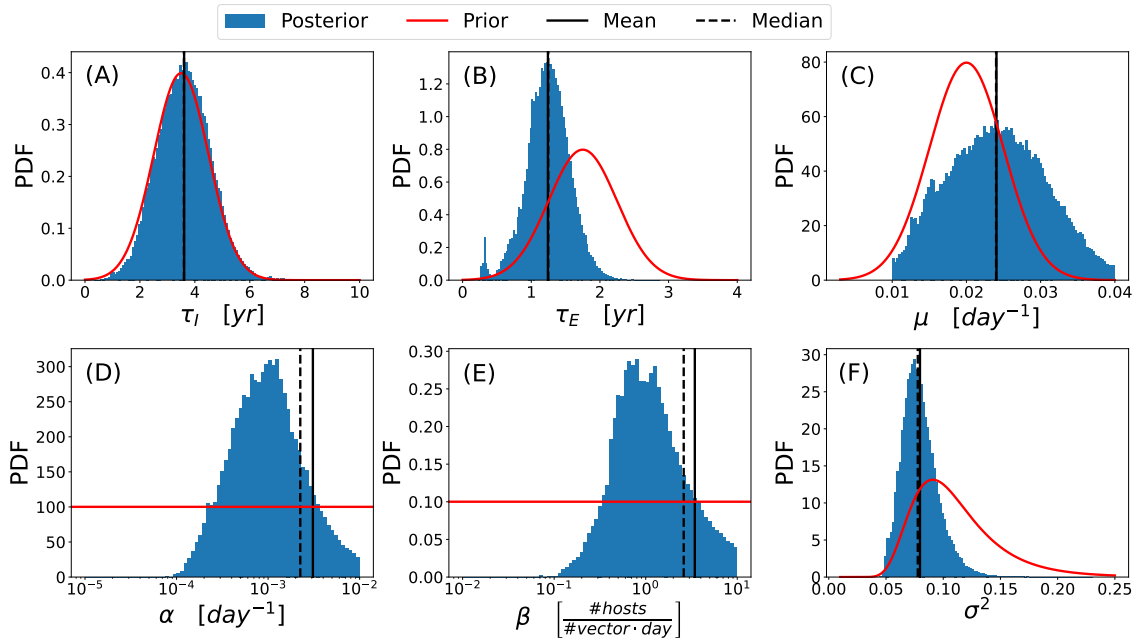
The posterior distributions of the fitted parameters including their estimated mean and median for ALS and OQDS are shown in Figs. 2 and 3, respectively, together with the assumed prior distributions. We observe that the literature-driven priors for the latent and infectious period,  $\tau_E$  and  $\tau_I$ , were already very good guesses and changed slightly converging to the appropriate distribution that better fitted the epidemic data for both ALS and OQDS (Fig. 2(A-B) and Fig. 3(A-B)). Similarly, the prior for the vector removal rate,  $\mu$ , obtained from field data, was good enough so that little changes were needed for convergence (Fig. 2(C) and Fig. 3(C)). On the other hand, we also observe that the completely uninformative priors for the transmission rates successfully converged to the posterior distributions (Fig. 2(D-E) and Fig. 3(D-E)).



**Figure 2:** Posterior (blue histograms) and prior (red line) distributions of the model parameters for ALSD. Solid and dashed black lines correspond to the mean and median of the posterior distributions. (A) Host infectious period  $\tau_I = 1/\gamma$ . (B) Host latent period  $\tau_E = 1/\kappa$ . (C) Vector removal rate  $\mu$ . (D) Vector infection rate  $\alpha$ . (E) Host infection rate  $\beta$ . (F) The variance of the field data  $\sigma^2$ .

The latter distributions are far from a Gaussian-like shape (note that the x-axis is log-scaled), being heavy-tailed. This kind of distribution highly distorts the statistical measures of mean, median and standard error, indicating that the estimates for transmission rates are not as robust as the estimates for the other parameters. These rather uninformative distributions are most probably arising because of the lack of data about the vector, i.e.  $S_v(t)$  and  $I_v(t)$ , to constrain the fits. In essence, many combinations of  $\alpha$  and  $\beta$  can similarly fit the host data while yielding quite different time series for  $S_v(t)$  and  $I_v(t)$ , which cannot be contrasted due to the lack of field data. Nevertheless, the obtained best-fit mean and median parameters, although quite different, are able to perfectly fit the data (Fig. 4). Finally, we also observe that the variance for the field data also converged to a bell-shaped distribution.

Mean and median parameter estimates, i.e. the best-fit parameter values for ALSD and OQDS, are summarized in Tables 1 and 2, respectively. As already seen from the posterior distributions, the best-fit values for  $\tau_E$ ,  $\tau_I$  and  $\mu$  are close to the ones given by literature and field data for both diseases. Conversely,  $\alpha$  and  $\beta$  are rather uninformative, as their 95% confidence intervals cover almost two orders of magnitude. This again indicates that without some data about the evolution of the vector states in time,  $S_v(t)$  and  $I_v(t)$ , it is nearly impossible to derive the proper values for these parameters.



**Figure 3:** Posterior (blue histograms) and prior (red line) distributions of the model parameters for OQDS. Solid and dashed black lines correspond to the mean and median of the posterior distributions. (A) Host infectious period  $\tau_I = 1/\gamma$ . (B) Host latent period  $\tau_E = 1/\kappa$ . (C) Vector removal rate  $\mu$ . (D) Vector infection rate  $\alpha$ . (E) Host infection rate  $\beta$ . (F) Variance of the field data  $\sigma^2$ .

**Table 1:** Estimated epidemiological parameters from Bayesian model fitting to the disease progression curve of ALSD in Mallorca.

Parameter	Definition	Units	Posterior Mean	Posterior Median	95% C.I.
$\tau_I$	Host infectious period	yr	13.84	13.82	[7.12, 20.47]
$\tau_E$	Host latent period	yr	4.46	4.47	[2.88, 5.99]
$\beta$	Host infection rate	$\frac{\text{\#host}}{\text{\#vector}\cdot\text{day}}$	0.062	0.02	[0.0061, 0.3013]
$\alpha$	Vector infection rate	$\text{day}^{-1}$	0.15	0.086	[0.0047, 0.54]
$\mu$	Vector removal rate	$\text{day}^{-1}$	0.0222	0.0221	[0.015, 0.030]
$R_0$	Basic reproductive number	-	133	25	-

**Table 2:** Estimated epidemiological parameters from Bayesian model fitting to the disease progression curve of OQDS in Apulia.

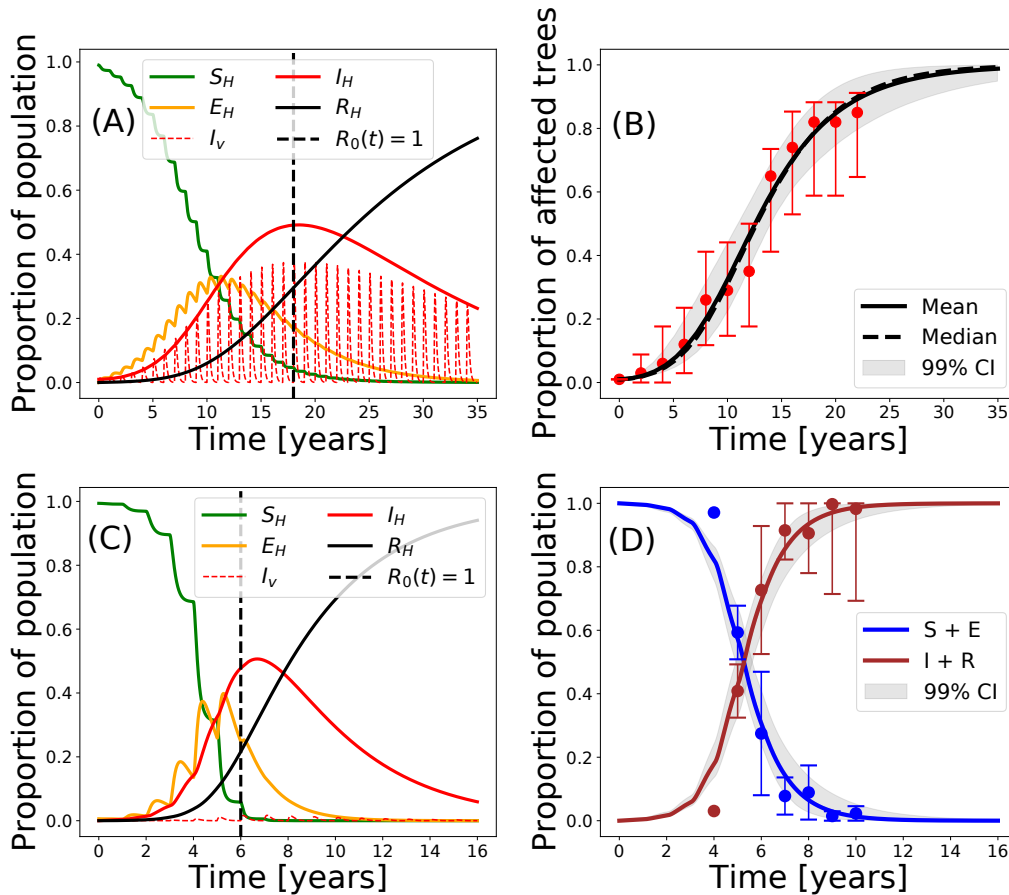
Parameter	Definition	Units	Posterior Mean	Posterior Median	95% C.I.
$\tau_I$	Host infectious period	yr	3.61	3.60	[2.06, 5.20]
$\tau_E$	Host latent period	yr	1.24	1.25	[0.70, 1.75]
$\beta$	Host infection rate	$\frac{\text{\#host}}{\text{\#vector}\cdot\text{day}}$	3.44	2.60	[0.55, 8.79]
$\alpha$	Vector infection rate	$\text{day}^{-1}$	0.0031	0.0022	[0.0005, 0.0084]
$\mu$	Vector removal rate	$\text{day}^{-1}$	0.0240	0.0240	[0.014, 0.035]
$R_0$	Basic reproductive number	-	33	21	-

Overall, the data falls within the 99% confidence limits of the fitted model for both the ALSD and OQDS outbreaks (Fig. 4(B,D)). We also computed the instantaneous reproductive number,  $R_0(t)$ , by using Eq. (3) with  $S_H(t)$  instead of only  $S_H(0)$  along the simulation. Noteworthy,  $R_0(t) = 1$  coincides with the stopping of new infections being produced, i.e. the number of exposed hosts does



not increase (Fig. 4(A,C)). This supports our approximate method for computing the reproductive number for Xf diseases (Appendix II, Eq. (3)). Due to the different time scales of both epidemics ( $\tau_I^{ALSD} + \tau_E^{ALSD} > \tau_I^{OQDS} + \tau_E^{OQDS}$ ), the OQDS outbreak dies out earlier than the one for ALSD.

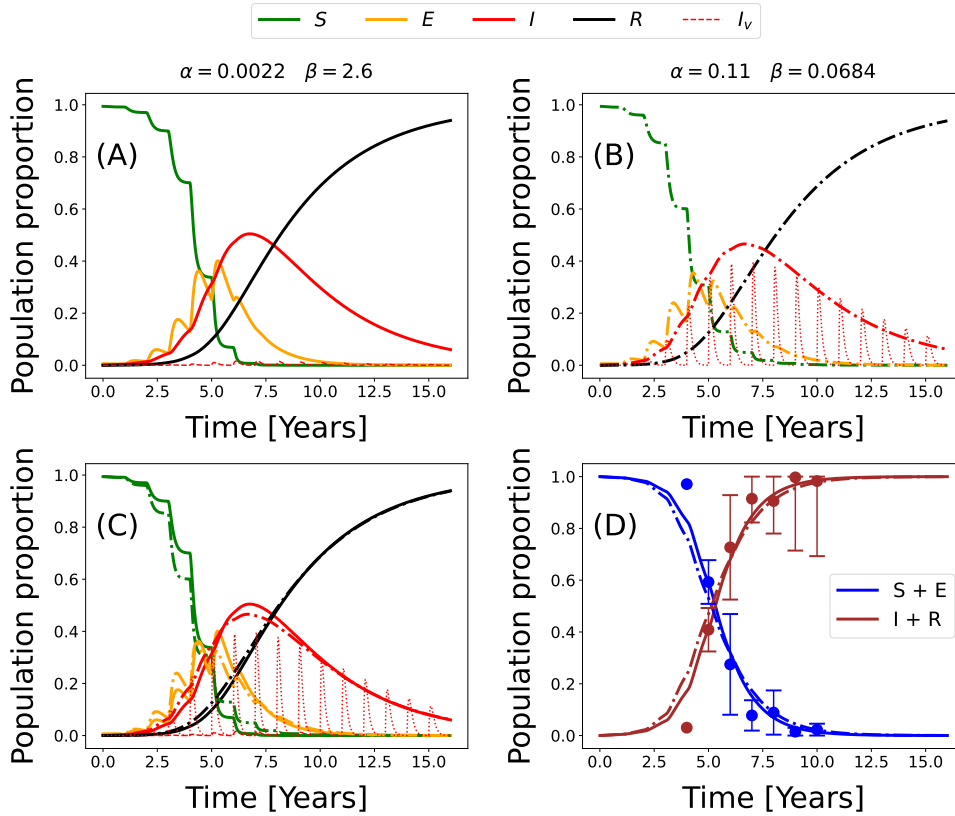
We notice that for ALSD a large proportion of the vector population gets infected every year (Fig. 4(A)), while a very small proportion is needed in OQDS to produce a lethal outbreak (Fig. 4(C)). However, this last statement is rather unrealistic, as around 50% of the vectors that are captured in Apulia are indeed infected by Xf (Cavalieri *et al.* 2019; Cornara *et al.* 2017b). Thus, the evolution of the infected vector population should be qualitatively similar to that obtained for ALSD (Fig. 4(C)). As previously explained, different suitable combinations of  $\alpha$  and  $\beta$  parameters should give rise to similar progression curves for the hosts while different ones for the vectors, but the realistic values for these parameters cannot be obtained from the Bayesian fit due to the lack of data of the vector states,  $S_v(t)$ ,  $I_v(t)$ .



**Figure 4:** (A) Simulation of the model with the best-fit parameters for ALSD. (B) Model fit to field data by means of the mean and median values of the posterior distributions of the parameters for ALSD. (C) Simulation of the model with the best-fit parameters for OQDS. (D) Model fit to field data by means of the mean and median values of the posterior distributions of the parameters for OQDS. The gray-shaded area corresponds to the 99% confidence interval. The error bars for the field data correspond to their 95% confidence interval obtained with a bootstrapping technique.

Nevertheless, by manually exploring other values for  $\alpha$  and  $\beta$  parameters, we can obtain a more biologically plausible scenario for the OQDS that is still able to fit the available data for the hosts. Fig. 5(A) shows a simulation of the model with previously inferred best-fit median parameters for OQDS. By changing the values of  $\alpha$  and  $\beta$ , we obtain a more realistic scenario, i.e. around a 50% of the vector population getting infected during the outbreak (Fig. 5(B)) (Cavalieri *et al.* 2019; Cornara *et al.* 2017b). Noteworthy, the  $\beta$  value obtained in this way is almost identical to the transmission rate recently reported by Bodino *et al.* 2021 for OQDS. This change in the

transmission parameters only affects the progression curve of the infected vector population, being the progression of the host compartments practically unchanged (Fig. 5(C)). Anyway, both sets of parameter values for  $\alpha$  and  $\beta$  can properly fit the field data, corresponding exclusively to the host population (Fig. 5(D)).

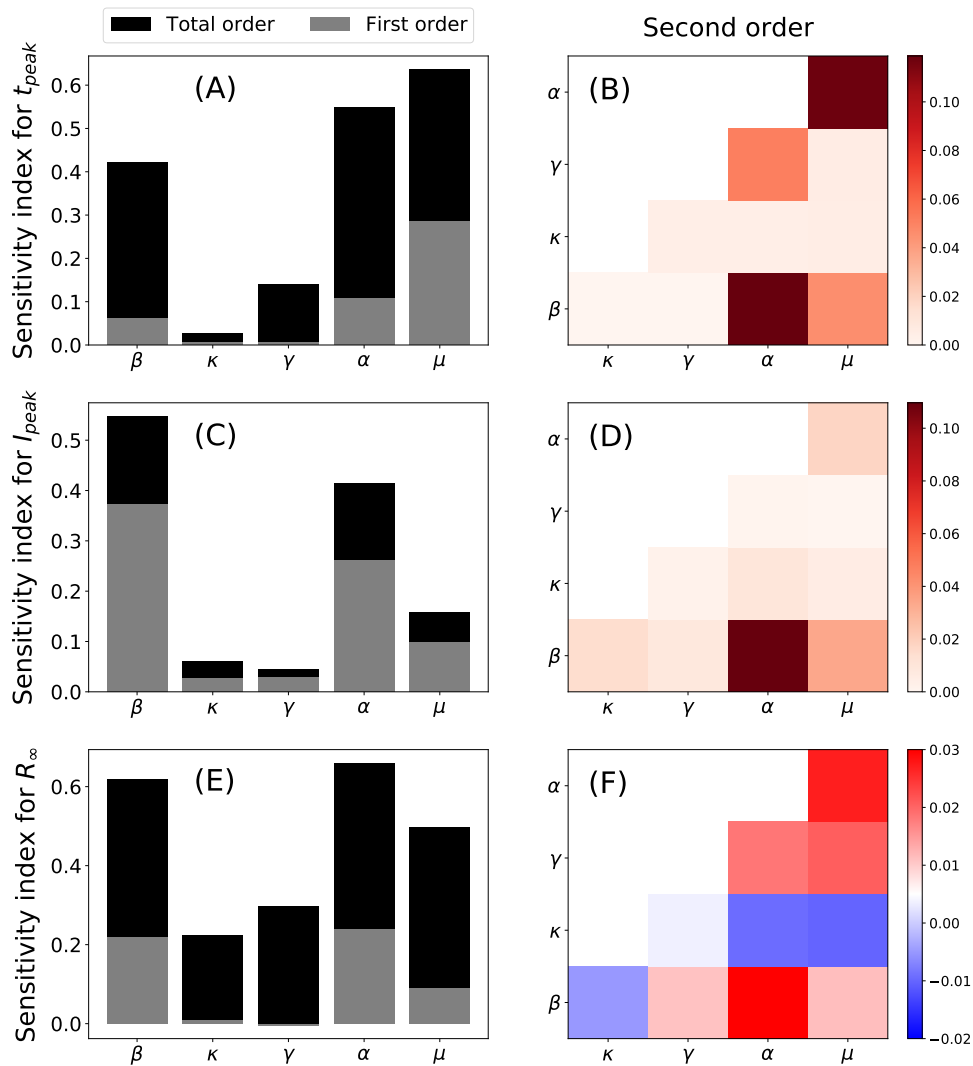


**Figure 5:** (A) Simulation of the model with the original best-fit parameters for OQDS. (B) Simulation of the model with the original best-fit parameters for OQDS but with different  $\alpha$ ,  $\beta$  values. (C) Comparison of the progression curves. Note that the curves for the hosts are very similar while the curve for the infected vector population is very different. (D) Comparison of the model fit to the data with both simulations. Solid lines correspond to results with the original best-fit parameters while dash-dot lines correspond to the results of the more realistic scenario with different  $\alpha$  and  $\beta$ .

The model adjusted to the progression curves of both diseases indicates that the transmission rate  $\alpha$  must be greater than  $\beta$  when the proportion of infected vectors is relatively high ( $> 30\%$ ). We checked if the relation between  $\alpha$  and  $\beta$  held when changing the assumed  $N_v(0) = N_H/2$ , obtaining that it kept approximately the same for very different values of the initial vector population.

## Global Sensitivity Analysis

We computed the sensitivity indices for the model parameters with respect to the more relevant quantities of interest, namely, the time at which the number of infectious hosts is maximal,  $t_{\text{peak}}$ , the maximum number of infectious hosts,  $I_{\text{peak}}$  and the final number of dead hosts,  $R_{\infty}$ . The results were obtained exploring the parameter space constrained to the intervals  $\{\beta \in (0.001, 0.1), \tau_E \in (3, 7), \tau_I \in (5, 25), \alpha \in (0.001, 1), \mu \in (0.01, 0.04)\}$  using  $10^4$  Quasi-Monte Carlo samples and are summarized in Fig. 6.



**Figure 6:** Global Sensitivity Analysis of the model parameters performed with the Sobol method with respect to the time at which the infectious population peaks,  $t_{peak}$  (A-B), the magnitude of this peak,  $I_{peak}$  (C-D) and the final number of dead hosts,  $R_{\infty}$  (E-F). The left column (A,C,E) shows the total and first-order indices and the right column (B,D,F) shows the second-order indices.

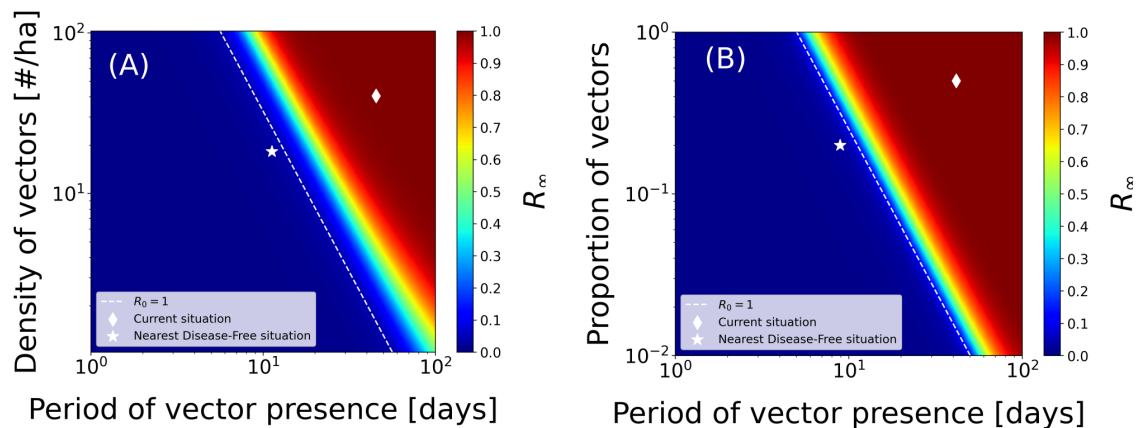
Parameters  $\alpha$ ,  $\beta$  and  $\mu$  are the most influential with regard to the time at which the infectious host population peaks,  $t_{peak}$ , the magnitude of the peak,  $I_{peak}$ , and the final number of dead hosts,  $R_{\infty}$ . The total output variance (total order indices) cannot be explained by the variances of the parameters alone (first order indices) (Fig. 6). Therefore, higher-order interactions among the parameters importantly affect the sensitivity of the quantities under study. Indeed, the contribution to the total output variance of  $\gamma$  and  $\kappa$  for  $t_{peak}$  and  $R_{\infty}$  come notably from higher-order interactions. This can be checked in panels (B,C,F) of Fig. 6, in which the contribution to the output variance from interactions between pairs of parameters (second order indices) is represented. Interactions among the parameters contribute to increasing the output variance with respect to  $t_{peak}$  and  $I_{peak}$ , while the effect is more heterogeneous in the case of  $R_{\infty}$ . In particular, the interactions between

$\alpha - \beta$  and  $\alpha - \mu$  produce the main contributions to the increase of output variance in all cases, while  $\kappa - \beta$ ,  $\kappa - \alpha$  and  $\kappa - \mu$  interactions decrease the output variance.

## Epidemic control through vector management

The sensitivity analysis clearly indicates that acting on  $\alpha$ ,  $\beta$  and  $\mu$  is the best strategy to lower disease incidence and mortality. However, controlling transmission rates is cumbersome so a different control strategy based only on vector control is considered in this section. In our model, there are two ways of implementing vector-population control: (i) decreasing the typical time,  $1/\mu$ , that vectors spend between crops each year by some mechanism (thus increasing  $\mu$ ) and (ii) reducing the initial number of vectors that invade crops each year (e.g. lowering  $N_v(0)$  via egg or nymph control (Lago *et al.* 2022)).

We analyzed the effect of vector management by simulating epidemic outbreaks using different values of  $\mu$  and  $N_v(0)$ , and keeping the rest of parameters as fitted for both ALSD and OQDS outbreaks (Fig. 7). In both epidemics, decreasing the presence time as well as the number of vectors contribute to controlling the epidemic by lowering  $R_0$  and, consequently, the final size of the epidemic,  $R_\infty$ . Furthermore, we observe that decreasing vector presence is more efficient than decreasing its annual initial population, i.e. we further reduce  $R_\infty$ , the final size of the epidemic, by applying a similar reduction in the residence time  $1/\mu$ . This could also be anticipated as  $R_0$  depends quadratically on  $1/\mu$  while only linearly on  $N_v(0)$  (Eq. (3)). However, the minimal intervention strategy, starting from the current situation in the  $(1/\tau, N_v(0))$  parameter space that yields an absolute control of the epidemic,  $R_0 < 1$ , involves a mixed strategy of lowering both  $1/\mu$  and  $N_v(0)$ .



**Figure 7:** Epidemic control through vector management for ALSD in Mallorca (A) and OQDS in Apulia (B). The white shaded line denotes  $R_0 = 1$ , the white diamond corresponds to the parameter values of the fitted model. The white star is the closest disease-free state to the current situation in this representation.

## DISCUSSION

In this work, we have developed a deterministic continuous-time compartmental model for *Xylella fastidiosa* vector-borne diseases in Europe. The model attempts to characterize the main biotic processes that lead to the development of epidemics, including the seasonal dynamics of the main vector, *P. spumarius*. We show how the model is sufficiently general to represent with some accuracy the parameters that determine the ALSD in Mallorca (Spain) and the OQDS in Apulia (Italy), both transmitted by *P. spumarius*. To our best knowledge, this is the first mathematical model describing Xf epidemics that considers the temporal pattern of vector abundance observed in field data, faithfully representing the known biological information about the pathosystem. It includes a dynamic approximation of the non-stationary populations of *P. spumarius*, mathematically represented by a sporadic source term through which vectors are born every year, and an exponential decay term. Due to the non-stationarity of the vector dynamics,  $R_0$  in the model cannot be computed

with standard methods such as the Next Generation Matrix (Diekmann *et al.* 2010). To circumvent this problem, we applied an approximate method to compute it as previously proposed by (Giménez-Romero *et al.* 2022a). We show that this approximate  $R_0$  correctly characterizes the epidemic, further validating the method proposed by (Giménez-Romero *et al.* 2022a).

Nonlinear mathematical models of disease transmission enhance our understanding of the different mechanisms operating in an epidemic, especially compared with correlative or machine learning methods, often very useful in practice but offering very little understanding. A key aspect to render these models useful is the determination of the parameters from available data. If this step can be properly performed, these models become very predictive and especially helpful to design disease control strategies. However, an appropriate calibration of the model relies on access to good-quality field data, which is often the bottleneck for the application of this kind of models. In the present study, the parameters have been obtained using a Bayesian inference framework, which relies on probability distributions rather than point-like measures. This way, mean or median values can be considered together with their confidence intervals able to characterize the robustness of the obtained parameters. In general, we obtained different values of the parameters for the ALSD and OQDS outbreaks in Mallorca and Apulia, respectively. The fitted values, however, are in good agreement with previous field-based measures for each disease while the differences observed between both outbreaks may reflect differences between the Xf subspecies and crops involved (deciduous vs. evergreen).

One of the conclusions of the study is that the available data for both diseases is not enough to obtain robust estimates for all of the model parameters. The lack of data about the vector population compartments yields many possible values for the parameters that regulate transmission,  $\alpha$  and  $\beta$ , provided that the progression of the host compartments correctly fits the field data. In other words, very infectious vectors (high  $\beta$ ) that hardly ever get infected (low  $\alpha$ ) can produce a similar outbreak within the host population to that produced by very low infectious vectors (low  $\beta$ ) that get infected very often (high  $\alpha$ ). The great difference in these situations would be that, in the former, the infected vector population would be very low, while in the latter, it would be quite high. This is a manifestation of parameter unidentifiability from the fit (Chowell 2017; Roosa *et al.* 2019), which stresses the importance of transmission and calls for detailed measurements of the vector population, and not just of the hosts. Furthermore, to compare transmission rates between different diseases caused by Xf (e.g.  $\beta$ ,  $\alpha$ ), it is necessary to know the vector-host population ratio of the pathosystem ( $N_v/N_H$ ), since  $\beta$  is expressed as a number of hosts per vectors per day. Although, in general, populations of *P. spumarius* in the canopy of olive trees are much larger than those found in the almond trees of the Balearic Islands during the months of July and August (López-Mercadal *et al.* 2021), our work is based on data from studies in which information of the vector populations is not provided. Without this information, therefore, conclusive results regarding transmission cannot be obtained.

In any case, our model shows that the vector-to-plant transmission process, mediated by  $\beta$ , is somehow different from that from the plant-to-vector one, mediated by  $\alpha$ . In essence,  $\beta$  must be smaller than  $\alpha$  in order to reproduce the observed outbreaks and have a sufficiently large vector population getting infectious, being this fact independent of the particular choice of  $N_v(0)/N_H$ . This heterogeneity can be caused by several factors: differences in the efficiency of plant-to-vector transmission with respect to vector-to-plant transmission, differences in contact rates, i.e. susceptible vectors contact trees at a different rate than infected vectors; vector feeding preferences, i.e. differences in the probability of contacting a susceptible host compared to an infectious host, etc. Indeed, our mathematical model assumes constant contact rates with no preferences over any host state, so that under these assumptions, it indicates that the probability of effectively transmitting the pathogen from plant to vector is greater than from vector to plant. However, this interpretation is subject to this particular assumption, so that to fully disentangle this question experimental work in form of transmission assays should be performed. Furthermore, we found that the timing and magnitude of the infectious host peak and the final number of dead hosts are mostly controlled by the vector-to-plant transmission rate,  $\beta$ , the plant-to-vector transmission rate,  $\alpha$  and the vector removal rate  $\mu$ . Because these parameters are strongly related to the vector, the analysis makes clear that enhancing the knowledge about the vector, as well as obtaining precise data, is crucial to improve the modeling of Xf diseases and pose important questions to be solved in specifically designed experiments.

The fact that the most influential parameters of the model are those related to the vector can be used to design appropriate disease control strategies. Because acting on transmission rates is rather cumbersome, we argue that control strategies should focus on reducing the vector population in crop fields. In our model, this depends on two parameters,  $\mu$ , the rate at which vectors die (or move to herbaceous vegetation and other non-host trees or exit the field) and  $N_v(0)$ , the number of newborns susceptible vectors every year (assumed constant in this study). Our results show that a mixed strategy acting on both parameters is optimal to lower disease prevalence and, eventually, eradicate the disease. Interestingly, we also show that acting on the vector removal  $\mu$  is more effective than controlling the newborn vector population  $N_v(0)$ . In fact, most control strategies carried out in practice for Xf diseases focus on the latter factor, reducing  $N_v(0)$  via egg or nymph control (Cornara *et al.* 2018b; Lago *et al.* 2022; López-Mercadal *et al.* 2022). However, our results indicate that alternative strategies based on increasing the removal (or dispersal) rate of vectors should be explored. Furthermore, the evolution of the population compartments of the hosts and vectors provides relevant information on the epidemiology of both diseases. In both cases, the newly defined basic reproductive number that accounts for a decaying vector population is very predictive of the moment in which new infections are not produced anymore, coinciding approximately with the peak of infectious hosts. Therefore, any intervention with control measures after this peak would have marginal effects on future disease progression.

Our mathematical model is still rather simple, implementing only a few relevant epidemic processes in contrast to the high complexity of the pathogen-vector-host interactions occurring in plant epidemics. Indeed, the model itself raises some questions about these interactions, for example, whether or not contact rates are homogeneous. Another simplification of the model is the fact that the spatial constraints and the intrinsic stochasticity of the transmission processes are neglected. A straightforward extension of the model would be to include a specific spatial setting and implement the explicit motion of the vector within a stochastic framework, such as Individual Based Models (Grimm *et al.* 2005). With this, the effectiveness of current and further control strategies could be tested and improved controlling for the motion of the vector. For instance, the control strategy based on the removal of symptomatic trees together with their surrounding trees at a given distance could be implemented in the model, evaluate the current effectiveness according to the present protocols and even provide improved parameters to be implemented in the field. Of course, implementing a model in which the spatial degrees of freedom are explicitly represented would require access to further information about vector mobility and spatially resolved data to confront the model, which is not currently available.

Mathematical models tested against experimental data increase our understanding of the system under study. They also help to identify critical parameters that require better prior information to adjust functions relating to different variables and make the model predictions more accurate to suggest and test control strategies (Cunniffe *et al.* 2015; Jeger *et al.* 2018). Our mathematical model suggests a certain lack of knowledge of the transmission processes and reveals that the currently available data is not enough to fit complex models dealing with the explicit dynamics of the vector population.

**Funding.** AGR and MAM acknowledge financial support from Grant No. PID2021-123723OB-C22 (CYCLE) funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe” and from Grant CEX2021-001164-M (María de Maeztu Program for Units of Excellence in R&D) funded by MCIN/AEI/10.13039/501100011033.

The authors declare no conflict of interest.

## References

- Abboud, C., O. Bonnefon, E. Parent, and S. Soubeyrand (2019). “Dating and localizing an invasion from post-introduction data and a coupled reaction–diffusion–absorption model”. *Journal of Mathematical Biology* **79**, pp. 765–789. DOI: [10.1007/s00285-019-01376-x](https://doi.org/10.1007/s00285-019-01376-x).
- Almeida, R. P. P. and L. Nunney (1987). “How do plant diseases caused by *Xylella fastidiosa* emerge?” *Plant Disease* **99** (11), pp. 1457–1467. DOI: [10.1094/PDIS-02-15-0159-FE](https://doi.org/10.1094/PDIS-02-15-0159-FE).
- Almeida, R. P. P. and A. H. Purcell (2003). “Biological traits of *Xylella fastidiosa* strains from grapes and almonds”. *Applied and Environmental Microbiology* **69**, pp. 7447–7452. DOI: [10.1128/AEM.69.12.7447-7452.2003](https://doi.org/10.1128/AEM.69.12.7447-7452.2003).
- Antonatos, S. et al. (2021). “Seasonal Appearance, Abundance, and Host Preference of *Philaenus spumarius* and *Neophilaenus campestris* (Hemiptera: Aphrophoridae) in Olive Groves in Greece”. *Environmental Entomology* **50**, pp. 1474–1482. DOI: [10.1093/ee/nvab093](https://doi.org/10.1093/ee/nvab093).
- Bacaër, N. (2007). “Approximation of the Basic Reproduction Number  $R_0$  for Vector-Borne Diseases with a Periodic Vector Population”. *Bulletin of Mathematical Biology* **69**, pp. 1067–1091. DOI: [10.1007/s11538-006-9166-9](https://doi.org/10.1007/s11538-006-9166-9).
- Beal, D. J., M. Cooper, M. P. Daugherty, A. H. Purcell, and R. P. P. Almeida (2021). “Seasonal Abundance and Infectivity of *Philaenus spumarius* (Hemiptera: Aphrophoridae), a Vector of *Xylella fastidiosa* in California Vineyards”. *Environmental Entomology* **50**, pp. 467–476. DOI: [10.1093/ee/nvaa178](https://doi.org/10.1093/ee/nvaa178).
- Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah (2017). “Julia: A fresh approach to numerical computing”. *SIAM Review* **59**, pp. 65–98. DOI: [10.1137/141000671](https://doi.org/10.1137/141000671).
- Bodino, N. et al. (2019). “Phenology, seasonal abundance and stage-structure of spittlebug (Hemiptera: Aphrophoridae) populations in olive groves in Italy”. *Scientific reports* **9**, pp. 1–17. DOI: [10.1038/s41598-019-54279-8](https://doi.org/10.1038/s41598-019-54279-8).
- Bodino, N. et al. (2021). “Temporal dynamics of the transmission of *Xylella fastidiosa* subsp. *pauca* by *Philaenus spumarius* to olive plants”. *Entomologia Generalis* **41**, pp. 463–480. DOI: [10.1127/entomologia/2021/1294](https://doi.org/10.1127/entomologia/2021/1294).
- Brauer, F., C. Castillo-Chavez, A. Mubayi, and S. Towers (2016). “Some models for epidemics of vector-transmitted diseases”. *Infectious Disease Modelling* **1**, pp. 79–87. DOI: [10.1016/j.idm.2016.08.001](https://doi.org/10.1016/j.idm.2016.08.001).
- Brunetti, M., V. Capasso, M. Montagna, and E. Venturino (2020). “A mathematical model for *Xylella fastidiosa* epidemics in the Mediterranean regions. Promoting good agronomic practices for their effective control.” *Ecological Modelling* **432**, p. 109204. DOI: [10.1016/j.ecolmodel.2020.109204](https://doi.org/10.1016/j.ecolmodel.2020.109204).
- Cavaliere, V. et al. (2019). “Transmission of *Xylella fastidiosa* Subspecies *Pauca* Sequence Type 53 by Different Insect Species”. *Insects* **10**, p. 324. DOI: [10.3390/insects10100324](https://doi.org/10.3390/insects10100324).
- Chew, Y. H. et al. (2014). “Mathematical Models Light Up Plant Signaling”. *The Plant Cell* **26**, pp. 5–20. DOI: [10.1105/tpc.113.120006](https://doi.org/10.1105/tpc.113.120006).
- Chiyaka, C., B. H. Singer, S. E. Halbert, J. G. Morris Jr, and A. H. C. van Bruggen (2012). “Modeling huanglongbing transmission within a citrus tree”. *Proceedings of the National Academy of Sciences* **109**, pp. 12213–12218. DOI: [10.1073/pnas.1208326109](https://doi.org/10.1073/pnas.1208326109).
- Chmiel, S. M. and M. C. Wilson (1979). “Estimation of the Lower and Upper Developmental Threshold Temperatures and Duration of the Nymphal Stages of the Meadow Spittlebug, *Philaenus spumarius*”. *Environmental Entomology* **8**, pp. 682–685. DOI: [10.1093/ee/8.4.682](https://doi.org/10.1093/ee/8.4.682).
- Chowell, G. (2017). “Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts”. *Infectious Disease Modelling* **2**, pp. 379–398. DOI: [10.1016/j.idm.2017.08.001](https://doi.org/10.1016/j.idm.2017.08.001).
- Cornara, D., D. Bosco, and A. Ferreres (2018a). “*Philaenus spumarius*: when an old acquaintance becomes a new threat to European agriculture”. *Journal of Pest Science* **91**, pp. 957–972. DOI: [10.1007/s10340-018-0966-0](https://doi.org/10.1007/s10340-018-0966-0).
- Cornara, D. et al. (2017a). “Spittlebugs as vectors of *Xylella fastidiosa* in olive orchards in Italy”. *Journal of Pest Science* **90**, pp. 521–530. DOI: [10.1007/s10340-016-0793-0](https://doi.org/10.1007/s10340-016-0793-0).
- Cornara, D. et al. (2017b). “Transmission of *Xylella fastidiosa* by naturally infected *Philaenus spumarius* (Hemiptera, Aphrophoridae) to different host plants”. *Journal of Applied Entomology* **141**, pp. 80–87. DOI: [10.1111/jen.12365](https://doi.org/10.1111/jen.12365).

- Cornara, D. et al. (2018b). “EPG combined with micro-CT and video recording reveals new insights on the feeding behavior of *Philaenus spumarius*”. *PLoS One* **13**, pp. 1–20. DOI: [10.1371/journal.pone.0199154](https://doi.org/10.1371/journal.pone.0199154).
- Cornara, D. et al. (2021). “Natural areas as reservoir of candidate vectors of *Xylella fastidiosa*”. *Bulletin of Insectology* **74**, pp. 173–180.
- Cunniffe, N. J. et al. (2015). “Thirteen challenges in modelling plant diseases”. *Epidemics* **10**, pp. 6–10. DOI: [10.1016/j.epidem.2014.06.002](https://doi.org/10.1016/j.epidem.2014.06.002).
- Daugherty, M. P. and R. P. P. Almeida (2019). “Understanding How an Invasive Vector Drives Pierce’s Disease Epidemics: Seasonality and Vine-to-Vine Spread”. *Phytopathology* **109**, pp. 277–285. DOI: [10.1094/PHYTO-07-18-0217-FI](https://doi.org/10.1094/PHYTO-07-18-0217-FI).
- Diekmann, O., J. A. P. Heesterbeek, and M. G. Roberts (2010). “The construction of next-generation matrices for compartmental epidemic models”. *Journal of The Royal Society Interface* **7**, pp. 873–885. DOI: [10.1098/rsif.2009.0386](https://doi.org/10.1098/rsif.2009.0386).
- Fierro, A., A. Liccardo, and F. Porcelli (2019). “A lattice model to manage the vector and the infection of the *Xylella fastidiosa* on olive trees”. *Scientific Reports* **9**, p. 8723. DOI: [10.1038/s41598-019-44997-4](https://doi.org/10.1038/s41598-019-44997-4).
- Freitag, J. H. (1951). “Host range of the Pierce’s disease virus of grapes as determined by insect transmission”. *Phytopathology* **41**, pp. 920–934.
- Ge, H., K. Xu, and Z. Ghahramani (2018). “Turing: a language for flexible probabilistic inference”. In: *International Conference on Artificial Intelligence and Statistics, AISTATS 2018, 9-11 April 2018, Playa Blanca, Lanzarote, Canary Islands, Spain*, pp. 1682–1690.
- Giménez-Romero, À. (2022). “A compartmental model for *Xylella fastidiosa* related diseases”. *GitHub Repo*.
- Giménez-Romero, À., R. Flaquer-Galmés, and M. A. Matías (2022a). “Vector-borne diseases with nonstationary vector populations: The case of growing and decaying populations”. *Physical Review E* **106**, p. 054402. DOI: [10.1103/PhysRevE.106.054402](https://doi.org/10.1103/PhysRevE.106.054402).
- Giménez-Romero, À. et al. (2022b). “Global predictions for the risk of establishment of Pierce’s disease of grapevines”. *Communications Biology* **5**, p. 1389. DOI: [10.1038/s42003-022-04358-w](https://doi.org/10.1038/s42003-022-04358-w).
- Grimm, V. and S. F. Railsback (2005). *Individual-based Modeling and Ecology*. Princeton (NJ, USA): Princeton University Press.
- Homan, M. D. and A. Gelman (2014). “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo”. *J. Mach. Learn. Res.* **15**, pp. 1593–1623.
- Hopkins, D. L. and A. H. Purcell (2002). “*Xylella fastidiosa*: Cause of Pierce’s Disease of Grapevine and Other Emergent Diseases”. *Plant Disease* **86**, pp. 1056–1066. DOI: [10.1094/PDIS.2002.86.10.1056](https://doi.org/10.1094/PDIS.2002.86.10.1056).
- Jeger, M. J. and C. Bragard (2019). “The Epidemiology of *Xylella fastidiosa*; A Perspective on Current Knowledge and Framework to Investigate Plant Host–Vector–Pathogen Interactions”. *Phytopathology* **109**, pp. 200–209. DOI: [10.1094/PHYTO-07-18-0239-FI](https://doi.org/10.1094/PHYTO-07-18-0239-FI).
- Jeger, M. J., J. Holt, F. Van Den Bosch, and L. V. Madden (2004). “Epidemiology of insect-transmitted plant viruses: modelling disease dynamics and control interventions”. *Physiological Entomology* **29**, pp. 291–304. DOI: [10.1111/j.0307-6962.2004.00394.x](https://doi.org/10.1111/j.0307-6962.2004.00394.x).
- Jeger, M. J., L. V. Madden, and F. Van Den Bosch (2018). “Plant virus epidemiology: Applications and prospects for mathematical modeling and analysis to improve understanding and disease control”. *Plant Disease* **102**, pp. 837–854. DOI: [10.1094/PDIS-04-17-0612-FE](https://doi.org/10.1094/PDIS-04-17-0612-FE).
- Jeger, M. J., F. van den Bosch, L. V. Madden, and J. Holt (1998). “A model for analysing plant-virus transmission characteristics and epidemic development”. *Mathematical Medicine and Biology: A Journal of the IMA* **15**, pp. 1–18. DOI: [10.1093/imammb/15.1.1](https://doi.org/10.1093/imammb/15.1.1).
- Lago, C. et al. (2022). “Degree-day-based model to predict egg hatching of *Philaenus spumarius* (Hemiptera: Aphrophoridae), the main vector of *Xylella fastidiosa* in Europe”. *bioRxiv*. DOI: [10.1101/2022.11.10.515963](https://doi.org/10.1101/2022.11.10.515963).
- Levin, S. A., ed. (1992). *Mathematics and Biology: The interface, challenges and opportunities*. Lawrence Berkeley Lab., CA (USA).
- López-Mercadal, J., P. Mercadal-Frontera, and M. Á. Miranda (2022). “Mechanical management of weeds drops nymphal density of *Xylella fastidiosa* vectors”. *bioRxiv*. DOI: [10.1101/2022.10.18.512680](https://doi.org/10.1101/2022.10.18.512680).



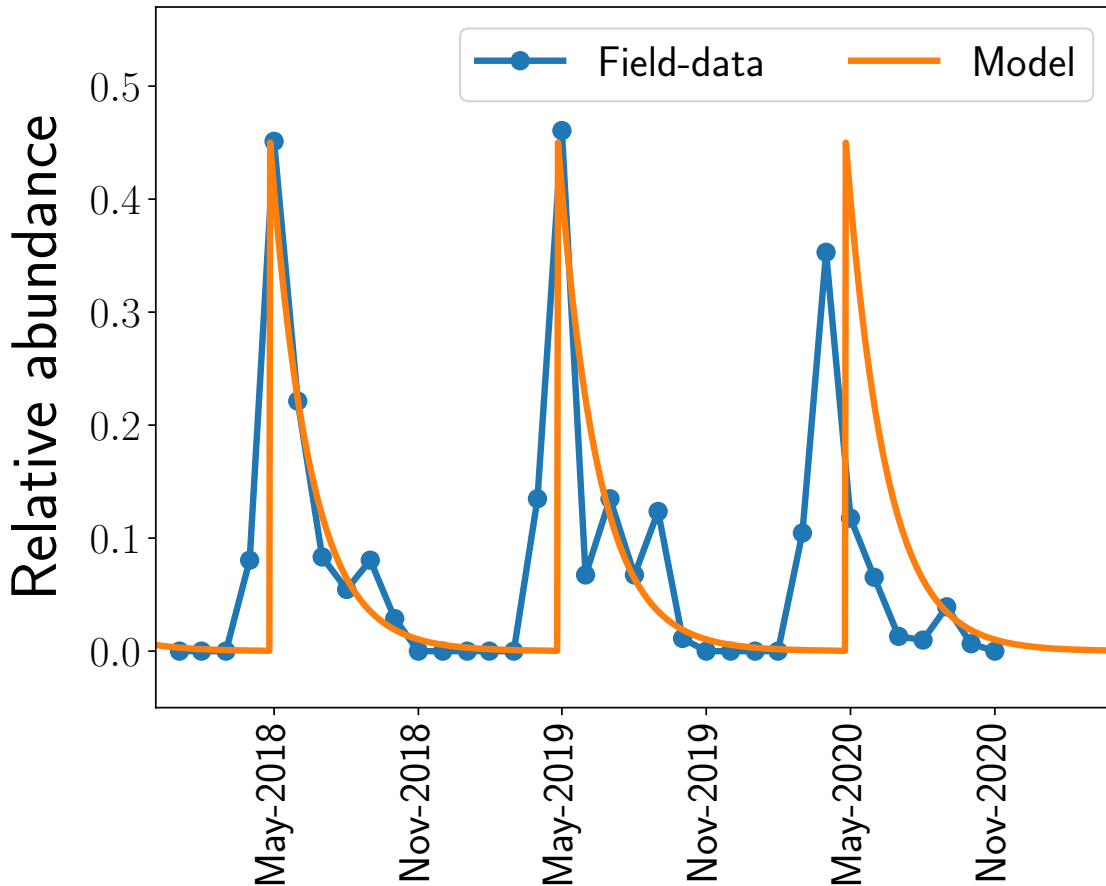
- López-Mercadal, J. et al. (2021). “Collection of data and information in Balearic Islands on biology of vectors and potential vectors of *Xylella fastidiosa* (GP/EFSA/ALPHA/017/01)”. *EFSA Supporting Publications* **18**, 6925E. DOI: [10.2903/sp.efsa.2021.EN-6925](https://doi.org/10.2903/sp.efsa.2021.EN-6925).
- Madden, L. V., M. J. Jeger, and F. van den Bosch (2000). “A Theoretical Assessment of the Effects of Vector-Virus Transmission Mechanism on Plant Virus Disease Epidemics”. *Phytopathology* **90**, pp. 576–594. DOI: [10.1094/PHYTO.2000.90.6.576](https://doi.org/10.1094/PHYTO.2000.90.6.576).
- Martcheva, M. (2015). *An Introduction to Mathematical Epidemiology*. New York (NY, USA): Springer. DOI: [10.1007/978-1-4899-7612-3](https://doi.org/10.1007/978-1-4899-7612-3).
- Moralejo, E. et al. (2019). “Insights into the epidemiology of Pierce’s disease in vineyards of Mallorca, Spain”. *Plant Pathology* **68**, pp. 1458–1471. DOI: <https://doi.org/10.1111/ppa.13076>.
- Moralejo, E. et al. (2020). “Phylogenetic inference enables reconstruction of a long-overlooked outbreak of almond leaf scorch disease (*Xylella fastidiosa*) in Europe”. *Communications Biology* **3**, p. 560. DOI: [10.1038/s42003-020-01284-7](https://doi.org/10.1038/s42003-020-01284-7).
- Morente, M. et al. (2018). “Distribution and relative abundance of insect vectors of *Xylella fastidiosa* in olive groves of the Iberian Peninsula”. *Insects* **9**, p. 175. DOI: [10.3390/insects9040175](https://doi.org/10.3390/insects9040175).
- Murray, J. D. (1989). *Mathematical Biology*. Berlin (Germany): Springer.
- Olmo, D. et al. (2021). “Landscape Epidemiology of *Xylella fastidiosa* in the Balearic Islands”. *Agronomy* **11**, p. 473. DOI: [10.3390/agronomy11030473](https://doi.org/10.3390/agronomy11030473).
- Purcell, A. H. and A. Finlay (1979). “Evidence for noncirculative transmission of Pierce’s disease bacterium by sharpshooter leafhoppers”. *Phytopathology* **69**, pp. 393–395. DOI: [10.1094/Phyto-69-393](https://doi.org/10.1094/Phyto-69-393).
- Rackauckas, C. and Q. Nie (2017). “DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia”. *The Journal of Open Research Software* **5**, p. 15. DOI: [10.5334/jors.151](https://doi.org/10.5334/jors.151).
- Redak, R. A. et al. (2004). “The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology”. *Annual Review of Entomology* **49**, pp. 243–270. DOI: [10.1146/annurev.ento.49.061802.123403](https://doi.org/10.1146/annurev.ento.49.061802.123403).
- Roosa, K. and G. Chowell (2019). “Assessing parameter identifiability in compartmental dynamic models using a computational approach: application to infectious disease transmission models”. *Theoretical Biology and Medical Modelling* **16**, p. 1. DOI: [10.1186/s12976-018-0097-6](https://doi.org/10.1186/s12976-018-0097-6).
- Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Hoboken (NJ, USA): Wiley.
- Saponari, M., D. Boscia, F. Nigro, and G. P. Martelli (2013). “Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy)”. *Journal of Plant Pathology* **95**. DOI: [10.4454/JPP.V95I3.035](https://doi.org/10.4454/JPP.V95I3.035).
- Saponari, M., A. Giampetruzzi, G. Loconsole, D. Boscia, and P. Saldarelli (2019). “*Xylella fastidiosa* in olive in Apulia: Where we stand”. *Phytopathology* **109**, pp. 175–186. DOI: [10.1094/PHYTO-08-18-0319-FI](https://doi.org/10.1094/PHYTO-08-18-0319-FI).
- Sarkar, S. et al. (2006). “Biodiversity conservation planning tools”. *Annual Review of Environment and Resources* **31**, pp. 123–59. DOI: [10.1146/annurev.energy.31.042606.085844](https://doi.org/10.1146/annurev.energy.31.042606.085844).
- Sobol, I. M. (2001). “Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates”. *Mathematics and Computers in Simulation* **55**, pp. 271–280. DOI: [10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6).
- Soubeyrand, S. et al. (2018). “Inferring pathogen dynamics from temporal count data: the emergence of *Xylella fastidiosa* in France is probably not recent”. *New Phytologist* **219**, pp. 824–836. DOI: <https://doi.org/10.1111/nph.15177>.
- Stevenson, J. F., M. A. Matthews, L. C. Greve, J. M. Labavitch, and T. L. Rost (2004). “Grapevine Susceptibility to Pierce’s Disease II: Progression of Anatomical Symptoms”. *American Journal of Enology and Viticulture* **55**, pp. 238–245. DOI: [10.5344/ajev.2004.55.3.238](https://doi.org/10.5344/ajev.2004.55.3.238).
- Teviotdale, B. L. and J. H. Connell (2003). “Almond Leaf Scorch”. *ANR University of California*, p. 8106. DOI: [10.3733/ucanr.8106](https://doi.org/10.3733/ucanr.8106).
- Vanhove, M. et al. (2019). “Genomic diversity and recombination among *Xylella fastidiosa* subspecies”. *Applied and Environmental Microbiology* **85**, e02972–18. DOI: [10.1128/AEM.02972-18](https://doi.org/10.1128/AEM.02972-18).
- Virtanen, P. et al. (2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. *Nature Methods* **17**, pp. 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).

- White, S. M., J. M. Bullock, D. A. P. Hooftman, and D. S. Chapman (2017). “Modelling the spread and control of *Xylella fastidiosa* in the early stages of invasion in Apulia, Italy”. *Biological Invasions* **19**, pp. 1825–1837. DOI: [10.1007/s10530-017-1393-5](https://doi.org/10.1007/s10530-017-1393-5).
- White, S. M., J. A. Navas-Cortés, J. M. Bullock, D. Boscia, and D. S. Chapman (2020). “Estimating the epidemiology of emerging *Xylella fastidiosa* outbreaks in olives”. *Plant Pathology* **69**, pp. 1403–1413. DOI: <https://doi.org/10.1111/ppa.13238>.

## Appendices

### I Vector population dynamics

Fig. S1 shows a time series for the population of *Philaenus spumarius* in Mallorca, taken from (López-Mercadal *et al.* 2021) (in blue). Superimposed (in orange) is the assumption used in our model Eq. (2), the  $\delta(t - nT)$ , i.e. every year susceptible vectors appear in the system.



**Figure S1:** Vector dynamics produced by the model compared to field-data from (López-Mercadal *et al.* 2021).

### II Determination of $R_0$

The handicap of determining the basic reproductive number of the model Eq. (2) is that the pre-pandemic fixed point given by  $I_H = I_v = 0$  and  $S_H = S_H(0)$  is not a fixed point of the system of differential equations, because vector population decays, so that the standard methods to compute  $R_0$  such as the Next Generation Matrix (Diekmann *et al.* 2010; Giménez-Romero *et al.* 2022a) do not apply. In (Giménez-Romero *et al.* 2022a) a method was suggested to determine the basic reproductive number in the case of compartmental models of vector-borne transmitted diseases in which the vector population grows or decays. It consists in averaging the instantaneous basic reproductive number over the time of a generation.

To proceed we consider that  $I_H = I_v = 0$ ,  $S_H = S_H(0)$  is indeed a fixed point of the system. Then, the basic reproductive number could be determined, e.g. as shown in (Brauer *et al.* 2016). First, an infectious host infects vectors at a rate  $\beta S_H(0)/N_H$  for a time  $1/\gamma$ . This produces  $\beta S_H(0)/\gamma N_H$  infected vectors. The second stage is that these infectious vectors infect hosts at a

rate  $\alpha N_v(0)/N_H$  for a time  $1/\mu$ , producing  $\alpha N_v/\mu N_H$  infectious hosts per vector. The net result of these two stages is

$$\tilde{R}_0 = \frac{\alpha\beta}{\mu\gamma} \frac{S_H(0)}{N_H^2} N_v(0) = R_0^* \cdot N_v(0) . \quad (1)$$

This result coincides with the value of  $R_0$  obtained using the standard NGM method, that can be applied in this case because we are assuming that we use a nongeneric initial condition that sits at the fixed point of the model.

In practice, our initial condition will never be a fixed point of the model, and, as mentioned above, we will obtain an approximate basic reproductive number, to which we will refer as  $R_0$  using the method suggested in (Giménez-Romero *et al.* 2022a), that consists in calculating the *average* number of secondary infections produced by an infectious host in *one generation*. One first defines an instantaneous basic reproductive number,

$$R_0^{(i)}(t) = \frac{\beta\alpha}{\mu\gamma} \frac{S_H(0)}{N_H^2} N_v(t) = R_0^* N_v(t) , \quad (2)$$

from which the average is simply computed as

$$R_0 = \left\langle R_0^{(i)}(t) \right\rangle \Big|_0^\tau = R_0^* \left\langle N_v(t) \right\rangle \Big|_0^\tau = R_0^* \frac{1}{\tau} \int_0^\tau N_v(t) dt . \quad (3)$$

In our model, the time-dependent vector population can be obtained from Eq. (2),

$$\dot{N}_v = \dot{S}_v + \dot{I}_v = -\mu N_v \implies N_v(t) = N_v(0)e^{-\mu t} , \quad (4)$$

and introducing this expression for  $N_v(t)$  in Eq. (3) the integral can be solved

$$R_0 = \frac{\beta\alpha S_H(0)}{\mu\gamma N_H^2} \frac{N_v(0)}{\mu\tau} (1 - e^{-\mu\tau}) = R_0^* \frac{N_v(0)}{\mu\tau} (1 - e^{-\mu\tau}) , \quad (5)$$

that is an approximated expression to the basic reproductive number for our model, in which the vector population is nonstationary, where, in Eq. (2) and Eq. (5) it has been defined,  $R_0^* = (\beta\alpha S_H(0))/(\mu\gamma N_H^2)$ .

Note that in our model one generation correspond to one year and that  $N_v(0)$  is reset every year.