

# <sup>1</sup> Field and landscape risk factors impacting <sup>2</sup> Flavescence dorée infection : Insights from spatial <sup>3</sup> Bayesian modelling in the Bordeaux vineyards

<sup>4</sup> Hola Kwame Adrakey<sup>1</sup>, Sylvie Malembic-Maher<sup>2</sup>, Adrien Rusch<sup>1</sup>, Jean-Sauveur  
<sup>5</sup> Ay<sup>3</sup>, Luke Riley<sup>4</sup>, Lovasoa Ramalanjaona<sup>1</sup>, and Frederic Fabre<sup>1</sup>

<sup>6</sup> <sup>1</sup>INRAE, Bordeaux Sciences Agro, UMR SAVE, Villenave d'Ornon F-33882,  
<sup>7</sup> France

<sup>8</sup> <sup>2</sup>INRAE, Univ. Bordeaux, UMR BFP, Villenave d'Ornon F-33882, France

<sup>9</sup> <sup>3</sup>INRAE, Institut Agro, Univ. Bourgogne Franche-Comté, UMR CESAEF,  
<sup>10</sup> F-21000, Dijon, France

<sup>11</sup> <sup>4</sup>INRAE, UR BioSP, Equipe OPE - Plateforme ESV, Avignon, France

## Abstract

Flavescence dorée (FD) is a quarantine disease threatening European vineyards. Its management is based on mandatory insecticide treatments and the uprooting of infected plants identified during annual surveys. Field surveys are currently not optimised, as the drivers affecting FD spread in vineyard landscapes remain poorly understood. We collated a georeferenced dataset of FD detection, collected from 34581 vineyard plots over 5 years in South-West France. Spatial models fitted with INLA (Integrated Nested Laplace Approximation) were used to identify local and landscape factors affecting FD detection and infection. Our analysis highlights the importance of sampling period on FD detection and of local practices and landscape context on FD infection. At field scale, altitude and cultivar choice were the main factors affecting FD infection. In particular, the odds ratio of FD infection in fields planted with the susceptible Cabernet Sauvignon, Cabernet Franc or Muscadelle varieties were about twice that in fields planted with the less susceptible Merlot. Field infection was also affected by the field's immediate surroundings (in 150 – 200 m radius circle), corresponding to landscapes of 7 – 12 ha. In particular, the probability of FD infection increased with the proportions of forest and urban land and with the proportion of susceptible cultivars, demonstrating that the cultivar composition impacts FD epidemiology at landscape scale. The satisfactory predictive performance of the model for identifying districts with prevalence of FD detection exceeding 10% of the fields suggests that it could be used to target areas in which future surveys would be most valuable.

**keywords**— Landscape epidemiology, vineyard disease, Distribution modelling, INLA, Varietal landscape, Landscape mosaic

## Introduction

An understanding of the contribution of environmental variables to the presence and spread of any pathogen is crucial for designing efficient surveillance and management disease strategies. Species distribution models are used for this purpose in epidemiology (Purse and Golding 2015, Bebber 2015). These correlative models can be used to disentangle the relative effects of multiple environmental variables (e.g., biotic and abiotic factors) on pathogen epidemiology. They also make it possible to make predictions, for all sites in a region of interest for which the factors studied have been mapped. It therefore provides a basis for the implementation of targeted surveillance and improvements in disease control, by maximising the detection of new cases (Parnell et al. 2014). Widely used for the mapping and management of infectious diseases in humans (Kraemer et al. 2016), distribution modelling has also been used to obtain information about the spatial distribution of emerging plant diseases, such as sudden oak disease in the USA (Meentemeyer et al. 2008, Vaclavik et al. 2010), citrus black spot in South

48 Africa (Martinez-Minaya et al. 2018) and *Xylella fastidiosa* in Europe (Godefroid et al. 2019,  
 49 Cendoya et al. 2020).

50 On the methodological side, the first step in any species distribution modelling approach  
 51 is to gather together diverse sources of information about potential risk factors (e.g., host  
 52 characteristics, cropping practices, climatic data, and land use) and georeferenced records  
 53 of the presence of the pathogen (Purse and Golding 2015, Kraemer et al. 2016). The next  
 54 step involves the use of statistical tools to investigate correlations between these variables  
 55 (Elith and Leathwick. 2009, Dormann et al. 2021). This step is not straightforward as the  
 56 analysis of spatial data is often complicated by a phenomenon known as spatial autocorrela-  
 57 tion (Dormann et al. 2007). Spatial autocorrelation occurs when the values of variables sam-  
 58 pled at nearby locations are not independent from each other. In such a case, one of the  
 59 key assumptions of standard statistical analyses, which is that residuals are independent and  
 60 identically distributed, is violated. The violation of this assumption may bias parameter es-  
 61 timates and can increase type I error rates (falsely rejecting the null hypothesis of no effect)  
 62 (Dormann et al. 2007). A variety of frequentist and Bayesian statistical approaches have been  
 63 developed over the last decades for modelling spatial data while accounting for spatial autocor-  
 64 relation (Dormann et al. 2007, Benguin et al. 2012). Among them, Integrated nested Laplace  
 65 approximation (INLA) (Rue et al. 2009, Lindgren et al. 2011) has emerged in the last decade as  
 66 a highly appealing alternative combining both outstanding computational speed and availability  
 67 in a user-friendly R interface in the R-INLA library (Zuur et al. 2017, Benguin et al. 2012).

68 Phytoplasmas are plant-pathogenic bacteria, pleiomorphic with no cell walls, that belong  
 69 to the class of Mollicutes (Namba 2019). They are obligate parasites invading the phloem  
 70 sieve tube elements of the host plants and colonizing the body of insect vectors. Phytoplas-  
 71 mas are transmitted by phloem feeding hemipteran insects (leafhoppers, planthoppers and  
 72 psyllids) (Weintraub and Beanland 2006) and by vegetative propagation of infected plant ma-  
 73 terial. They are associated with diseases that cause severe economic impacts on many crops  
 74 worldwide (Namba 2019). Flavescence doree (FD), one of the most damaging diseases in Euro-  
 75 pean vineyards, is caused by the Flavescence dorée phytoplasma (taxonomic subgroups 16SrV-C  
 76 and 16SrV-D). FD phytoplasma is transmitted from grapevine to grapevine by the leafhopper  
 77 vector *Scaphoideus titanus* (Chuche and Thiéry 2014). Typical symptoms are leaf yellowing or  
 78 reddening, with downward rolling, incomplete lignification of canes, abortion of flowers, and  
 79 grape wilting. FD disease emerged in South-West France in the 1950s, following the acciden-  
 80 tal introduction of *S. titanus* from North America (Caudwell 1957, Papura et al. 2012). But  
 81 FD phytoplasma was demonstrated to be European, originating from wild plant reservoirs: the  
 82 alder tree *Alnus glutinosa* (Fagales, Betulaceae) and the climbing shrub *Clematis vitalba* (Ranunc-  
 83 culales, Ranunculaceae) from which the phytoplasma was originally transmitted to cultivated  
 84 grapevines (Malembic-Maher et al. 2020).

85 At the field scale, the epidemiology of FD is firstly driven by the dynamics of its am-  
 86 pelophagous vector *S. titanus*. This species is univoltine. The eggs hatch in April on grapevine,  
 87 and there are then five nymphal instars before the first adults appear, usually in June and  
 88 July. The adults live for about one month. The fertilised females lay eggs in the late summer,  
 89 from August to September. Phytoplasmas are acquired passively, from the first larval stage  
 90 onwards, through feeding of the phloem sap of infected grapevine plants where the phytoplasma  
 91 multiply. Once infected, the insects carry and transmit the phytoplasma for the rest of their  
 92 lives (Weintraub and Beanland 2006). The flight activity of the vector depends on vine density  
 93 and canopy architecture (Lessio and Alma. 2004, Lession et al. 2015). The epidemiology of FD  
 94 is also driven by the differential propensity of vine cultivars to act as a source of inoculum  
 95 (Bressan et al. 2005, Galetto et al. 2014). No cultivars are resistant to FD. They rather dis-  
 96 play a continuum from low (e.g. Merlot) to high (e.g. Cabernet-Sauvignon) susceptibility to  
 97 the presence and multiplication of phytoplasma (Eveillard et al. 2016, Ripamonti et al. 2021,  
 98 Oliveira et al. 2019). Once the disease has become chronically installed in a region, *Vitis* spp.  
 99 growing in abandoned vineyards or close to cultivated vineyards can constitute an important  
 100 source of primary infection (Tramontini et al. 2020). Indeed, they provide a reservoir for both  
 101 the phytoplasma and its vector (Lession et al. 2014, Ripamonti et al. 2020). Although dispersal  
 102 abilities of the vector is rather small (< 30 m), longer range dispersals have been observed, sug-

103 gesting that landscape structure may also affect the epidemiology of FD (Lession et al. 2014).  
 104 In the last decades, FD has spread throughout European vineyards (Jeger et al. 2016). Due  
 105 to the severe economic consequences of the disease, FD phytoplasma is classified as a quarantine  
 106 organism in Europe since 1993. There is currently no means of curing plants of FD phytoplasma.  
 107 The disease is therefore controlled principally by four mandatory measures: (i) the planting of  
 108 disease-free material, (ii) the application of insecticides to kill the vector, (iii) the establishment  
 109 of annual vineyard surveys for monitoring plant infection and (iv) the uprooting of infected  
 110 plants. Two factors complicate the detection of FD in the field. First, FD cocirculates in Euro-  
 111 pean vineyards with Bois noir (BN), another phytoplasma disease that has similar symptoms but  
 112 does less economic damage (Quaglino et al. 2013, Tramontini et al. 2020). Molecular tests in-  
 113 volving the detection of pathogen DNA in a real-time multiplex PCR assay are therefore required  
 114 to confirm the presence of FD (Pelletier et al. 2009). Second, typical FD symptoms appear only  
 115 during the summer of the year following inoculation (Schvester et al. 1969, Morone et al. 2007,  
 116 Tramontini et al. 2020). Plants newly infected during the spring and early summer of year  $n$   
 117 therefore constitute a source of inoculum before the symptoms of the disease become visible,  
 118 late in the summer of year  $n + 1$ , when the monitoring campaigns are performed. Once the  
 119 diagnosis of FD is confirmed, the infected plants must be removed and insecticide treatments  
 120 are mandatory within defined perimeters, with no financial compensation for the growers. More  
 121 than 70% of French vineyards are treated with insecticides against *S. titanus*, with consequences  
 122 for both the environment and human health (Desneux et al. 2007, Tang et al. 2021). Surveil-  
 123 lance has been intensified over the last decade under the supervision of legal authorities, leading  
 124 to a reduction of the number of annual treatments.

125 In this study, we applied a distribution modelling approach to improve our understanding  
 126 of FD epidemiology and, more specifically, gain a quantitative view of the field and landscape  
 127 factors impacting the probabilities of FD detection and infection at field scale. Based on the  
 128 literature, we first hypothesized that the probability of FD detection improves as fall approaches.  
 129 We also hypothesized that key local factors would impact the probability of FD infection with  
 130 a higher probability of infection associated to older and denser plantation as well as cultivation  
 131 of more susceptible cultivars. We also hypothesized that the probability of infection would  
 132 increase with the proportion of susceptible cultivars, the proportion of semi-natural habitats or  
 133 the proportion of urban areas in the surrounding landscape as these habitats may act as FD  
 134 reservoirs. In addition, we hypothesized that higher level of fragmentation of these potential  
 135 reservoirs at the landscape scale would limit vector spillover and therefore be associated with  
 136 lower probability of infection (Tscharntke and Brandl 2004). To test these hypotheses, we fitted  
 137 distribution models using INLA to a spatial dataset collected from 34581 vineyard fields over 5  
 138 years of mandatory vineyard surveys. Finally, we investigated whether the model could identify  
 139 sites with higher probability of infection for targeted surveillance to improve current management  
 140 strategies.

## 141 Materials and methods

### 142 Flavescence dorée monitoring and detection

143 In the vineyards of Bordeaux in South-West France, FD surveys are performed by professional  
 144 organisations known as GDONs (Groupements de Défense contre les Organismes Nuisibles),  
 145 under the supervision of the French Ministry of Agriculture. The "GDON des Bordeaux"  
 146 has been in charge of FD monitoring since 2011 in an area of 364718 ha of which 83912 ha  
 147 (23%) were cultivated with grapevine in 2016 according to the "Casier Viticole Informatisé"  
 148 (CVI). CVI is a GIS database created by the French directorate general of customs which  
 149 provides a comprehensive history of each legal piece of land cultivated with grapevine. It  
 150 gathers information on the spatial boundary, size, year and density of plantation and cultivars  
 151 used in each legal field. The region monitored by GDON is subdivided into 347 "districts", with  
 152 a mean area of 1048 ha and 10% and 90% quantiles of 331 ha and 1849 ha, respectively (Fig.  
 153 1A). Although these districts serve as a basis for the practical organization of FD monitoring  
 154 performed by the GDON in the vineyards, the basic observational unit is the field. In the

following, we considered the annual survey realized from 2012 to 2016. Each year, the monitoring strategy of GDON was nearly identical. It was conducted by a team of trained inspectors between August and October, the period of the year most favourable for the detection of FD symptoms (Tramontini et al. 2020). The fields selected each year  $n$  by the GDON can be separated into 2 categories. The first category consists of the fields already surveyed and detected infected the previous year  $n-1$ . Once an infected field is detected, GDON inspectors check the following year that mandatory control measures were applied. These re-inspection data were not considered in our study. The second category consists of the fields not surveyed in the previous years. Approximately 10% of the vineyard area was newly inspected each year. These fields were considered in our study. The survey involved the inspection of most of the fields in large parts of newly selected districts each year, most often regardless of prior information on the presence of FD (note that "spontaneous" reporting of FD by the winegrowers is unusual). Overall a mean of 6916 fields are newly inspected each year, a number ranging from 5916 in 2013 to 8712 in 2012. The fields had a mean area of 0.77 ha, with 10% and 90% quantiles of 0.16 ha and 1.6 ha, respectively.

When the inspection teams survey a given area, they inspect a high proportion of fields. We quantify this sampling effort by the proportion of the vineyard sampled by GDON inspectors. For each of the 34581 fields sampled, the sampling effort was assessed by the ratio between (i) the total vineyard area sampled by GDON inspectors in a radius of 1 km around each field (as assessed from the area of the polygons drawn by inspectors during their survey, see below) and (ii) the total size of the vineyard in this same radius (as given by the area of legal piece of land cultivated with grapevine provided by the CVI). The mean sampling effort was 0.58, with 10% and 90% quantiles of 0.38 and 0.74, respectively. As previously stated, the basic observational unit was the field (Fig. 1B). For each field surveyed, inspectors walk through the whole field for visual detection of the symptoms and draw its polygon on an GIS software. If no symptoms are detected, the inspectors do not sample any plants and the field is declared as "uninfected". If symptoms are detected, symptomatic plants are marked for uprooting and symptomatic leaves from one to five plants per field are collected and pooled into a single sample. The detection of the FD and BN phytoplasmas in samples is performed by accredited laboratories, with a real-time PCR triplex molecular test derived from that described by (Pelletier et al. 2009). Each field is then classified as either infected with FD (if the molecular test is positive for FD phytoplasma, meaning that at least one of the plant in the sample is infected with FD) or uninfected with FD. Overall, FD was detected in 7.6% of the 34581 fields surveyed and BN detected in 4.5%.

## Explanatory variables at field scale

In addition to the georeferenced records of FD detection, we collected a set of potentially important local explanatory variables characterizing the production situations of each field (Table 1). First, the **altitude** was obtained from topography data provided by IGN (Institut national de l'information géographique et forestière). Second, the **AOC** (Appellation d'Origine Contrôlée), a French certification defining the geographic area of production for particular wine label and common and specific guidelines (e.g. proportions of particular cultivars, growing practices, wine-making practices, etc.) was noted, to provide information about the socioeconomic context of production. The AOCs were grouped into six levels on the basis of geographic proximity. Third, the CVI was used to determine, for each field, the **cultivars** grown (7 levels, Fig. 1 B), plant **density** and the **age** of the plantation. The protocol for extracting these variables from the CVI is described in Supplementary Material (Text S1, Table S1). Fourth, the type of viticultural **Practice** (organic or conventional) in each field was provided by the GDON. This information was available only for 2016, but we assumed that growers maintained the same practices over the entire time period considered here.

## Explanatory variables at landscape scale

We created annual land covers maps of the region monitored by GDON des Bordeaux. These maps are raster maps with a 10-metre resolution. The maps were created by overlaying the CVI database and the France land cover map for 2017 (OSO map, (Inglada et al. 2017)). We used

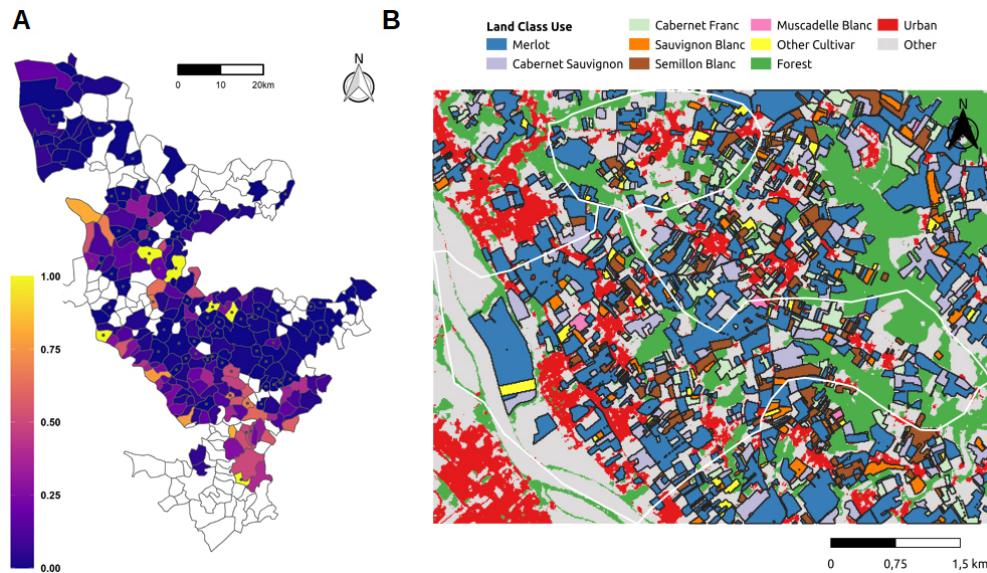


Figure 1: Field-based Flavescence dorée (FD) data at the regional and district scales. A: Map of the study region in South-West France. The map shows the borders of the 347 districts located within the GDON des Bordeaux area. The colour scale indicates the prevalence of FD detection per district in each of the 239 districts in which FD surveys were performed from 2012 to 2016. Unsurveyed districts are shown in white and districts with less than 10 fields are shown with a dot. B: Maps at landscape scale (over 2 entire districts depicted by white border) illustrating the vineyard plots, classified according to their cultivar (seven levels), and the other categories of land use (three levels).

207 the OSO map for 2017 because (i) the landscape was expected to be constant within the time  
 208 frame of the study and (ii) the map resolution was more precise in 2017 (pixels of 10 m per side)  
 209 than previous years (pixels of 20 m per side). For each year from 2012 to 2016, the CVI database  
 210 was used to map the pixels representing grapevine cultivars. Seven classes depicting the cultivar  
 211 used in each pixel during the year considered were distinguished (as listed in Table 1) while  
 212 an independent eighth class indicates whether the pixel belongs to an organic or conventional  
 213 vineyard farm (data provided by GDON, see previous section). Each pixel of the GDON region  
 214 not attributed to a vineyard class was classified into 3 separate classes (forest, urban or other  
 215 land use) depending on its classification in the OSO map (Fig. 1B). The resulting land cover  
 216 map for 2016 is illustrated in Figure S1.

217 These annual land cover maps were used to characterize the landscape surrounding each  
 218 field. We calculated composition and configuration metrics in circles of 15 increasing radius (50,  
 219 100, 150, 200, 250, 300, 500, 1000, 1500, 2000, 2500, 3000, 4000, 5000 and 6000 m) centered  
 220 on the barycentre of each field  $i$  ( $i = 1, \dots, 34581$ ) (Table 1). Thereafter, “scale” defined the  
 221 radii around field barycentres (*i.e.* the extent of the landscapes). For each field  $i$  and scales  
 222 considered, we calculated the percentage of land occupied by vineyard **L\_vineyard** (whatever  
 223 the cultivar grown), forest **L\_forest** and urban areas **L\_urban**. These metrics are assessed as  
 224 the proportion of pixels attributed to each class within the circle. We also calculated the total  
 225 number of patches **L\_np** of these 3 classes, a simple measure of landscape fragmentation. A  
 226 patch refers here to a continuous area of the landscape attributed to a single class use. Finally,  
 227 we calculated the percentage of land (i) cultivated with the cultivar Merlot **L\_merlot** and  
 228 (ii) organically managed **L\_organic** within the total area devoted to grapevine cultivation  
 229 in the circle considered. These metrics were computed with the R package *landscapemetrics*  
 230 (Hesselbarth et al. 2019).

Table 1: Local and landscape explanatory variables used to characterize the vineyard fields.

Type	Variables	Description	Mean (sd)/levels <sup>1</sup>	Source
Local variables	<b>Area</b>	Field area	0.77 ha (0.77)	GDON
	<b>Year</b>	Year of the first field inspection	Five levels	
	<b>Season</b>	Summer or Autumn inspection	Two levels <sup>2</sup>	
	<b>Week</b>	Week number within year	Eleven levels <sup>2</sup>	
	<b>Practice</b>	Organic or conventional practices	Two levels <sup>2</sup>	
	<b>AOC</b>	Appellation d'origine contrôlée	Six levels <sup>3</sup>	INAO
	<b>Age</b>	Age of the plantation at inspection	23 years (14)	CVI
	<b>Cultivar</b>	Grape cultivar	Seven levels <sup>4</sup>	
	<b>Density</b>	Density of the plantation	4131 plants/ha (1008)	
	<b>Altitude</b>	Altitude	59.5 m (29)	IGN
Landscape scale	<b>L_urban</b>	Percentage of urban area	8.2 (10)	CESBIO
	<b>L_forest</b>	Percentage of forest area	5.6 (9.8)	
	<b>L_vineyard</b>	Percentage of vineyard area	66.7 (20.2)	
	<b>L_merlot</b>	Percentage of merlot area in the vineyard	62.2 (32)	CVI
	<b>L_organic</b>	Percentage of organic practice area in the vineyard	8 (21.3)	
	<b>L_np</b>	Number of vineyard, forest and urban patches	12.5 (5.5)	

<sup>1</sup> Values for the landscape covariates are given for a scale of 150m.

<sup>2</sup> Season : Summer here is the period from July August (28719 fields) and Autumn corresponds to September-October period (5862 fields). Week : 11 levels from week 32 to week 42. Practice : Organic (2745 fields) or conventional (31836 fields) farming practices.

<sup>3</sup> The 11 controlled designation of origin located in the GDON des Bordeaux area were grouped by spatial proximity in six levels in order to have at least 700 fields in each level. These levels are AOC1 (Côtes de Bordeaux-Saint-Macaire, 1936), AOC2 (Premières Côtes de Bordeaux, 5032 fields), AOC3 (Loupiac, Sainte-Croix-du-Mont and Graves supérieures, 725 fields), AOC4 (Fronsac and Cannon Fronsac, 743 fields), AOC5 (Côtes de Bordeaux Blaye, 5858 fields) and AOC6 (Bordeaux, 20287 fields).

<sup>4</sup> Cultivars include merlot (21181 fields), cabernet-sauvignon (5266 fields), cabernet-franc (2491 fields), muscadelle (270 fields), sauvignon (1985 fields), semillon (2587 fields) and other cultivars (801 fields) not previously listed.

## 231 Statistical analysis

### 232 Bayesian inference with INLA

233 Integrated nested Laplace approximation (INLA) is a computationally efficient method for fitting  
 234 complex spatial models within the Bayesian paradigm (Rue et al. 2009, Lindgren et al. 2011).  
 235 It offers a faster alternative to Markov Chain Monte Carlo (MCMC) methods as, in essence,  
 236 INLA replaces stochastic sampling with deterministic approximation based on a clever use of the  
 237 Laplace approximation and on numerical integration (Benguin et al. 2012). INLA may be used  
 238 to fit a large class of latent Gaussian models in a Bayesian framework. Moreover, the method is  
 239 implemented in the R package INLA (<https://www.r-inla.org/>) (Lindgren and Rue 2015) which  
 240 allows fitting models almost as easily as the base R functions for generalized linear models.

241 Latent Gaussian models include in particular several classes of models accounting for spatial  
 242 autocorrelations. Indeed, combined with the stochastic partial differential equation approach  
 243 (SPDE), one can accommodate with INLA all kinds of geographically referenced data, includ-  
 244 ing areal and geostatistical ones, as well as spatial point process data (Lindgren et al. 2011,

245 Lindgren and Rue 2015). Here, we employ INLA to fit spatial logistic regression in a continuous  
 246 spatial domain. References such as (Benguin et al. 2012) and (Zuur et al. 2017) provide  
 247 comprehensive text on the subject for applied ecologists and phytopathologists.

248 **Spatial logistic regression**

249 Let  $y_i$  be the FD detection status of the field  $i$  at location  $s_i$  on its year of first inspection  $year_i$ ,  
 250 i.e. the field  $i$  is classified as either infected (value 1) or uninfected (value 0). We assumed that  
 251  $y_i$  follows a Bernoulli random variable and modelled the probability  $\pi_i$  of FD detection in field  
 252  $i$  using a logit link function. The generic model can be written as

$$\begin{cases} y_i \sim \text{Binomial}(1, \pi_i), & i = 1, \dots, 34581 \\ \text{logit}(\pi_i) = \beta X_i + a(year_i) + u(s_i) \\ u(s) \sim \text{GMRF}(0, \mathbf{Q}^{-1}(\phi, \sigma_u^2)) \\ a(year_i) \sim \text{Normal}(0, \sigma_{year}^2) \end{cases} \quad (1)$$

253 In this model,  $X_i$  is the vector of explanatory variables for field  $i$ , and  $\beta$  is the vector of  
 254 associated coefficients to be estimated. All the explanatory variables  $X_i$  listed in Table 1 were  
 255 treated as fixed effects except the year of first field inspection treated as independent random  
 256 effects  $a(year_i)$ . The term  $u(s_i)$  is the spatial random effect for field  $i$ . It endows the model with  
 257 a spatial dependence between neighbouring fields not explained by the explanatory variables. In  
 258 R-INLA, a computationally convenient Gaussian Markov random field (GMRF) representation  
 259 is used to accurately approximate a Gaussian random field with spatial variance and autocor-  
 260 relation characterised by the Matérn covariance function (Lindgren et al. 2011). This approxi-  
 261 mation relies on basis functions anchored at a set of discrete points corresponding the nodes of  
 262 a mesh dividing the study area into a large number of non-overlapping triangles (Fig. S2). Its  
 263 use allows inference about the latent spatial field over the entire continuous domain of interest,  
 264 where the spatial interpolation between the Gaussian variables located at the nodes is linear.

265 The GMRF can be regarded as a multivariate Gaussian distribution (with dimension equal  
 266 to the number of nodes) with zero mean and sparse precision matrix  $\mathbf{Q}$  that depends on two  
 267 positive hyperparameters  $\sigma_u^2, \phi$  describing the variance and the correlation range (the distance  
 268 beyond which pairs of observations are approximately not spatially correlated any more), re-  
 269 spectively (Lindgren et al. 2011, Krainski et al. 2018). The spatial dependence is encoded in  
 270 the matrix  $\mathbf{Q}$  of size  $N \times N$  by using the Matérn covariance function based on the first-order  
 271 Bessel function  $K_1$ . Accordingly, the covariance between  $u(s_i)$  and  $u(s_j)$  depends on the  
 272 two unknown hyperparameters and on the Euclidean distance  $d(i, j)$  between the two fields  
 273 through  $cov_{\text{Matern}}(u(s_i), u(s_j)) = \sigma_u^2 \sqrt{8} (d(i, j)/\phi) K_1(\sqrt{8}(d(i, j)/\phi))$ . We remark that different  
 274 parametrizations exist for the Matérn covariance, and the preceding one offers the advantage  
 275 of relatively simple intuitive interpretation. The corresponding correlation function is obtained  
 276 by replacing  $\sigma_u^2$  by 1 in  $cov_{\text{Matern}}$ . In our model (1), we refer to the sparse precision matrix  
 277 resulting from this covariance model as  $Q(\phi, \sigma_u^2)$ .

278 Bayesian inference requires to specify prior distributions for the model parameters and hy-  
 279 perparameters (i.e. parameters of prior distributions). These choices are less crucial with  
 280 large datasets as the effect of prior choice on the posterior estimates is expected to wane as  
 281 the sample size increases. For the parameters  $\beta_j$  and the hyperparameter  $\sigma_{year}^2$ , we used  
 282 the default internal vague priors recommended in R-INLA,  $\text{Normal}(0, precision = 10^{-3})$  and  
 283  $\text{Log-Gamma}(1, 0.00005)$ , respectively. We used penalized complexity priors (PC-priors) for the  
 284 range and the variance of the spatial random effect (Simpson et al. 2017, Fuglstad et al. 2019).  
 285 PC-priors can be defined via intuitive probability statements. As few fields are less than  
 286 0.25 km apart, we assume that the probability that  $\phi$  is  $< 0.25$  km is 0.01, leading to a  
 287  $PC - prior(0.25, 0.01)$  for  $\phi$ . Moreover, we assume that the probability that  $\sigma_u$  is  $> 1$  is  
 288 0.01, leading to a  $PC - prior(1, 0.01)$  for  $\sigma_u$ .

**289 Model comparison**

290 The baseline spatial model  $M_0$  corresponds to equation 1 with 13 explanatory variables, including  
 291 7 local variables *Area*, *Age*, *Practice*, *Density*, *Altitude*, *Cultivar* and *AOC* characterizing  
 292 the field and 6 landscape variables *L\_Organic*, *L\_Vineyard*, *L\_np*, *L\_Forest*, *L\_Merlot* and  
 293 *L\_Urban* characterizing the landscape within a zone of radius  $r$ . No interaction between ex-  
 294 planatory variables were considered and all continuous explanatory variables were standardized.  
 295 In all, 16 models  $M_0$  corresponding to landscape scales  $r$  ranging from 0 to 6000 m were consid-  
 296 ered. The case  $r = 0$  corresponds to a case in which only local variables are considered. These  
 297 16 spatial models were compared to their 16 non-spatial counterparts for which the spatial  
 298 random effect  $u(s_i)$  was removed. We estimated the Watanabe Information Criterion (WAIC)  
 299 (Watanabe 2010, Gelman et al. 2014), which penalises model complexity, and the Deviance In-  
 300 formation Criterion (DIC) (Spiegelhalter et al. 2014) to compare all the models fitted.

301 In addition to the baseline model  $M_0$ , we considered a model  $M_1$  to refine the study of the  
 302 period of field inspection. In model  $M_1$ , the variable *Season* is eliminated from the set of fixed  
 303 effects  $X_i$  and replaced by a random walk of order 1 indexed by the weeks of inspection. The  
 304 weeks range from week 32 (second week of august) to week 42 (third week of october). The  
 305 notation  $k_i$  denotes thereafter the week  $k$  of inspection of field  $i$  ( $k = 32, \dots, 42$ ). In model  
 306  $M_1$ , the logit equation in 1 is replaced by  $\text{logit}(\pi_i) = \beta X_i + a(\text{year}_i) + u(s_i) + w(k_i)$  with  
 307  $w_k = w_{k-1} + \nu_k$  ( $k = 33, \dots, 42$ ) and  $\nu_k \sim \text{Normal}(0, \sigma_k^2)$ . In this model,  $\sigma_k^2$  is an additional  
 308 hyperparameter for which we used the default internal INLA prior  $PC - \text{prior}(0.5, 0.01)$ .

**309 Preliminary analysis**

310 A set of preliminary analysis was realized with the baseline model  $M_0$  (not shown). First,  
 311 a test for multicollinearity was performed, by fitting the model  $M_0$  without spatial random  
 312 effect for the 16 landscape scales considered. The variance inflation factors for each variable  
 313 were  $< 3$  (except for *Practice* and *L\_Organic* at the landscape scale of 50 m radius with values  
 314  $< 5$ ), indicating an absence of collinearity between the explanatory variables (Zuur et al. 2009).  
 315 Second, we compared the model  $M_0$  with a binomial distribution to an alternative model re-  
 316 lying on a zero-inflated binomial (ZIB) distribution, to account for a possible excess of zeros  
 317 (Martin et al. 2005), as 92.5% of the fields were FD-negative. We found that the binomial  
 318 models outperformed the ZIB models, suggesting the effectiveness of the surveillance. Finally,  
 319 we investigated the effect of six meshes on parameter estimations. The meshes differ in the  
 320 largest triangle edge length (parameter `max.edge`) and the minimum allowed distance between  
 321 points (parameter `cutoff`) (Fig. S2). Overall, the parameters estimates were consistent from the  
 322 coarsest to the finest meshes (Fig. ??).

**323 Predictive performance of the models**

324 The predictive performance of the baseline model  $M_0$  was estimated at two scales (field, district)  
 325 by cross-validation. The model was trained on 80% of the data (training set) and tested on the  
 326 remaining 20% (testing set). We repeated this partitioning for cross-validation 50 times. We  
 327 first assessed the ability of the model to predict the FD detection status of individual fields with  
 328 precision-recall curve (PR curve). PR curve were preferred over ROC curve as they are more  
 329 informative when evaluating binary classifiers on imbalanced datasets which is the case in our  
 330 study with 7.5% of FD detection (Saito and Rehmsmeier 2015). For a given threshold (between  
 331 0 and 1) above which a field is classified as infected by the model, the recall, also known as  
 332 sensitivity, is the probability that a field is classified as infected by the model when FD was  
 333 detected in that field. Similarly, the precision, also known as positive predictive value, is the  
 334 probability that FD is detected in a field when it is classified as infected the model. A PR curve  
 335 is then a plot of the precision (y-axis) and the recall (x-axis) for all possible thresholds between 0  
 336 and 1. Predictive performance was summarised by the area under the PR curve (AUC). We also  
 337 evaluated predictive performance at the district scale. Using the same 80% – 20% rule applied  
 338 at the district scale, we evaluated the ability of the model to predict if the proportion (thereafter  
 339 termed prevalence) of FD detection in a district exceeded a given threshold. Thresholds were

<sup>340</sup> varied from 3% to 15%, and, for each threshold used, we calculated precision and recall for each  
<sup>341</sup> partitioning.

## <sup>342</sup> Results

### <sup>343</sup> Model selection, goodness-of-fit and spatial field

<sup>344</sup> The baseline model  $M_0$  was fitted to 16 landscape scales ranging from 0 (only local variables  
<sup>345</sup> considered) to 6000 metres. As a first step, we compared the 16 spatial models  $M_0$  to their  
<sup>346</sup> 16 non-spatial counterparts for which the spatial random effect  $u(s_i)$  was removed. The AIC  
<sup>347</sup> and DIC values of the spatial model were much lower at all scales (by at least 400 points),  
<sup>348</sup> highlighting the importance of the spatial structure underlying FD detection in our data. The  
<sup>349</sup> spatial model fitted with landscape variables at the 150 m scale was the best of the models  
<sup>350</sup> fitted for both metrics considered (Fig. 2 A). Similar results were obtained with model  $M_1$ . The  
<sup>351</sup> second best landscape scale was 200 m for both models, parameter estimates being consistent  
<sup>352</sup> between  $M_0$  and  $M_1$  at these two scales (Fig. S5). Thereafter, we then focused on the results  
<sup>353</sup> obtained with a landscape scale of 150 metres, corresponding to a landscape area of 7.06 ha  
<sup>354</sup> containing 6 fields on average (90% confidence interval [2 – 12] fields). Note also that, at this  
<sup>355</sup> scale, the continuous explanatory variables were weakly correlated (Fig. S4).

<sup>356</sup> The goodness-of-fit of the spatial model for a 150 m landscape scale was highly satisfactory.  
<sup>357</sup> In particular, the coefficient of determination for the linear regression between the observed and  
<sup>358</sup> adjusted prevalences of FD detection at the district scale was 0.99. Furthermore, this regression  
<sup>359</sup> line is very close to the  $y = x$  line, suggesting that the model explains most of the variability of  
<sup>360</sup> the data at district scale (Fig. 2B).

<sup>361</sup> Compared to the non-spatial model, the spatial model strongly reduces spatial autocorre-  
<sup>362</sup> lation as evidenced by comparing the spatial variograms of the residuals of both models (Fig.  
<sup>363</sup> 2C). However, a weak spatial autocorrelation remains for distances of up to 3.6 km, the esti-  
<sup>364</sup> mated range (95% confidence interval [3.1 – 4.2]). This is distance beyond which the correlation  
<sup>365</sup> between the FD detection status of two fields becomes negligible. The shape of the Matérn  
<sup>366</sup> function is itself instructive (Fig. 2D). In particular, the coefficient of correlation between the  
<sup>367</sup> detection statuses of fields located up to 1 km apart is  $> 0.5$ .

<sup>368</sup> The distribution of spatial random effects provided evidence of a strong spatial pattern  
<sup>369</sup> (Fig. 3A). Positive effects are clustered in the southern and western regions of the GDON  
<sup>370</sup> des Bordeaux area, whereas negative effects are clustered in the north-western and the eastern  
<sup>371</sup> regions. On the logit scale, the spatial random effect ranges from  $-1.33$  (5% quantile) to  $1.44$   
<sup>372</sup> (95% quantile). These values correspond to odds ratios of 0.25 and 4.23, respectively. Moreover,  
<sup>373</sup> the variation on the standard deviation is due to the clustering of FD detection cases over the  
<sup>374</sup> study area (Fig. 3B).

### <sup>375</sup> Intra- and inter-annual effects on Flavescence doree detection

<sup>376</sup> Inferences indicate that the probability of inspectors detecting FD during August is lower (odds  
<sup>377</sup> ratio 0.24) than in September-October (Fig. 4A, last panel). The random walk used in model  
<sup>378</sup>  $M_1$  allows to refine the estimation of the effect of the weeks of field inspections and highlights  
<sup>379</sup> a sharp increase from the first week of September (Fig. 5A). Moreover, the probability of FD  
<sup>380</sup> detection was remarkably consistent for each of the five years surveyed except 2014, when it was  
<sup>381</sup> somewhat significantly lower (Fig. 4A, last panel).

### <sup>382</sup> Flavescence doree risk factors at the field scale

<sup>383</sup> Local variables describe the physical and agronomic characteristics of each field in the vineyard  
<sup>384</sup> (Table 1). Their effects (posterior mean, 95% credible intervals and probability of being positive)  
<sup>385</sup> are displayed in Fig. 4A. The corresponding odds ratio are obtained as the exponent of the  
<sup>386</sup> posterior means. The probability of FD detection increases with the age of the plantation and  
<sup>387</sup> its area but slightly decreases with planting density. No significant effect of the farming practice

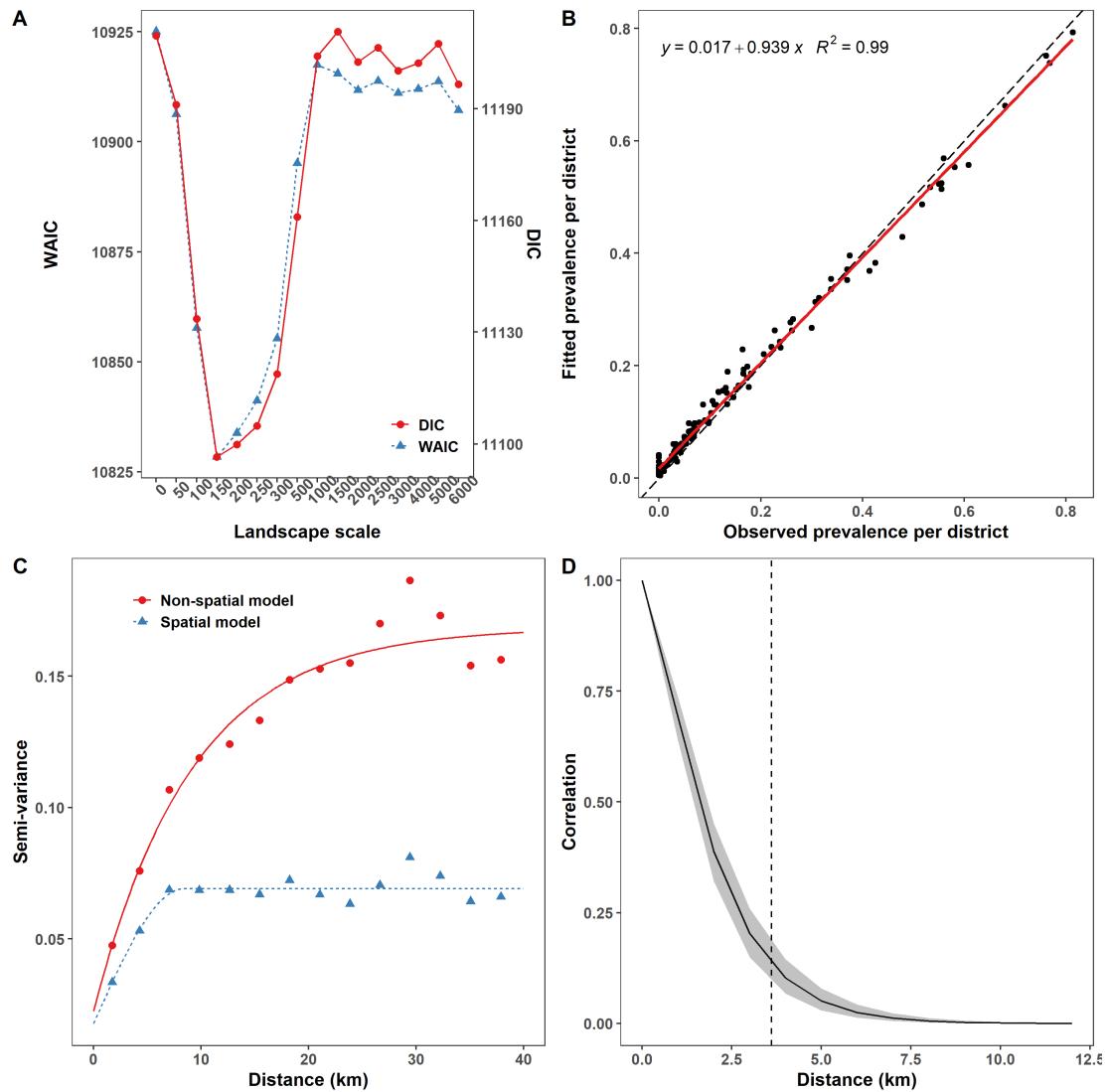


Figure 2: Comparison and fit of the spatial logistic models  $M_0$ . A: Watanabe information criterion (WAIC) and Deviance information criterion (DIC) for the spatial model  $M_0$  in 16 landscapes of increasing scales ranging from 0 m (only local explanatory variables considered) to 6000 m. B: Posterior mean estimate of the adjusted prevalence of FD detection against the observed prevalence aggregated at district scale. The fitted line between the adjusted and observed prevalences is shown in red and the  $y = x$  line is shown in black. C: Sample variogram of the Pearson's residuals obtained for the spatial and non-spatial model  $M_0$  at 150 m. D: The Matérn correlation function and its 95% credible band obtained with model  $M_0$  at 150 m. The dashed vertical line indicates the estimated range.

388 was observed. The probability of FD detection decreases significantly with altitude (Fig. 4B).  
 389 A 20 m increase in altitude is associated to an odds ratio of 0.72.

390 We analysed the effect of cultivar choice at field scale using Merlot, the most widespread  
 391 cultivar in Bordeaux, as the reference (Fig. 4A, second panel from bottom). The odds ratio of  
 392 FD detection in fields planted with Cabernet Sauvignon, Cabernet Franc and Muscadelle were,  
 393 on average, 2.76, 2.29 and 2.08, respectively, compared to fields planted with Merlot. These  
 394 differences were all highly significantly positive. Conversely, the odds ratio of FD detection in  
 395 a field planted with Semillon were 0.76. No significant differences were found between Merlot  
 396 and the cultivars Sauvignon and others (a class containing all the other minority cultivars).

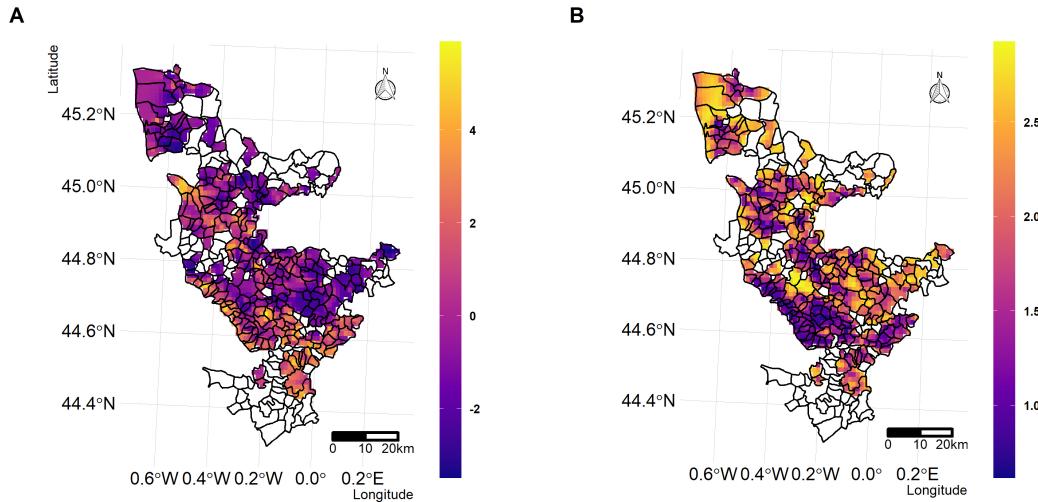


Figure 3: Spatial random effect estimated in the GDON des Bordeaux area for the baseline model  $M_0$  for a landscape scale of 150 m. No data were available for the districts in white. A: Posterior mean of the spatial random effect. B: Posterior standard deviation of the spatial random effect.

Finally, the variable AOC was used as a proxy for the socioeconomic conditions of vine cultivation, with AOC Bordeaux as the reference. For the six AOC levels considered, the probability of FD detection was significantly higher only in the south-western part of the GDON des Bordeaux area containing the AOCs Loupiac, Graves supérieures and Sainte-Croix-du-Mont (Fig. 4A, fourth panel). The odds ratio of FD detection in these AOCs were 2.42 compared to the AOC Bordeaux.

#### Flavescence doree risk factors at the landscape scale

The effects of the variables describing the composition and configuration of the landscape surrounding a field within a landscape scale of 150 m are displayed in Fig. 4A (third panel). The probability of FD detection increases with the proportions of forest and urban land and, more slightly, with the proportion of organic fields. Conversely, the probability of FD detection decreases substantially with the proportion of vineyards. These effects are visualised in Fig. 4B. For an increase of 20 percentage points in the proportion of forest, urban areas, or of fields with organic practices, the odds ratio of detection are 1.28, 1.34 and 1.13, respectively whereas a 20 percentage points increase in the proportion of vineyards is associated to an odds ratio of 0.84.

We also tested the effect of the varietal composition of the landscape throughout the proportion of Merlot area in the vineyard. The probability of FD detection decreases substantially with increases in the proportion of fields planted with Merlot (Fig. 5 B). An increase of 20 percentage points in the proportion of fields planted with Merlot, for a landscape scale of 150 m is associated to an odds ratio of 0.93.

#### Predictive performance of the model

We evaluated the ability of the model to predict the FD detection status of individual fields. The corresponding PR curve indicates a moderate predictive performance (Fig. S6A), as summarised by the AUC of 0.59 (95% credible interval [0.54, 0.64]). The predictive performance was better at district scale. Specifically, we evaluated the ability of the model to predict if the prevalence of FD detection in a district exceeds a given threshold (Fig. S6B). Recall decreases with the threshold, whereas precision remains constant. For a threshold of 10%, the posterior mean recall (*i.e.* sensitivity) is 89%. This is the probability of the model correctly classifying a district as

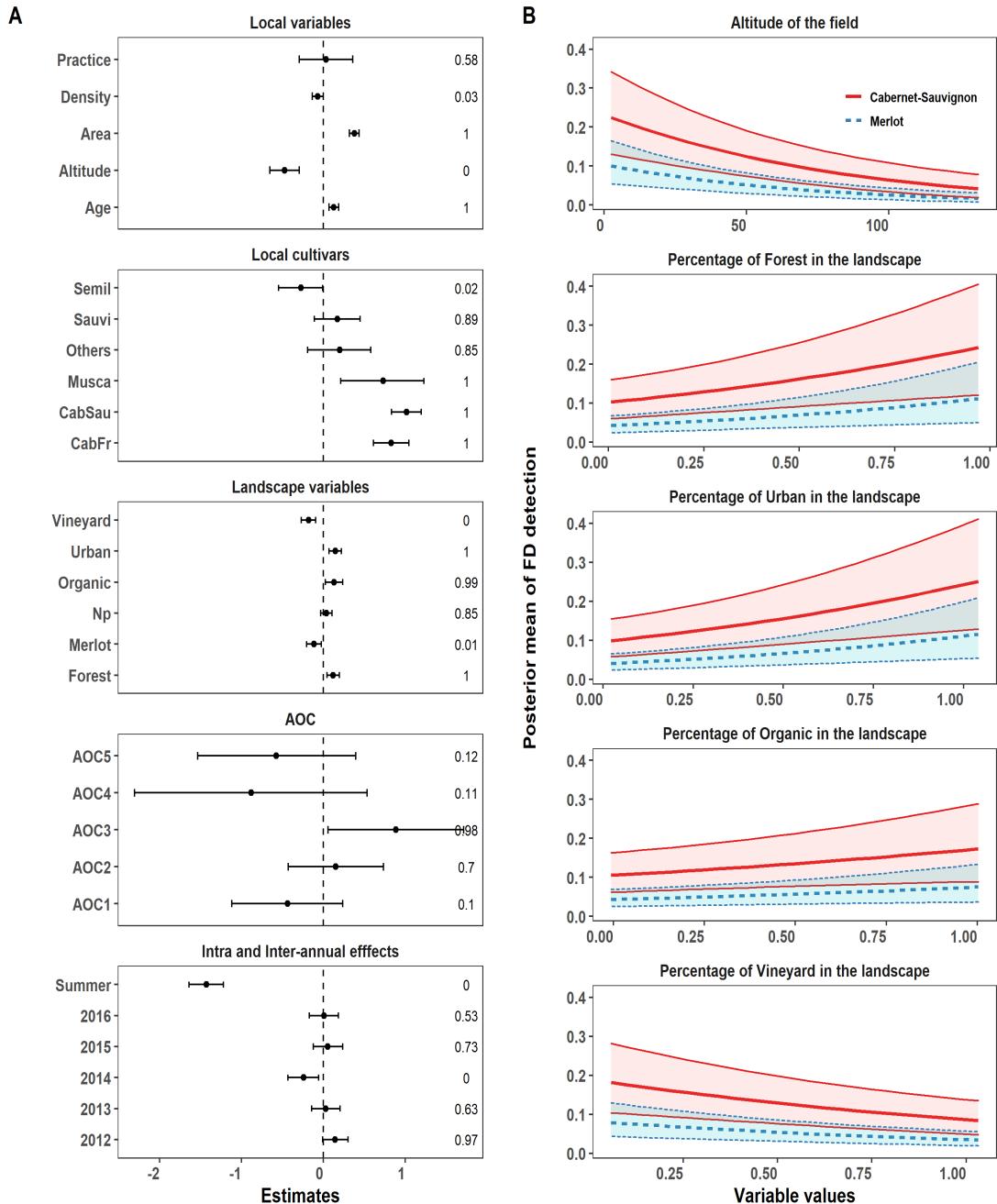


Figure 4: Estimation of the parameters of the baseline model  $M_0$  for a landscape scale of 150 m. A: Fixed and random effects of the local and landscape explanatory variables. For each variables, the posterior mean (dots) and 95% credible intervals (solid lines) are displayed, together with the posterior probability of the effect being positive. The dashed line corresponds to the value 0. B: Effects of altitude, proportions of forest, urban area, organic fields and vineyard in the landscape on the probability of FD detection for the two most widespread cultivars (Merlot and Cabernet Sauvignon) in vineyard plots under conventional practices (mean probability and 95% credible bands). The effects were estimated by fixing all other numerical variables at their mean values, the variable AOC to Bordeaux, the variable Practice to conventional and considering a sampling period during the autumn. All variables and their levels are described in Table 1. We dropped the "L\_" from the landscape variables to lighten notations.

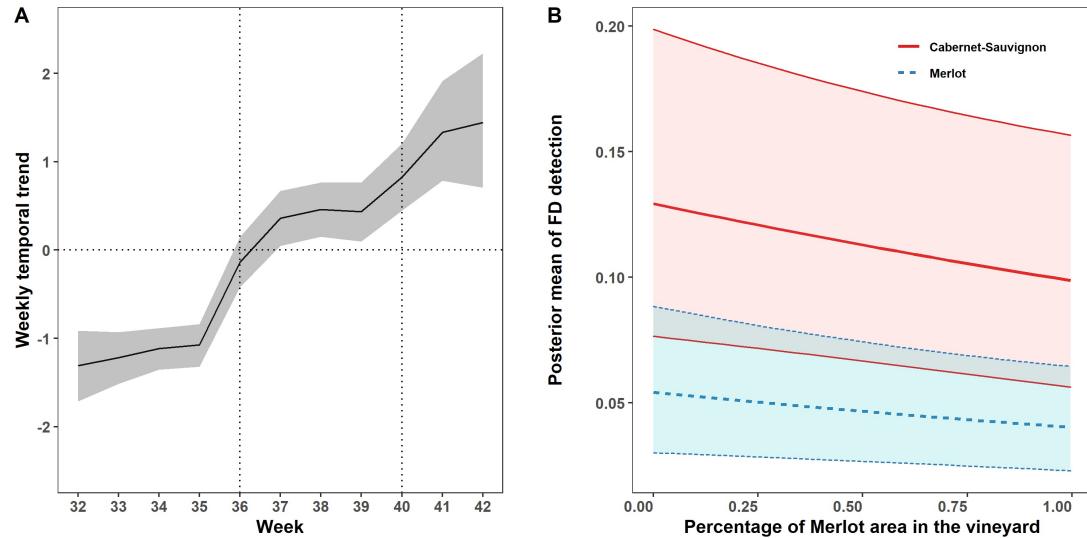


Figure 5: Effects of the week of field inspection and of the cultivar composition of the landscape on FD detection, for a landscape scale of 150 m. A: Effects of the week of field inspected on FD detection as estimated with model  $M_1$  at 150m. In model  $M_1$ , the fixed effect variable *Season* is replaced by a random walk of order 1 indexed by weeks. The mean standardized temporal effect (from week 32, second week of august, to week 42, third week of october) is displayed along with its 95% credible band. B: Effect of the proportion of Merlot in the landscape on the probability of FD detection, using the model  $M_0$  for the two most widespread cultivars (Merlot and Cabernet Sauvignon) in vineyard plots under conventional practices (mean probability and 95% credible bands). The effects were estimated by fixing all other numerical variables at their mean values, the variable AOC to Bordeaux, the variable Practice to conventional and considering a sampling period during the autumn.

425 having a prevalence of FD detection exceeding 10% of the fields. The posterior mean precision  
 426 (i.e. positive predictive value) is then 50%. This is the probability that prevalence of FD  
 427 detection exceeds 10% in a district when the district was classified as such by the model.

## 428 Discussion

429 In this study, we used species distribution models fitted in a Bayesian framework to identify  
 430 key biotic and abiotic factors driving the detection and epidemiology of Flavescence dorée, a  
 431 major quarantine disease damaging European vineyards. To this end and based on our eco-  
 432 logical knowledge of the disease, we integrated data from different sources (e.g., FD survey  
 433 data, agricultural management and environmental explanatory variables) to build a quantita-  
 434 tive understanding of FD infection. Our study makes use of a large spatial dataset gathering  
 435 34581 observations of the infection status of vineyard plots. It relies on an accurate moni-  
 436 toring of symptoms realized by trained inspectors completed with molecular detection of FD  
 437 phytoplasma on symptomatic plants performed by accredited laboratories. Due to the costs  
 438 of laboratory testing, no tests are conducted on leaves from non-symptomatic plants, opening  
 439 possibly the way to false negative results due to cryptic infection ((Parnell et al. 2017)). To this  
 440 respect, our analysis highlights the importance of sampling period on FD detection (variable  
 441 *Season* in model  $M_0$  and *Week* in model  $M_1$ ). Both models indicate that the probability of  
 442 FD detection substantially increases after week 36 (first week of September, Figs. 4A, 5A). The  
 443 visual identification of FD symptoms is then much easier as this period is the ideal time for the  
 444 expression of certain specific symptoms, such as grape shrivelling and the non-lignification of  
 445 canes (Tramontini et al. 2020). Accordingly, from September onwards, the risk of false negative  
 446 results is low. Furthermore, recent studies show that non-symptomatic plants sampled at the end

of the summer in fields with high FD prevalence were actually not infected by the phytoplasma (Eveillard et al. 2016, Ripamonti et al. 2020). In particular, the study of (Eveillard et al. 2016) was conducted in the Bordeaux region over hundreds of Cabernet-Sauvignon and Merlot plants, the two most widespread cultivars in our dataset (76 % of the samples). Moreover, inspectors collected in each field a wide diversity of symptomatic leaves possibly related to phytoplasmas. Among the 4554 molecular detection performed in as many fields, 47% were positive to FD. Accordingly, based on these premises, we are pretty confident that our analysis provide important information on how local farming practices and landscape context affect the probability of FD infection. At local scale, our results reveal in particular that the probability of detecting infected fields increases with field area, the age of the plantation, and with specific cultivars (i.e., Cabernet Sauvignon, Cabernet Franc and Muscadelle). At the landscape scale, the probability of FD infection increases with the proportion of susceptible cultivars as well as with the proportion of forest, urban land and organic farming but decreases with the proportion of vineyard.

At field scale, cultivar choice, but not other viticultural practices considered, affected the probability of field infection. Cultivar choice was a major determinant of FD infection. No major resistance gene effective against FD is currently available, but substantial differences in cultivar susceptibility have been reported (Oliveira et al. 2019, Eveillard et al. 2016). In the GDON des Bordeaux region, the most frequently grown cultivars are Merlot (in 61% of the fields analysed), Cabernet Sauvignon (in 15% of the fields), Cabernet Franc and Semillon (in 7% of the fields each). Previous studies involving controlled inoculations in greenhouses or monitoring in experimental plots have shown that Cabernet Sauvignon and Cabernet Franc are more susceptible to FD than Merlot, displaying a higher incidence of the disease, more severe symptoms and higher rates of phytoplasma multiplication (Eveillard et al. 2016). Our results therefore confirmed this ranking in production situations for the first time. They also suggest that Semillon has low susceptibility to FD, whereas Muscadelle is highly susceptible to FD. In addition, our analyses revealed that the altitude substantially affects FD infection : the higher the altitude of the field, the lower its probability of infection (Fig. 4B). This effect fits the empirical observation that, in the region studied, FD is more likely to occur in wetland areas, which are frequently found at lower altitudes, typically close to rivers. The slightly lower goodness-of-fit obtained by replacing altitude by an wetland index (Merot et al. 2003) have led us to favour this variable which is, also, easier to access. Identifying the underlying biological factors deserve further studies.

Our study reveals that field infection by FD were also affected by the landscape context within the 150 – 200 m surrounding focal fields. The best fit were obtained with the landscape scale of 150 m, followed by 200 m. These landscape extent corresponds to areas ranging from 7 ha (150 m scale) to 12.5 ha (200 m scale) and contains, in average, 6-10 fields in the study area (the mean field area is 0.76 ha). This short range may result from the low dispersal capacity of the vector of the FD phytoplasma. Indeed, 80% of *S. titanus* adults disperse within 30 m of their source, although long-range dispersal over distances of up to 330 m has occasionally been observed (Lesson et al. 2014). Several effects of landscape composition were identified (Fig. 4B). Increases in the proportions of two generic land-use classes, forest and urban land, increased the probability of FD infection. The interpretation of such effects is not straightforward because generic land-use classes provide only an imperfect description of the presence of host habitats involved in epidemiological dynamics (Vanwambeke et al. 2019). In our case study, it is tempting to consider the positive effect of the proportion of forest as favoring wild reservoirs of FD, such as wild alders and *Clematis sp.* However, recent studies have demonstrated that the transfer of the phytoplasma from these plants to grapevine is rare, so their impact as chronic reservoirs is likely to be minimal (Filippin et al. 2009, Malembic-Maher et al. 2020). The hypothesis involving vines, whether abandoned or cultivated, is more likely. Indeed, abandoned *Vitis sp.* plants in forest margins and outskirts constitute an important source of FD inoculum but also of insect vectors (Lesson et al. 2014, Ripamonti et al. 2020). To this respect, it could be interesting to refine landscape description in order to investigate the role of the size of forest patches as smaller patches offer more surface on their outskirts compared to larger ones. Similarly, inventories and cartography of *Vitis sp.* in two districts of Bordeaux area have shown that cultivated vines in individual gardens or uncultivated *Vitis sp.* in wasteland are frequent

502 in urbanized areas neighboring vineyards (data not published). They could also constitute an  
 503 important source of FD inoculum and insect vectors, especially as residents and gardeners have  
 504 little awareness of the problem of FD. Alternatively, it is also likely that fewer insecticide treat-  
 505 ments are performed by farmers in vineyards close to private houses or public facilities but also  
 506 by amateur gardeners, which might also explain this result.

507 In addition, we identify that the probability of FD infection decreases with the proportion of  
 508 vineyards in the landscape but also with planting density. This effect, which appears counter-  
 509 intuitive at first glance, may come from several processes. It may reflect a direct dilution effect  
 510 of the available inoculum or of the vector population within larger cultivated vineyards. In  
 511 line with this hypothesis, (Delaune et al. 2021) have recently shown that univoltine pest species  
 512 such as *S. titanus* are negatively correlated with the host crop area in the landscape during the  
 513 ongoing growing season. This effect may also reflect that larger cultivated vineyards can benefit  
 514 from a more homogeneous insecticide protection (Meehan et al. 2011). Alternatively, it might  
 515 be an indirect effect mediated by the economic value of vineyards. Indeed, areas with highly  
 516 valued vineyards are generally more specialised, with more financial incentives for farmers to  
 517 control the infection (Ay and Gozlan 2020).

518 Interestingly, effects of cultivars detected at the field scale were also found at the land-  
 519 scape scale suggesting up-scaling of these effects that may be mediated by dispersal abili-  
 520 ties of *S. titanus* populations. We show that the probability of FD infection decreases with  
 521 increasing proportions of Merlot (Fig. 5B). Several studies have shown that the landscape  
 522 composition in plant species (*i.e.* between-species diversity) affect plant disease epidemiology  
 523 (Plantegenest et al. 2007, Meentemeyer et al. 2012). However, studies on the effect of varietal  
 524 (*i.e.* within-species diversity) composition at the landscape scale are much rarer and mainly  
 525 based on theoretical analyses (Rimbaud et al. 2021). Few empirical studies have suggested such  
 526 an effect and they mainly considered regional or national scales (Priestley and Bayles 1980,  
 527 Finckh and Wolfe. 2017, Papaïx et al. 2011)). By contrast, our study suggests an effect of the  
 528 varietal composition of the landscape at an intermediate spatial scale, typically over an area of  
 529 7 – 12 ha, containing about 6-10 fields. Finally, while no effect of farming practices (organic  
 530 *versus* conventional) was detected at the field scale, our result suggest that the probability of  
 531 FD infection may slightly increase with the proportion of organic fields in the landscape (Fig.  
 532 4B). Note first that this variable indicates if organic practices are used at farm scale rather  
 533 than for each individuals fields (fields being attached to their farm in our study). It could  
 534 possibly explain this mismatch between the results at local and landscape scales. The positive  
 535 effect at landscape scale could suggest a lower efficacy of vector control in organic systems. The  
 536 natural products used in organic vineyards to control *S. titanus* are as effective as synthetic  
 537 insecticides in controlling nymphs but less effective in controlling adults, the most dispersive  
 538 instar (Tacoli et al. 2017, Guesberti et al. 2008, Chuche and Thiéry 2014). Combined with the  
 539 stricter conditions governing field applications, it might account for this effect. However, this  
 540 hypothesis needs further investigations as we did not have the information about the number of  
 541 mandatory treatments actually realized. Moreover, other practices used in conventional and or-  
 542 ganic viticulture (*e.g.* soil tillage, cover crops or canopy management) that were not considered  
 543 here might have affected the spread of FD and the dynamics of its vector (Muneret et al. 2018).

544 The local and landscape factors identified can be used to improve vineyard surveys, thereby  
 545 decreasing the economic and environmental impacts of managing this disease. The use of the  
 546 model as a predictive tool is potentially interesting for the targeting of new districts in which  
 547 future surveys are most likely to be valuable, as suggested by its to the ability of the model to  
 548 discriminate districts with prevalence of FD detection exceeding 10% of the fields (Fig. S6B).  
 549 Out-of-sample predicted probabilities of FD presence could be used in a benefit-cost framework  
 550 in order to determine the optimal size of future surveys or to improve the current control policy  
 551 (Ay and Gozlan 2020). The grubbing costs associated to FD presence would be derived from  
 552 farm's gross margins or vineyard prices and compared to the monitoring costs from professional  
 553 organisations such as GDON. Moreover, the Bayesian method used (INLA) allows uncertainty  
 554 to be quantified and visualised in risk map outputs. This is important for fostering dialogue  
 555 with stakeholders and policy-makers. However, and importantly, using the model as a predic-  
 556 tive tool must firstly be restricted to make interpolation to unsampled sites within the GDON

des Bordeaux area and not as a tool able to make prediction in remote vineyards. Moreover, given the importance of the spatial effects (as discussed below), the model has also to be fitted again annually by taking into account new FD survey in order to update the spatial field to the current state of the epidemic in the area. A short-term operational output of the result to improve FD monitoring concerns the effect of the sampling period. It suggests that the harvest period in Bordeaux (September) is a more favourable period for maximising the efficacy of vineyard surveys (Fig. 4A). Note that part of the size of this effect could also be explained by the GDON internal organization of field surveys as the last weeks of survey, mostly after week 38 (*i.e.* over 2.8% of the sample), are realized by more experienced teams of inspectors while sometimes targeting districts closer to previously identified infected fields. This result strongly suggests that equipping grapevine harvesting machines with cameras and automatic symptom recognition tools could greatly improve FD detection, especially since harvesting machines travel through a large part of the vineyard every year. Recently, the use of convolutional neural networks has shown promise for detecting FD symptoms from in-field images (Boulent et al. 2020). Finally, our results may have important consequences for the management of vineyard landscapes echoing the land sparing / land sharing debate (Phalan et al. 2011). Our study suggests that promoting landscapes with large amount of vineyards and lower proportions of forests limit FD infections while maintaining semi-natural habitats in such landscapes are known to be of major importance for biodiversity conservation or mitigation of climate change effects (Rusch et al. 2022, Barbaro et al. 2021). Policy makers and land use planners may therefore include such issue if they are to design multifunctional vineyard landscapes.

In addition to the effects of the fixed factors, a large proportion of the biological process of interest was captured by the spatial effects included in our models. In our case study of more than 35000 observations, the INLA method proved effective for implementing complex Bayesian hierarchical models, taking spatial autocorrelation into account. The spatial component, as revealed by the Matérn function (Fig. 2D), had a strong effect, suggesting that some biotic (*e.g.* specific cultural practices) and abiotic (*e.g.* climatic variables) factors are missing or misspecified (*e.g.* linear terms where nonlinear terms are required) (Elith and Leathwick. 2009). However, the strong spatial pattern observed may also be the result of several epidemics developing simultaneously at local level, over small scales, consistent with the short dispersal distances of the vector, whereas the distribution model is based primarily on a static vision of the disease risk factor in an environment at equilibrium (Elith and Leathwick. 2009). This assumption does not hold for invading pathogens (Purse and Golding 2015). In recent years, the idea of coupling niche modelling with spatially explicit models of disease dispersal to provide better information about potential disease spread and ultimately improve epidemiological surveillance strategies has been explored by several authors (Purse and Golding 2015, Parnell et al. 2017, Martinetti and Soubeyrand 2019, Rimbaud et al. 2018, Cunniffe et al. 2016, Hyatt-Twynam et al. 2017, Adrakey et al. 2017). This approach merits implementation in this context, to improve prediction accuracy at field scale. This implementation could make use of existing models of FD epidemics at field scale (Lesson et al. 2015). The insights gained from our correlative study could be incorporated into mechanistic models of epidemic spread (Meentemeyer et al. 2011, Hartemink et al. 2011). Moreover, it should also be recognised that correlative models, which can easily take into account large numbers of predictors, sometimes outperform mechanistic models and can have a major impact on policy decisions (Leach and Scoones. 2013). From this point of view, it may also be useful to explore the rapidly developing research fields of machine learning and data mining. These methods have been successfully applied to the surveillance of *X. fastidiosa* (Martinetti and Soubeyrand 2019). This possibility is particularly attractive, given that machine learning methods have recently been adapted to take spatial autocorrelation into account (Georganos et al. 2021).

606

## 607 Authors' contributions

608 HA, SM and FF conceived the ideas and designed methodology; HA, SM, LR and FF col-  
609 lected the data; HA, LR, AR, JSA and FF analysed the data; HA and FF led the writing

610 of the manuscript. All authors contributed critically to the drafts and gave final approval for  
611 publication.

## 612 Acknowledgments

613 Authors thank people who sampled or provide data: Sophie Bentejac, Morgane Le Goff and  
614 Charlotte Labit (GDON des Bordeaux), Dominique Vergnes (FREDON Aquitaine), Gontran  
615 Casella (Syndicat des Bordeaux et Bordeaux Supérieur) and Soungalo Dembele (Master student  
616 involved in early work on GIS data base). We also thanks Sophie Bentejac, Charlotte Labit and  
617 Thomas Opitz (INRAE) for their critical and careful reading of the manuscript. We thank the  
618 team at INRAE in charge of the Migale server used to perform calculations. We also thanks  
619 Ruairi Donnelly for reviewing an early draft. This study receives financial support under research  
620 contracts Co-Act and RISCA (Plan National Dépérissage du Vignoble).

## 621 References

- 622 (Adrakey et al. 2017) Adrakey, H., Streftaris, G., Cunniffe, N., Gottwald, T., Gilligan, C., and  
623 Gibson, G. 2017. Evidence-based controls for epidemics using spatio-temporal stochastic  
624 models in a bayesian framework. *Journal of Royal Society Interface*. 14: 20170386.
- 625 (Ay and Gozlan 2020) Ay, J.-S., and Gozlan, E. 2020. Disease dispersion as a spatial interaction:  
626 The case of Flavescence Dorée. *Natural Resource Modeling*. 33: e12265.
- 627 (Barbaro et al. 2021) Barbaro, L., Assandri, G., Brambilla, M., Castagneyrol, B., Froidevaux,  
628 J., Giffard, B., Python, J., Puig- Montserrat, X., Torre, I., Calatayud, F., Gaüzère, P.,  
629 Guenser, J., Macia'-Valverde, F.-X., Mary, S., Raison, L., Sirami, C., and Rusch, A. 2021.  
630 Organic management and landscape heterogeneity combine to sustain multifunctional bird  
631 communities in European vineyards. *Journal of Applied Ecology*. 58(6): 1261–1271.
- 632 (Bebber 2015) Bebber, D. P. 2015. Range-Expanding Pests and Pathogens in a Warming World.  
633 *Annual Review of Phytopathology*. 53(1): 335–356.
- 634 (Beguin et al. 2012) Beguin, J., Martino, S., Rue, H., and Cumming, S. G. 2012. Hierarchi-  
635 cal analysis of spatially autocorre- lated ecological data using integrated nested laplace  
636 approximation. *Methods in Ecology and Evolution*. 3: 921–929.
- 637 (Boulent et al. 2020) Boulent, J., St-Charles, P.-L., Foucher, S., and Théau, J. 2020. Automatic  
638 Detection of Flavescence Dorée Symptoms Across White Grapevine Varieties Using Deep  
639 Learning. *Frontiers in Artificial Intelligence*. 3: 564878.
- 640 (Bressan et al. 2005) Bressan, A., Spiazzi, S., Girolami, V., and Boudon-Padieu, E. 2005. Ac-  
641 quisition efficiency of Flavescence dorée phytoplasma by *Scaphoideus titanus* Ball from in-  
642 fected tolerant or susceptible grapevine cultivars or experimental host plants. *Vitis*. 44(3):  
643 143–146.
- 644 (Caudwell 1957) Caudwell, A. 1957. Deux années d'études sur la flavescence dorée nouvelle  
645 maladie grave de la vigne. *Annales de l'Amelioration des Plantes*. 4: 359–393.
- 646 (Cendoya et al. 2020) Cendoya, M., Martinez-Minaya, J., Dalmau, V., Ferrer, A., Saponari, M.,  
647 Conesa, D., Lopez-Quilez, A., and Vicent, A. 2020. Spatial Bayesian Modeling Applied to  
648 the Surveys of *Xylella fastidiosa* in Alicante (Spain) and Apulia (Italy). *Frontiers in Plant  
649 Science*. 11: 1204.
- 650 (Chuche and Thiéry 2014) Chuche, J., and Thiéry, D. 2014. Biology and ecology of the Flaves-  
651 cence dorée vector *Scaphoideus titanus*: A review. *Agronomy for Sustainable Development*.  
652 34(2): 381–403.
- 653 (Cunniffe et al. 2016) Cunniffe, N. J., Cobb, R. C., Meentemeyer, R. K., Rizzo, D. M., and  
654 Gilligan, C. A. 2016. Modeling when, where, and how to manage a forest epidemic, moti-  
655 vated by sudden oak death in California. *Proceedings of the National Academy of Sciences  
656 of the United States of America*. 113(20): 5640–5645.
- 657 (Delaune et al. 2021) Delaune, T., Ouattara, M. S., Ballot, R., Sausse, C., Felix, I., Maupas,  
658 F., Chen, M., Morison, M., Makowski, D., and Barbu, C. 2021. Landscape drivers of pests  
659 and pathogens abundance in arable crops. *Ecography*. 44(10): 1429–1442.
- 660 (Desneux et al. 2007) Desneux, N., Decourtey, A., and Delpuech, J.-M. 2007. The sublethal  
661 effects of pesticides on beneficial arthropods. *Annual Review of Entomology*. 52: 81–106.
- 662 (Dormann et al. 2007) Dormann, C. F., McPherson, J. M., Araùjo, M. B., Bivand, R., Bolliger,  
663 J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R.,  
664 Peres-Neto, P. , Reineking, B., Schröder, B., Schurr, F. M., and Wilson, R. 2007. Methods  
665 to account for spatial autocorrelation in the analysis of species distributional data: A  
666 review. *Ecography*. 30(5): 609–628.

- 667 (Dormann et al. 2021) Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C.,  
 668 Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B., and Singer, A. 2012.  
 669 Correlation and process in species distribution models: Bridging a dichotomy. *Journal of  
 670 Biogeography*. 39(12): 2119–2131.
- 671 (Elith and Leathwick. 2009) Elith, J., and Leathwick, J. R. 2009. Species Distribution Models:  
 672 Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology,  
 673 Evolution, and Systematics*. 40(1): 677–697.
- 674 (Eveillard et al. 2016) Eveillard, S., Jollard, C., Labroussaa, F., Khalil, D., Perrin, M., Desqué,  
 675 D., Salar, P., Razan, F., Hévin, C., Bordenave, L., Foissac, X., Masson, J. E., and  
 676 Malembic-Maher, S. 2016. Contrasting Susceptibilities to Flavescence Dorée in *Vitis  
 677 vinifera*, Rootstocks and Wild *Vitis* Species. *Frontiers in Plant Science*. 7: 1762.
- 678 (Filippin et al. 2009) Filippin, L., Jovic, J., Cvrkovic, T., Forte, V., Clair, D., Tosevski, I.,  
 679 Boudon-Padieu, E., Borgo, M., and Angelini, E. 2009. Molecular characteristics of phy-  
 680 toplasmas associated with Flavescence dorée in clematis and grapevine and preliminary  
 681 results on the role of *Dictyophara europaea* as a vector. *Plant Pathology*. 58(5): 826–837.
- 682 (Finckh and Wolfe. 2017) Finckh, M. R., and Wolfe, M. S. 2017. Chapter 4.4: Biodiversity En-  
 683 hancement. In: *Plant Diseases and Their Management in Organic Agriculture*, eds Maria  
 684 R. Finckh, Ariena H. C. van Bruggen, and Lucius Tamm. The American Phytopathological  
 685 Society, p. 153–174.
- 686 (Fuglstad et al. 2019) Fuglstad, G. A., Simpson, D., Lindgren, F., and Rue, H. 2019. Con-  
 687 structing priors that penalize the complexity of gaussian random fields. *Journal of the  
 688 American Statistical Association*. 114: 445–452.
- 689 (Galetto et al. 2014) Galetto, L., Miliordos, D., Roggia, C., Rashidi, M., Sacco, D., Marzachi,  
 690 C., and Bosco, D. 2014. Acquisition capability of the grapevine Flavescence dorée by the  
 691 leafhopper vector *Scaphoideus titanus* Ball correlates with phytoplasma titre in the source  
 692 plant. *Journal of Pest Science*. 87(4): 671–679.
- 693 (Gelman et al. 2014) Gelman, A., Hwang, J., and Vehtari, A. 2014. Understanding predictive  
 694 information criteria for bayesian models. *Statistics and Computing*. 24: 997–1016.
- 695 (Georganos et al. 2021) Georganos, S., Grippa, T., Gadiaga, A. N., Linard, C., Lennert, M.,  
 696 Vanhuyse, S., Mboga, N., Wolff, E., and Kalogirou, S. 2021. Geographical random forests:  
 697 A spatial extension of the random forest algorithm to address spatial heterogeneity in  
 698 remote sensing and population modelling. *Geocarto International*. 36(2): 121-136.
- 699 (Godefroid et al. 2019) Godefroid, M., Cruaud, A., Streito, J.-C., Rasplus, J.-Y., and Rossi, J.-  
 700 P. 2019. *Xylella fastidiosa*: Climate suitability of European continent. *Scientific Reports*.  
 701 9(1): 8844.
- 702 (Gusberti et al. 2008) Gusberti, M., Jermini, M., Wyss, E., and Lindner, C. 2008. Efficacité  
 703 d'insecticides contre *Scaphoideus titanus* en vignobles biologiques et effets secondaires.  
 704 Revue suisse de viticulture, arboriculture, horticulture. 40(3): 173–177.
- 705 (Hartemink et al. 2011) Hartemink, N., Vanwambeke, S. O., Heesterbeek, H., Rogers, D., Mor-  
 706 ley, D., Pesson, B., Davies, C., Mahamdallie, S., and Ready, P. 2011. Integrated Mapping  
 707 of Establishment Risk for Emerging Vector-Borne Infections: A Case Study of Canine  
 708 Leishmaniasis in Southwest France. *PLOS One*. 6(8): e20817.
- 709 (Hesselbarth et al. 2019) Hesselbarth, M. H., Sciajini, M., With, K. A., Wiegand, K., and  
 710 Nowosad, J. 2019. Landscapemetrics: An open-source r tool to calculate landscape metrics.  
 711 *Ecography*. 42: 1648–1657.
- 712 (Hyatt-Twynam et al. 2017) Hyatt-Twynam, S. R., Parnell, S., Stutt, R. O. J. H., Gottwald,  
 713 T. R., Gilligan, C. A., and Cunniffe, N. J. 2017. Risk-based management of invading plant  
 714 disease. *New Phytologist*. 214(3): 1317–1329.

- 715 (Inglada et al. 2017) Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I.  
 716        2017. Operational High Resolution Land Cover Map Production at the Country Scale Using  
 717        Satellite Image Time Series. *Remote Sensing*. 9(1): 95.
- 718 (Jeger et al. 2016) Jeger, M., Bragard, C., Caffier, D., Candresse, T., Chatzivassiliou, E.,  
 719        Dehnen-Schmutz, K., Giloli, G., Miret, J. A. J. et al. 2016. Risk to plant health of Flaves-  
 720        cence dorée for the EU territory. *EFSA Journal*. 14(12): e04603.
- 721 (Kraemer et al. 2016) Kraemer, M. U. G., Hay, S. I., Pigott, D. M., Smith, D. L., Wint, G.  
 722        R. W., and Golding, N. 2016. Progress and Challenges in Infectious Disease Cartography.  
 723        *Trends in Parasitology*. 32(1): 19–29.
- 724 (Krainski et al. 2018) Krainski, E., Gomez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D.,  
 725        Simpson, D., Lindgren, F., and Rue, H. 2018. *Advanced Spatial Modeling with Stochastic  
 726        Partial Differential Equations Using R and INLA*. Chapman and Hall/CRC.
- 727 (Leach and Scoones. 2013) Leach, M., and Scoones, I. 2013. The social and political lives of  
 728        zoonotic disease models: Narratives, science and policy. *Social Science and Medicine*. 88:  
 729        10–17.
- 730 (Lessio and Alma. 2004) Lessio, F., and Alma, A. 2004. Dispersal patterns and chromatic re-  
 731        sponse of *Scaphoideus titanus* Ball (Homoptera Cicadellidae), vector of the phytoplasma  
 732        agent of grapevine flavescence dorée. *Agricultural and Forest Entomology*. 6(2): 121–128.
- 733 (Lesson et al. 2015) Lessio, F., Portaluri, A., Paparella, F., and Alma, A. 2015. A mathematical  
 734        model of flavescence dorée epidemiology. *Ecological Modelling*, 312, 41–53.
- 735 (Lesson et al. 2014) Lessio, F., Tota, F., and Alma, A. 2014. Tracking the dispersion of  
 736        *Scaphoideus titanus* Ball (Hemiptera: Cicadellidae) from wild to cultivated grapevine: Use  
 737        of a novel mark-capture technique. *Bulletin of Entomological Research*. 104(4): 432–443.
- 738 (Lindgren et al. 2011) Lindgren, F., Rue, H., and Lindstrom, J. 2011. An explicit link between  
 739        gaussian fields and gaussian markov random fields: The spde approach (with discussion).  
 740        *Journal of the Royal Statistical Society, Series B*. 73(4): 423–398.
- 741 (Lindgren and Rue 2015) Lindgren, F., and Rue, H. 2015. Bayesian Spatial Modelling with  
 742        R-INLA. *Journal of Statistical Software*. 63: 1–25.
- 743 (Malembic-Maher et al. 2020) Malembic-Maher, S., Desqué, D., Khalil, D., Salar, P., Bergey,  
 744        B., Danet, J.-L., Duret, S., et al. 2020. When a Palearctic bacterium meets a Nearctic insect  
 745        vector: Genetic and ecological insights into the emergence of the grapevine Flavescence  
 746        dorée epidemics in Europe. *PLOS Pathogens*. 16(3): e1007967.
- 747 (Martin et al. 2005) Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A.,  
 748        Low-Choy, S. J., Tyre, A. J., and Possingham, H. P. 2005. Zero tolerance ecology: Improv-  
 749        ing ecological inference by modelling the source of zero observations. *Ecology Letters*. 8:  
 750        1235–1246.
- 751 (Martinetti and Soubeyrand 2019) Martinetti, D., and Soubeyrand, S. 2019. Identifying Look-  
 752        outs for Epidemio-Surveillance: Application to the Emergence of *Xylella fastidiosa* in  
 753        France. *Phytopathology*, PHYTO-07-18-023.
- 754 (Martinez-Minaya et al. 2018) Martinez-Minaya, J., Conesa, D., Lopez-Quilez, A., and Vicent,  
 755        A. 2018. Spatial and climatic factors associated with the geographical distribution of citrus  
 756        black spot disease in South Africa. A Bayesian latent Gaussian model approach. *European  
 757        Journal of Plant Pathology*. 151(4): 991– 1007.
- 758 (Meehan et al. 2011) Meehan, T., Werling, B., Landis, D. and Gratton, C. 2011. Agricultural  
 759        Landscape Simplification and Insecticide Use in the Midwestern United States. *Proceedings  
 760        Of The National Academy Of Sciences*. 108: 11500-11505.

- 761 (Meentemeyer et al. 2008) Meentemeyer, R. K., Anacker, B. L., Mark, W., and Rizzo, D. M.  
 762 2008. Early Detection of Emerging Forest Disease Using Dispersal Estimation and Ecolog-  
 763 ical Niche Modeling. *Ecological Applications*. 18(2): 377–390.
- 764 (Meentemeyer et al. 2011) Meentemeyer, R. K., Cunniffe, N. J., Cook, A. R., Filipe, J. A. N.,  
 765 Hunter, R. D., Rizzo, D. M., and Gilligan, C. A. 2011. Epidemiological modelling of inv-  
 766 asion in heterogeneous landscapes: Spread of sudden oak death in California (1990–2030).  
 767 *Ecosphere*. 2(2): 1–24.
- 768 (Meentemeyer et al. 2012) Meentemeyer, R. K., Haas, S. E., and Väclavík, T. 2012. Landscape  
 769 Epidemiology of Emerging Infectious Diseases in Natural and Human-Altered Ecosystems.  
 770 *Annual Review of Phytopathology*. 50(1): 379–402.
- 771 (Merot et al. 2003) Merot, P., Squividant, H., Aurousseau, P., Hefting, M., Burt, T., Maitre, V.,  
 772 Kruk, M., Butturini, A., Thenail, C., and Viaud, V. 2003. Testing a climato-topographic  
 773 index for predicting wetlands distribution along an European climate gradient. *Ecological  
 774 Modelling*, 163(1), 51–71.
- 775 (Morone et al. 2007) Morone, C., Boveri, M., Giosue, S., Gotta, P., Rossi, V., Scapin, I., and Marzachi, C.  
 776 2007. Epidemiology of Flavescence dorée in Vineyards in Northwestern Italy. *Phytopathol-  
 777 ogy*, 97(11), 1422–1427.
- 778 (Muneret et al. 2018) Muneret, L., Mitchell, M., Seufert, V., Aviron, S., Djoudi, E. A., Pétill-  
 779 lon, J., Plantegenest, M., Thiéry, D., and Rusch, A. 2018. Evidence that organic farming  
 780 promotes pest control. *Nature Sustainability*. 1(7): 361–368.
- 781 (Namba 2019) Namba, S. 2019. Molecular and biological properties of phytoplasmas. *Proceed-  
 782 ings of the Japan Academy. Series B, Physical and Biological Sciences*. 95(7): 401–418.
- 783 (Oliveira et al. 2019) Oliveira, M. J. R., Roriz, M., Vasconcelos, M. W., Bertaccini, A., and  
 784 Carvalho, S. M. P. 2019. Conven- tional and novel approaches for managing “flavescence  
 785 dorée” in grapevine: Knowledge gaps and future prospects. *Plant Pathology*. 68(1): 3–17.
- 786 (Papaïx et al. 2011) Papaïx, J., Goyeau, H., Du Cheyron, P., Monod, H., and Lannou, C. 2011.  
 787 Influence of cultivated landscape composition on variety resistance: An assessment based  
 788 on wheat leaf rust epidemics. *New Phytologist*. 191(4): 1095–1107.
- 789 (Papura et al. 2012) Papura, D., Burban, C., van Helden, M., Giresse, X., Nusillard, B., Guille-  
 790 maud, T., and Kerdelhué, C. 2012. Microsatellite and Mitochondrial Data Provide Evidence  
 791 for a Single Major Introduction for the Nearctic Leafhopper *Scaphoideus titanus* in Europe.  
 792 *PLOS One*. 7(5):
- 793 (Parnell et al. 2014) Parnell, S., Gottwald, T. R., Riley, T., and van den Bosch, F. 2014. A  
 794 generic risk-based surveying method for invading plant pathogens. *Ecological Applications*.  
 795 24(4): 779–790.
- 796 (Parnell et al. 2017) Parnell, S., van den Bosch, F., Gottwald, T. R., and Gilligan, C. A. 2017.  
 797 Surveillance to Inform Control of Emerging Plant Diseases: An Epidemiological Perspec-  
 798 tive. *Annual Review of Phytopathology*. 55(1): 591–610.
- 799 (Pelletier et al. 2009) Pelletier, C., Salar, P., Gillet, J., Cloquemin, G., Very, P., Foissac, X., and  
 800 Malembic-Maher, S. 2009. Triplex real-time PCR assay for sensitive and simultaneous de-  
 801 tection of grapevine phytoplasmas of the 16SrV and 16SrXII-A groups with an endogenous  
 802 analytical control. *Vitis*. 48(2): 87–95.
- 803 (Phalan et al. 2011) Phalan, B., Onial, M., Balmford, A., and Green, R. E. 2011. Reconciling  
 804 Food Production and Biodiversity Conservation: Land Sharing and Land Sparing Com-  
 805 pared. *Science*. 333(6047): 1289–1291.
- 806 (Plantegenest et al. 2007) Plantegenest, M., Le May, C., and Fabre, F. 2007. Landscape epi-  
 807 demiology of plant diseases. *Journal of the Royal Society Interface*. 4: 963–972.

- 808 (Priestley and Bayles 1980) Priestley, R. H., and Bayles, R. A. 1980. Varietal diversification as  
809 a means of reducing the spread of cereal diseases in the united kingdom. *Journal of the*  
810 *National Institute of Agricultural Botany.* 15: 205–214.
- 811 (Purse and Golding 2015) Purse, B. V., and Golding, N. 2015. Tracking the distribution and  
812 impacts of diseases with biological records and distribution modelling. *Biological Journal of the Linnean Society.* 115(3): 664–677.
- 813 (Quaglino et al. 2013) Quaglino, F., Zhao, Y., Casati, P., Bulgari, D., Bianco, P. A., Wei, W.,  
814 and Davis, R. E. 2013. ‘Candidatus Phytoplasma solani’, a novel taxon associated with  
815 stolbur- and bois noir-related diseases of plants. *International Journal of Systematic and*  
816 *Evolutionary Microbiology.* 63: 2879–2894.
- 817 (Rimbaud et al. 2021) Rimbaud, L., Fabre, F., Papaix, J., Moury, B., Lannou, C., Barrett, L.  
818 G., and Thrall, P. H. 2021. Models of Plant Resistance Deployment. *Annual Review of*  
819 *Phytopathology.* 59(1): 125–152.
- 820 (Rimbaud et al. 2018) Rimbaud, L. [Loup], Bruchou, C., Dallot, S., Pleydell, D. R. J., Jacquot,  
821 E., Soubeyrand, S., and Thébaud, G. 2018. Using sensitivity analysis to identify key factors  
822 for the propagation of a plant epidemic. *Royal Society Open Science.* 5(1): 171435.
- 823 (Ripamonti et al. 2021) Ripamonti, M., Pegoraro, M., Morabito, C., Gribaudo, I., Schubert,  
824 A., Bosco, D., and Marzachi, C. 2021. Susceptibility to flavescence dorée of different *vitis*  
825 *vinifera* genotypes from north-western italy. *Plant Pathology.* 70(3): 511–520.
- 826 (Ripamonti et al. 2020) Ripamonti, M., Pegoraro, M., Rossi, M., Bodino, N., Beal, D., Panero,  
827 L., Marzach'i, C., and Bosco, D. 2020. Prevalence of Flavescence Dorée Phytoplasma-  
828 Infected *Scaphoideus titanus* in Different Vineyard Agroecosystems of Northwestern Italy.  
829 *Insects.* 11(5): 301.
- 830 (Rue et al. 2009) Rue, H., Martino, S., and Chopin, N. 2009. Approximate bayesian inference  
831 for latent gaussian models using integrated nested laplace approximations (with discussion).  
832 *Journal of the Royal Statistical Society, Series B.* 71(2): 319–392.
- 833 (Rusch et al. 2022) Rusch, A., Beaumelle, L., Giffard, B., and Alonso Ugaglia, A. 2022. Chapter  
834 Seven - Harnessing biodiversity and ecosystem services to safeguard multifunctional  
835 vineyard landscapes in a global change context. In , *Advances in Ecological Research*, eds.  
836 D. A. Bohan, A. J. Dumbrell, and A. J. Vanbergen . Academic Press, p. 305–335.
- 837 (Saito and Rehmsmeier 2015) Saito, T., and Rehmsmeier, M. 2015. The Precision-Recall Plot  
838 Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced  
839 Datasets. *PLOS ONE,* 10(3): e0118432.
- 840 (Schvester et al. 1969) Schvester, D., Carle, P., and Moutous, G. 1969. Nouvelles données sur  
841 la transmission de la flavescence dorée de la vigne par *Scaphoideus littoralis* Ball. *Annales*  
842 *de Zoologie Ecologie Animale.* 1(4): 445–465.
- 843 (Simpson et al. 2017) Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. 2017.  
844 Penalising Model Component Complexity: A Principled, Practical Approach to Construct-  
845 ing Priors. *Statistical Science.* 32(1): 1–28.
- 846 (Spiegelhalter et al. 2014) Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and der Linde, A.  
847 V. 2014. Bayesian measures of model complexity and fit (with discussion). *Journal of the*  
848 *Royal Statistical Society.* 64: 583–639.
- 849 (Tacoli et al. 2017) Tacoli, F., Mori, N., Pozzebon, A., Cargnus, E., Da Via, S., Zandigiacomo,  
850 P., Duso, C., and Pavan, F. 2017. Control of *Scaphoideus titanus* with Natural Products  
851 in Organic Vineyards. *Insects.* 8(4): 129.
- 852 (Tang et al. 2021) Tang, F. H. M., Lenzen, M., McBratney, A., and Maggi, F. 2021. Risk of  
853 pesticide pollution at the global scale. *Nature Geoscience.* 14(4): 206–210.

- 855 (Tramontini et al. 2020) Tramontini, S., Delbianco, A., and Vos, S. 2020. Pest survey card  
856 on flavescence dorée phytoplasma and its vector *Scaphoideus titanus*. EFSA Supporting  
857 Publications. 17(8): 1909E.
- 858 (Tscharntke and Brandl 2004) Tscharntke, T., and Brandl, R. 2004. Plant-insect interactions in  
859 fragmented landscapes [Publisher: Annual Reviews]. Annual Review of Entomology, 49(1),  
860 405–430.
- 861 (Vaclavik et al. 2010) Václavík, T., Kanaskie, A., Hansen, E. M., Ohmann, J. L., and Meen-  
862 temeyer, R. K. 2010. Predicting potential and actual distribution of sudden oak death  
863 in Oregon: Prioritizing landscape contexts for early detection and eradication of disease  
864 outbreaks. Forest Ecology and Management. 260(6): 1026–1035.
- 865 (Vanwambeke et al. 2019) Vanwambeke, S. O., Linard, C., and Gilbert, M. 2019. Emerging  
866 challenges of infectious diseases as a feature of land systems. Current Opinion in Environ-  
867 mental Sustainability. 38: 31–36.
- 868 (Watanabe 2010) Watanabe, S. 2010. Asymptotic equivalence of bayes cross validation and  
869 widely applicable information criterion in singular learning theory. Journal of Machine  
870 Learning Research. 11: 3571–3594.
- 871 (Weintraub and Beanland 2006) Weintraub, P. G., and Beanland, L. 2006. Insect Vectors of  
872 Phytoplasmas. Annual Review of Entomology. 51(1): 91–111.
- 873 (Zuur et al. 2017) Zuur, A. F., Ieno, E. N., and Saveliev, A. A. 2017. *Beginner's guide to spatial,*  
874 *temporal and spatial-temporal ecological data analysis with r-inla*, Highland Statistics Ltd.
- 875 (Zuur et al. 2009) Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., and Smith, G. M. 2009.  
876 *Mixed effects models and extensions in ecology with R*, New York, USA: Springer Science.

## 877 Supplementary Materials

878 **Supplementary Materials for "Field and landscape risk factors impacting Flaves-**  
 879 **cence dorée infection : Insights from spatial Bayesian modelling in the Bordeaux**  
 880 **vineyards" by Adrakey HK., Malembic-Maher S., Rusch A., Ay JS., Riley L., Ramalanjaona**  
 881 **L. and Fabre F. 2022. Phytopathology, early view.**

### 882 Text S1: Attributing CVI variables to field sampled by GDON

883 The Flavescence Doree (FD) dataset analysed provides the FD detection status of 34581 fields  
 884 sampled between 2012 and 2016. During their field trips, trained inspectors delimit in a GIS  
 885 database with a spatial polygon which they identify as an homogeneous area of vine cultivation,  
 886 thereafter termed GDON field. GDON fields are supposed to share the same year of plantation,  
 887 density of plantation and cultivar. Unfortunately, GDON inspectors do not have access to these  
 888 variables. However, they are available in the CVI (Casier Viticole Informatisé), a database  
 889 created by the French directorate general of customs. The CVI gives a comprehensive history  
 890 of each legal piece of land referenced in an administrative land register (Cadastral in French).  
 891 However, matching these information to GDON fields is not always straightforward as the  
 892 polygons drawn by GDON' inspectors do not necessarily strictly overlap those of the CVI.  
 893 Initial work (Table S1) focused on mapping information from the CVI to each GDON field -  
 894 enabling extraction of information such as the field's age and the majority cultivar grown as  
 895 shown in Fig. 1.

896 Most of the CVI fields are associated to a single agronomic field as defined by an unique set  
 897 of values for the year and plantation, plantation density and cultivar. Out of the 34581 GDON  
 898 fields, 15204 are associated to a single agronomic field. However, a CVI field can also be divided  
 899 into several agronomic fields. The difficulty here is that agronomic fields are not associated to  
 900 their own spatial polygon. Only their areas are known. A case by case study is then necessary  
 901 in order to assign the more likely values of year of plantation, plantation density and cultivar to  
 902 each GDON field. The cases are detailed in Table S1 along with their frequency of occurrence  
 903 among the 34581 GDON fields analysed. For year of plantation and plantation density, the  
 904 process consists in assigning to each GDON field metrics corresponding to mean weighted by  
 905 the surface of agronomic fields of the CVI field where the GDON field of interest is located. For  
 906 example an estimate of the planting year in the case 3 (Table S1) is derived as weighted mean of  
 907 all agronomic fields planting years weighted by their respective areas. Similarly, for the cultivar,  
 908 the process consists in assigning to GDON field the cultivar grown on the largest agronomic  
 909 field of the CVI. This process occurs in 11927 GDON fields (Table S1, cases 4 to 6). In these  
 910 cases, the largest agronomic field occupy on average 63% of the CVI field (quantile 10% = 0.43,  
 911 quantile 90% = 0.85).

912 Also, the spatial polygons of the GDON fields do not always exactly match those of the  
 913 CVI fields as they are drawn independently (initially taken from an ortho-photograph and later  
 914 adjusted during inspections). Accordingly, noting that GDON fields are in general smaller than  
 915 CVI fields, we assigned the information issued from a CVI field to a GDON field if its centroid  
 916 lies within the polygon of the CVI field.

Table S1: Columns 1-6: Summary of scenarios resulting from the information provided by the CVI and their interpretation. Column 7: The number of GDON fields to which CVI legal fields belonging to each scenario were assigned, with the sum corresponding to the size of the population under study.

Cases	Sum of agronomic field area within the administrative field	Number of agronomic fields within the administrative field	Year of planting	Cultivars	Interpretation	Number of occurrence in the Gdon data (2012-2016)
1	$\leq$ cadastral land area	1			Administrative field having only one agronomic field and with same or less area compared to the crop area	15203
2		$\geq 1$	Identical	Identical	Administrative field having multiple agronomic fields with the same planting year and cultivars and their total area is smaller than the cadastral one	1373
3			Different	Identical	Administrative field having multiple agronomic fields with the same cultivars but different planting year and their total area is smaller than the cadastral one.	6105
4			Identical	Different	Administrative field having multiple agronomic fields with different cultivars but the same planting year and their total area is smaller than the cadastral one	895
5			Different	Different	Administrative field having multiple agronomic fields with different cultivars, different planting year and their total area is smaller than the cadastral one	10976
6	$\geq$ cadastral land area	$\geq 1$	Different	Different	Administrative field having multiple agronomic fields with the same planting year and their total area is greater than the cadastral one	19
7			Identical	Identical	Administrative field having multiple agronomic fields with the same planting year and cultivars and their total area is greater than the cadastral one	8
8		1	Identical	Identical	Administrative field having only one agronomic field and with same or bigger area compared to the crop area	2

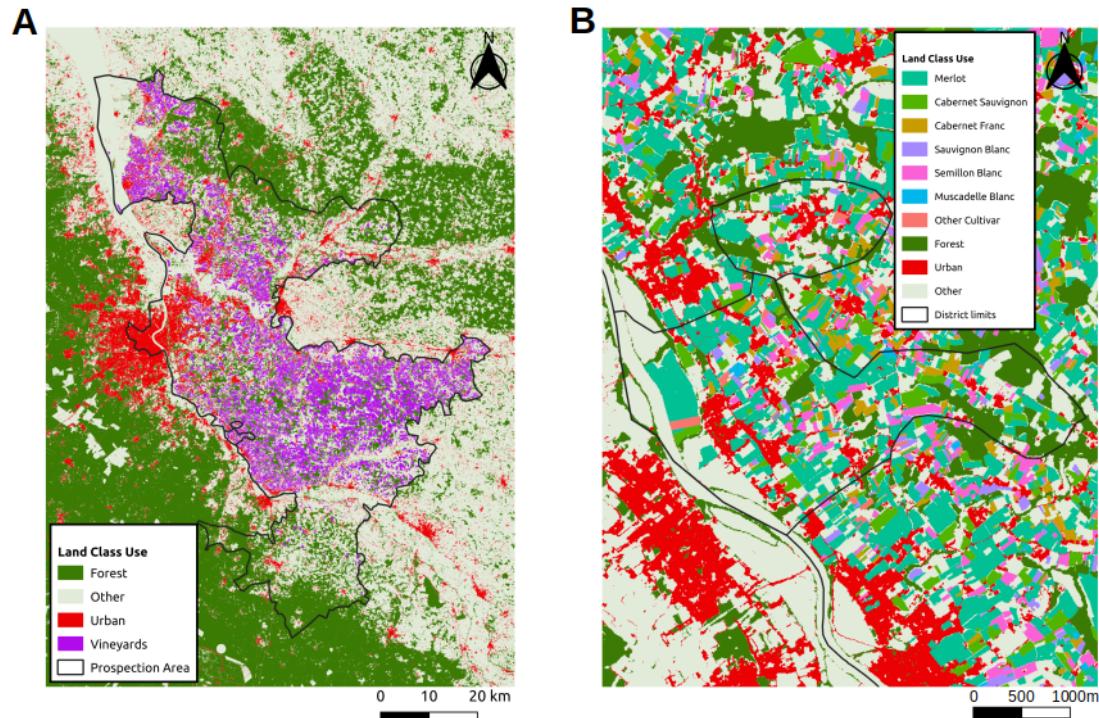


Figure S1: Land cover map for the year 2016. The land cover maps used to characterize the landscape surrounding each field are raster maps with a 10-meters resolution. Each pixel was classified into 1 of 11 categories. The first 8 categories describe the vineyard field : cultivars (categories 1 to 7) and cultural practice (category 8 : organic or conventional ; not illustrated here as it overlap with the cultivars). The 3 last categories corresponds to other land use : forest, urban or other land use. The class forest merges the classes 31 (Broad-leaved Forests) and 32 (Coniferous Forests) of the OSO map for 2017. Similarly, the class urban merges the classes 41 (Continuous Urbain Fabric), 42 (Discontinuous Urbain Fabric) and 43 (Industrial and Commercial Unites) of the OSO map for 2017. The panel A illustrates the 4 classes of land use vineyard, forest, urban and others over the whole area mointored by GDON des Bordeaux. The panel A illustrates the 4 class of land use vineyard, forest, urban and others over the whole area survey by GDON des Bordeaux. The panel B also details in a smaller area the cultivars used in the vineyard plots.

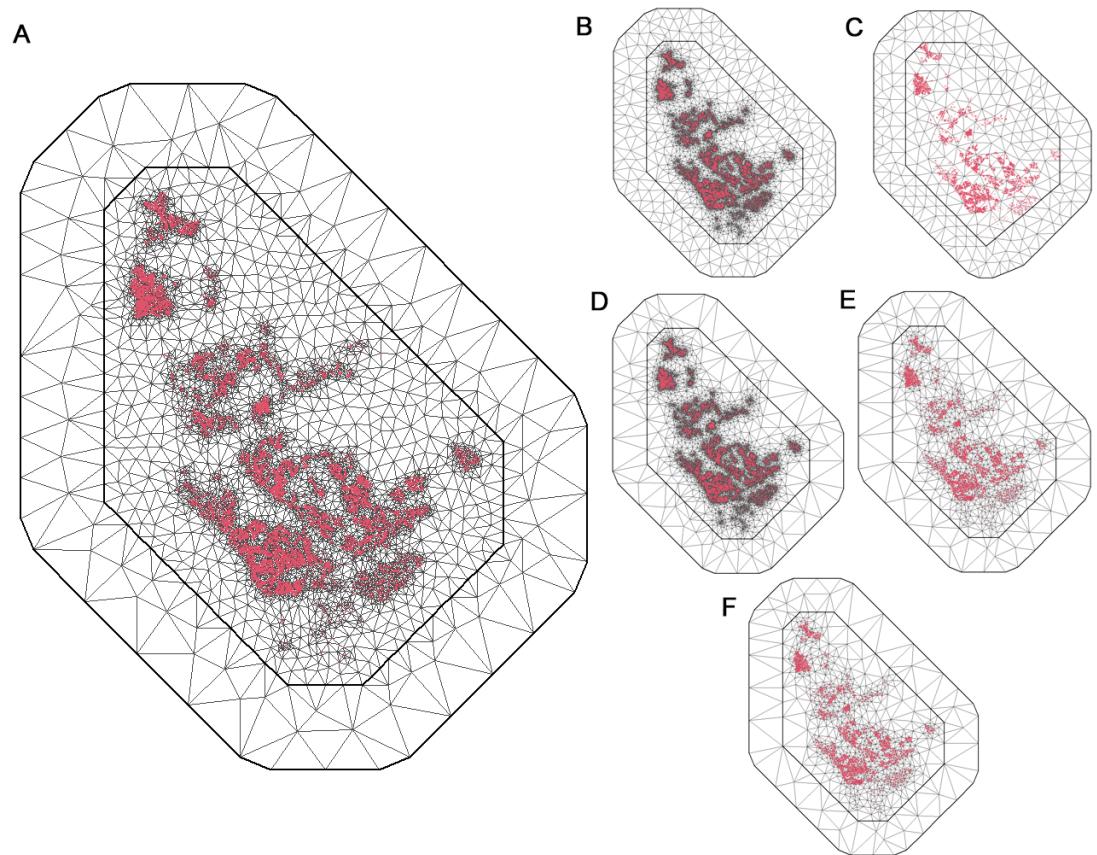


Figure S2: Boundary of the study region and the mesh created to approximate the solution of the SPDE to obtain the spatial model. A thick line separates the outer offset from the inner offset and the boundary of the study region. The mesh is extended beyond the study area to avoid a boundary effect. The meshes differ in the parameters `max.edge` (the largest allowed triangle edge length) and `cutoff` (the minimum allowed distance between points). A more precise mesh is expected to provide a better estimation of the spatial effect (the prediction will be smoother) but this comes at the cost of longer computational times. A: `max.edge=c(5,25)` and `cutoff=0.25`. This is the mesh used in the main text. B: `max.edge=c(10,10)` and `cutoff=0`. C: `max.edge=c(10,10)` and `cutoff=10`. D: `max.edge=c(50,50)` and `cutoff=1e-12`. E: `max.edge=c(75,75)` and `cutoff=1`. F: `max.edge=c(25,50)` and `cutoff=1`. Values are in kilometers. The corresponding number of vertices associated from panel A to F are 8845, 58209, 317, 58078, 1296 and 1296, respectively.

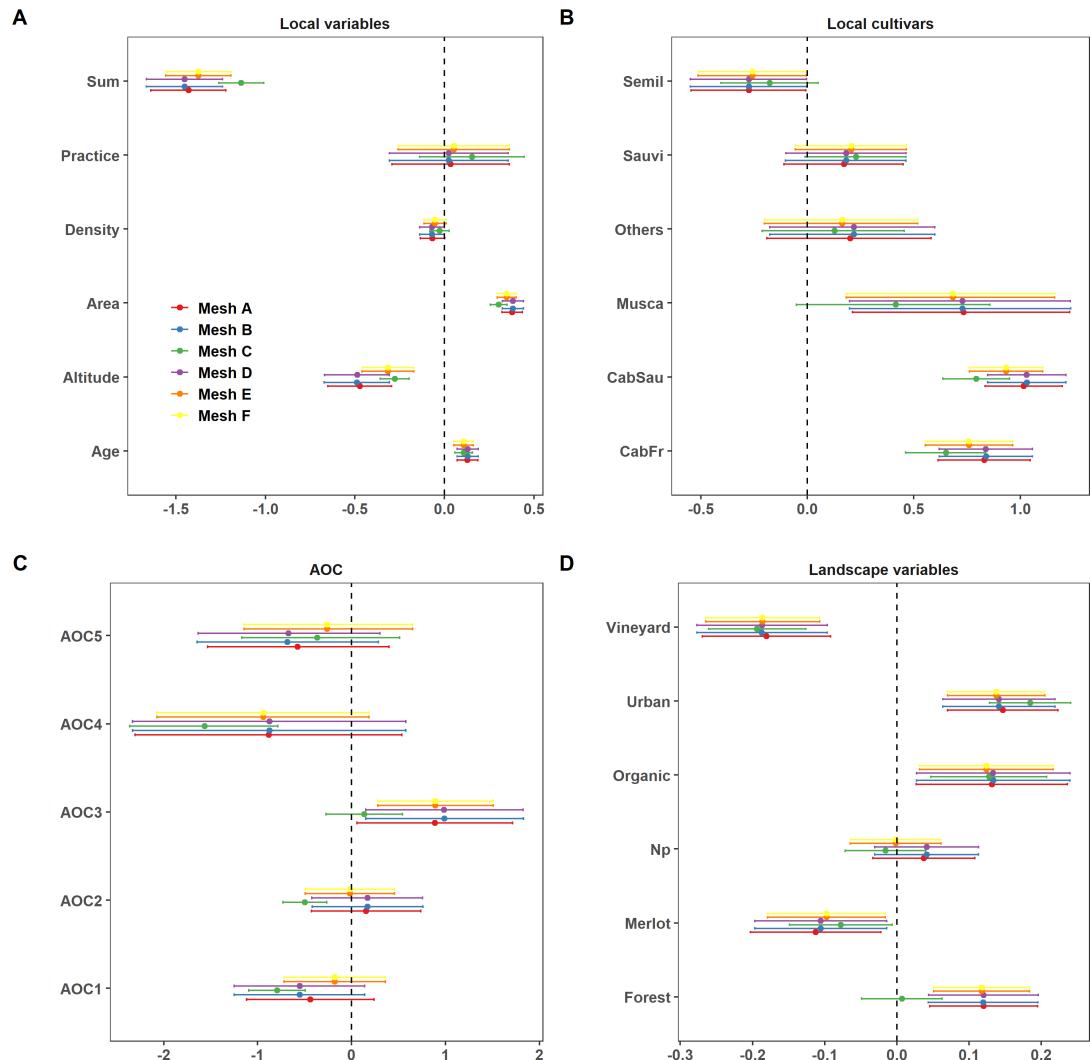


Figure S3: Estimation of the parameters associated to fixed effects of the baseline model  $M_0$  (at a landscape scale of 150 m) for six meshes. The panels displayed the posterior mean (dots) and 95% credible intervals (solid lines) of every local covariates, cultivars, AOC and landscape covariates for the six meshes used to define the spde model (Fig. S2). MeshA corresponds to the mesh used to derive results in the main text (Fig. S2A). Dashed line corresponds to the values 0. For clarity, we dropped the "L\_" from the landscape variables.

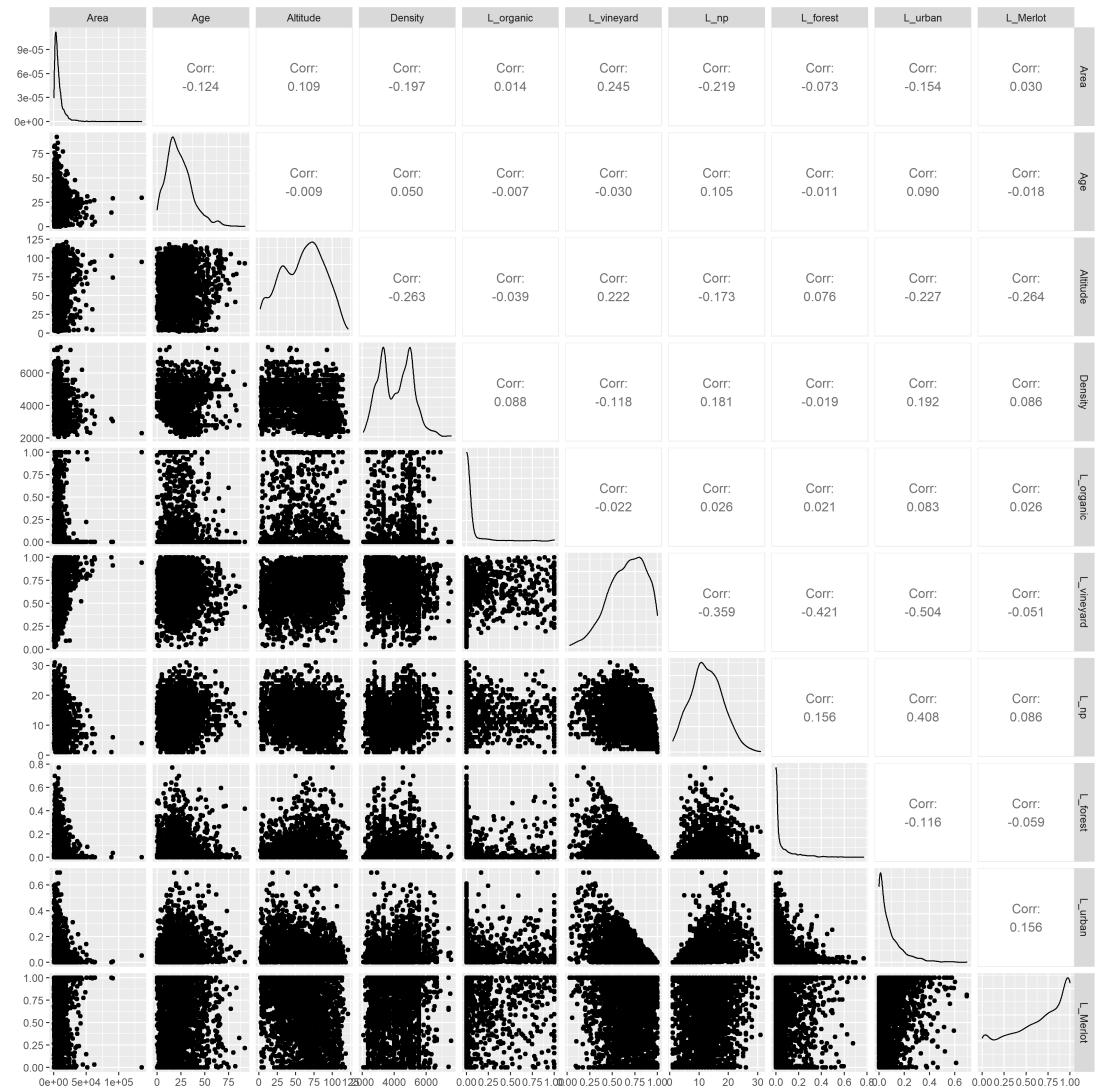


Figure S4: Correlation matrix between the continuous explanatory variables used in the model  $M_0$  at the landscape scale of 150 m. A random sample of 2500 out of the 34581 observations is used to draw the pairwise scatter plots.

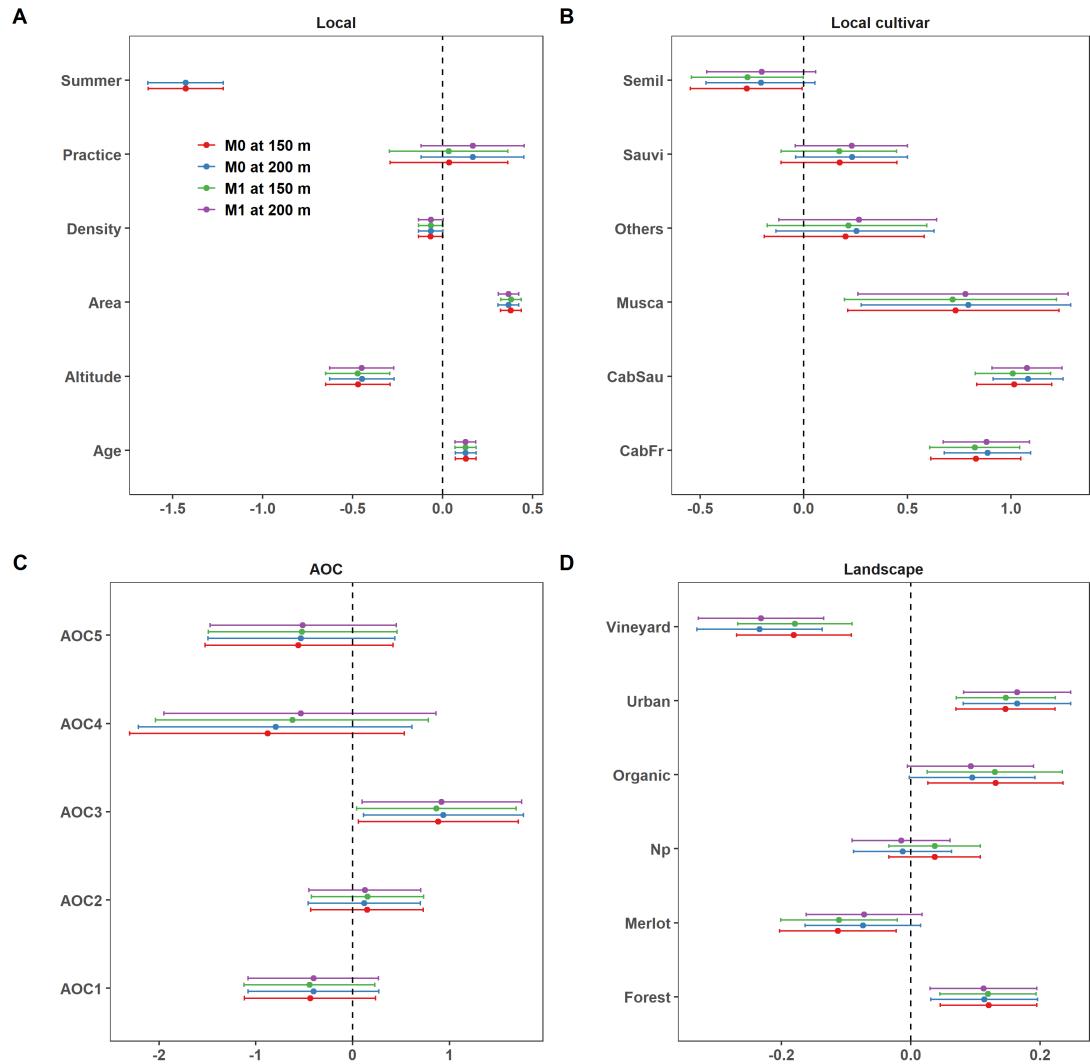


Figure S5: Estimation of the parameters associated to fixed effects of models  $M_0$  and  $M_1$  at the landscape scales of 150 m and 200 m. The landscape scale of 150 m is associated the lowest AIC and DIC for both models  $M_0$  and  $M_1$ , followed by the landscape scale of 200 m. The panels displayed the posterior mean (dots) and 95% credible intervals (solid lines) of every local covariates, cultivars, AOC and landscape covariates for the four combinations (two models times two scales). Dashed line corresponds to the values 0. For clarity, we dropped the "L\_" from the landscape variables

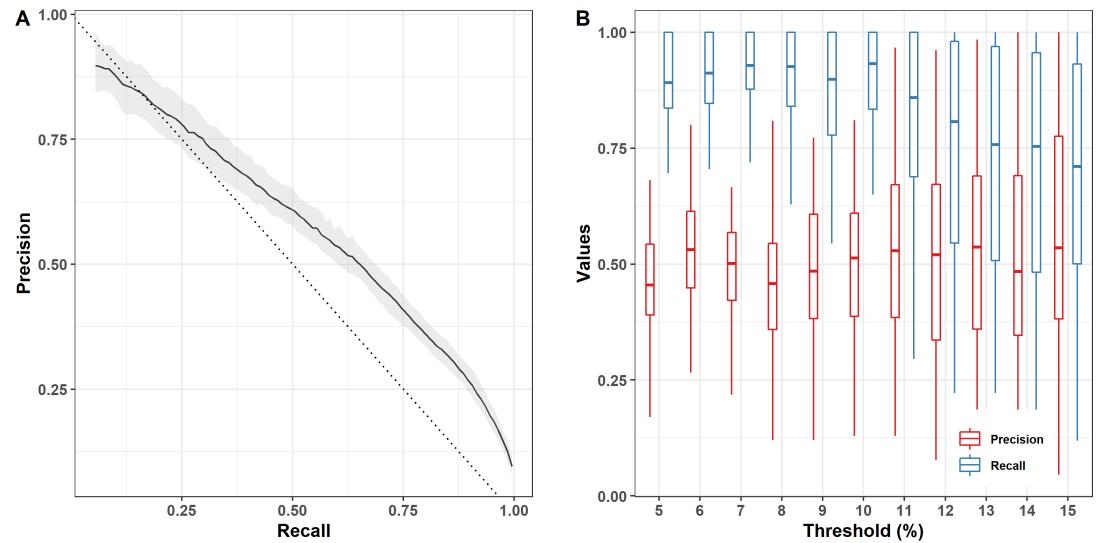


Figure S6: Predictive performance of the baseline model  $M_0$  in a landscape scale of 150 m. A: Predictive performance at field scale, as estimated by the mean precision-recall curve (in black) and its 80% credible band (in grey), based on 50 cross-validation partitions. B: Predictive performance at district scale, as estimated by the ability of the model to predict if the prevalence of FD detection in a district exceeds a threshold. For each threshold, the marginal posterior distribution of precision and recall were obtained from 50 cross-validation partitions.