ORIGINAL PAPER

Tracking the outbreak: an optimized sequential adaptive strategy for *Xylella fastidiosa* delimiting surveys

E. Lázaro · **M. Sesé** · **A. López-Quílez** · **D. Conesa** · **V. Dalmau** ·
A. Ferrer · **A. Vicent**

Abstract The EU plant health legislation enforces the implementation of intensive surveillance programs for quarantine pests. After an outbreak, surveys are implemented to delimit the extent of the infested zone and to manage disease control. Surveillance in agricultural and natural environments can be enhanced by increasing the survey efforts. Budget constraints often limit inspection and sampling intensities, thus making it necessary to adapt and optimize surveillance strategies. A sequential adaptive delimiting survey involving a three-phase and a two-phase design with increasing spatial resolution was developed and implemented for the *Xylella fastidiosa* demarcated area in Alicante, Spain. Inspection and sampling intensities were optimized using simulation-based methods. Sampling intensity thresholds were evaluated by quantifying their effect on the estimation of *X. fastidiosa* incidence. This strategy made it possible to sequence inspection and sampling taking into account increasing spatial resolutions, and to adapt the inspection and sampling intensities according to the information obtained in the previous, coarser, spatial resolution. The proposed strategy was able to efficiently delimit the extent of *Xylella fastidiosa*, while improving on the efficiency and maintaining the efficacy of the official survey campaign. From a methodological perspective, our approach provides new insights into alternative delimiting designs and new reference sampling intensity values.

E. Lázaro (✉) · M. Sesé · A. Vicent
Centre de Protecció Vegetal i Biotecnologia, Institut Valencià d'Investigacions Agràries (IVIA), Moncada, Spain
e-mail: lazaro_ele@gva.es

M. Sesé · A. López-Quílez · D. Conesa
Departament d'Estadística e Investigació Operativa, Universitat de València, València, Spain

V. Dalmau · A. Ferrer
Servei de Sanitat Vegetal, Conselleria d'Agricultura, Desenvolupament Rural, Emergència Climàtica i Transició Ecológica, Generalitat Valenciana, Silla, Spain

Introduction

Emerging infectious diseases of plants are an increasing threat to wild and cultivated plants worldwide. After a pathogen or pest has been found in an area, the implementation of delimiting surveillance actions is pivotal to establish management actions such as eradication or containment programs to prevent

further spread (Mastin et al. 2020). European Union Regulation (EU) 2016/2031 (EU 2016b) provides the general framework for plant pest surveys under a risk-based approach. It is clearly established that the focus must be on maximizing detection of infected host plants. Thus, there is a need to adapt and optimize delimiting strategies to each epidemiological situation in order to set the number of host plants to be inspected and define where to sample, minimizing the impacts of the control measures (Vicent and Blasco 2017).

The bacterium *Xylella fastidiosa* (*Xf*) is a xylem-inhabiting pathogen (Wells et al. 1987) that causes different diseases in a wide range of cultivated, ornamental, and forest plant species (EFSA 2020c; Saponari et al. 2019). The species is naturally trans-mitted by xylem sap-feeding insects. The natural spread of *X. fastidiosa* depends on these insects, which can propagate the disease from plant to plant depend-ing on their dispersal capability and transmission efficiency (Bodino et al. 2020). Human-assisted spread is mainly driven by the movement of infected plant materials (EFSA 2018).

The species *Xf* is a priority "pest" in the EU (EU 2016b, 2019a, b). Since 2013, the presence of *Xf* has been confirmed in Italy (2013), France (2015), Spain (2016), and Portugal (2018). In 2019, the presence of *Xf* was officially reported in Israel (EPPO 2019). After its first detection in the EU, additional emergency measures were enforced for *Xf* in order to prevent further spread under Decision 2015/789/EU (EU 2015b, a, 2016a, 2017, 2018b, a) (hereinafter referred to as "the Decision").

Among other actions, the Decision established the implementation of two different surveillance actions, detection and delimiting surveys, depending on the pathogen status. Detection surveys are aimed at detecting the pathogen and determining the status of the "pest". Delimiting surveys are conducted to demarcate the geographic extent of the pathogen after an outbreak. A temporary demarcated area split into an infested zone (i.e., infected zone in the Decision) and a buffer zone must be established following a new detection. Delimiting surveys for demarcated areas should be conducted on an annual basis (EU 2016b) in order to update their boundaries and their epidemio-logical status for the adequate implementation of control measures, such as the removal of infected plants and vector control.

The current situation of *Xf* in the EU is raising major concerns that are leading to a phytosanitary emergency. Two of the most relevant crops in the Mediterranean areas of the EU, olives and almonds, are severely affected by *Xf* diseases. Grapevines were also found to be affected in the Balearic Islands, Spain, and other major crops such as citrus are at risk. Furthermore, the insect vector *Philaenus spumarius* (Hemiptera: Aphrophoridae) is widespread in affected regions in Spain and Italy, as well as other areas of the EU territory (Saponari et al. 2014; Cornara et al. 2017; EFSA 2018). The efficacy of any eradication or containment measures relies heavily on a precise delimitation of the infested and buffer zones to tailor control interventions. Therefore, there is a need to design delimiting surveys that maintain an accept-able level of risk of overlooking the detection of positives (Hauser et al. 2016).

During the last two decades innovative surveillance methods have been developed taking optimization-based tools into consideration with the aim of improving survey efficiency (Epanchin-Niell and Liebhold 2015; Hauser et al. 2016; Moore and McCarthy 2016; Yemshanov et al. 2017a, b; Büyüktahtakın and Haight 2018). Priority has been given to the development of innovative and advanced modelling methods for detection, delimiting, and control, with ongoing research activity to estimate disease spread and to find optimal surveillance and/or management scenarios (White et al. 2017, 2019; Mar-tinetti and Soubeyrand 2019; Schneider et al. 2020).

Recent advances in surveillance methodologies focus not only on improving the sampling design but also on developing more sophisticated data analyses to overcome deficiencies in the survey design (Pacifici et al. 2016). In relation to the sampling methodology, there are several strategies such as sequential sam-pling (Chaudhuri and Stenger 2005), stratified sam-pling (Edwards et al. 2005) or adaptive sampling (Brown et al. 2013) that can increase the information content and provide a more efficient estimation when disease distribution is spatially correlated (Pacifici et al. 2016).

On the other hand, the use of statistical models can overcome deficiencies in the survey design and data collection. For instance, spatial autocorrelation and other factors associated with imperfect survey tech-niques such as observer bias, sampling gaps or missing data can now be included in the modeling process

given the developments made in statistical and computing methods (Latimer et al. 2006; Banerjee et al. 2014; Martínez-Minaya et al. 2018).

Considering the survey protocol outlined in the Decision as the point of reference, here we develop an alternative strategy for performing delimiting surveys in the demarcated area of Alicante. The outbreak in Alicante was first reported in June 2017 and *Xf* subsp. *multiplex* was identified (Schaad et al. 2004; Arias-Giraldo et al. 2020; Landa et al. 2020). Since then delimiting surveys have been conducted to update the boundaries of the demarcated area and eradication measures have been applied, including the removal of all host plants within the infested zones and vector control. This outbreak, which in 2019 covered a demarcated area of 140,000 ha, represents one of the largest plant disease eradication campaigns ever attempted in Europe with circa 4,800 orchards and 75,000 trees already destroyed (GVA 2020).

The emergency measures enforced in Alicante due to *Xf* have encountered serious difficulties to be implemented (DG-SANTE 2018, 2019). The delimiting survey efforts that were required were difficult to reach because the inspection and sampling intensities established by the Decision were higher than the survey and laboratory capacities available. Additionally, eradication measures have generated public opposition especially with regard to the removal of trees. Thus, there is a need to tailor the delimiting survey strategy to the specific epidemiological framework in order to optimize an accurate demarcation of the affected area, which in turn ensures the effectiveness of control strategies.

In our study, we propose an alternative delimiting strategy with a sequential adaptive scheme that combines different survey phases with increasing spatial resolutions to delineate the infested zones in the survey area. The sequential adaptive strategy was simulated considering a two-phase and a three-phase design for the particular case of *Xf* in Alicante. For each phase, considering the sampling of both asymptomatic and symptomatic plants, optimum inspection and sampling intensities were estimated from the 2018 official delimiting survey campaign using simulation-based optimization methods. The sequential introduction of increasing survey resolutions will allow for greater accuracy in delimiting infested zones and the implementation of control measures in a more targeted way. The performance of the alternative delimiting strategy was compared with the Decision delimiting strategy (hereinafter referred to as the "official strategy") in terms of survey efforts.

An additional objective was to determine the influence of sampling intensity on the estimates of the spatial distribution of *Xf* incidence in the demarcated area in Alicante. A Bayesian hierarchical spatial model was selected to infer this metric using the data from the official survey. Under the selected spatial modeling framework, several subsets of the data were considered to infer incidence. These data subsets were generated from the official survey data restricted by the optimum sampling intensity estimated through the sequential adaptive strategy. A comparison of the performance of the models fitted under these database subsets was carried out by quantifying differences in incidence estimates and their derived quantities using several discrepancy measures.

Material and methods

We used the 2018 official survey data collected in the demarcated area for *Xf* in Alicante as the basis to feed an optimization algorithm specified under a sequential adaptive strategy. Algorithm conditions were set to ensure minimum inspection and sampling intensities for different survey resolutions while maintaining the delimitation efficacy reached by the official strategy. The sequential adaptive strategy was defined considering a three-phase and a two-phase design. Three-phase design combined the delimitation at 1 km^2, 0.25 km^2 and 0.01 km^2 grid resolutions while the two-phase set them at 1 km^2 and 0.01 km^2. Survey efforts were calculated and compared based on the estimates of the inspection and sampling intensities for each of the resolutions. Furthermore, we assessed the influence of different sampling intensity values by comparing the estimation of the distribution of *Xf* incidence with a Bayesian hierarchical spatial model.

Study area, official delimiting survey and database

The data used in the present study were from the 2018 official delimiting survey campaign for *Xf* in Alicante, Spain, (î "reference database"). In compliance with the Decision, the demarcated area (i.e., the survey area) consisted of an overall buffer formed by buffer

zones with a radius of 5 km around 71 individual infested zones. An infested zone was defined as the area within a radius of 100 m around a plant positive for *Xf* based on real-time PCR (EPPO 2019).

The demarcated area was surveyed including visual inspections followed by sampling and laboratory testing of individual host plants, preferentially when *Xf*-like symptoms were observed. The official strategy was organized at two spatial resolutions: (i) 100×100 m grid cells (0.01 km²) in the first km radius around each infested zone, and (ii) on 1×1 km grid cells (1 km²) in the rest of the buffer zone. Thus, the survey area was subdivided into 656 cells of 1 km² and 17,700 cells of 0.01 km² covering an extension of 83,300 ha. Altogether 8,142 samples were collected and tested. The grid layout of the 2018 Alicante survey is shown in Fig. 1a and the distribution of the samples can be seen in Fig. 1b.

A description of the data categorized by grid resolution is provided in Table 1. A 500 m × 500 m (0.25 km²) grid was included to be considered later in the sequential adaptive design (see next section). For the 1 km² grid, the percentage of cells sampled reached 100% with a rate of positives for *Xf* of 8.52%, and median and maximum sampling intensity values of 5 and 109 samples/cell. For the 0.25 km² and 0.01 km² grids, the percentages of sampled cells were 49.34% and 4.02% with a rate of positives for *Xf* of 2.67% and 0.19%, respectively. The median sampling intensity was 1 sample/cell in both grid resolutions, with a maximum of 101 samples/cell for the 0.25 km² grid and 59 samples/cell for the 0.01 km² grid.

Data related to sampling date, plant species, presence of *Xf*-like symptoms (i.e., symptomatic vs. asymptomatic), and GPS coordinates in the UTM system were also collected. In total, 124 species were sampled with *Olea europaea*, *Prunus dulcis*, *Ficus carica*, *Rosmarinus officinalis*, and *Vitis* spp. accounting for 62.06% of the total sample size (Table 2). The vast majority of the selected host plant species are known to be infected by multiple *Xf* subspecies, not only by subsp. *multiplex* but also by *fastidiosa* and *pauca* (EFSA 2020c) as it recommended by EFSA (2019a, 2020b). As indicated above, plants expressing *Xf*-like symptoms were preferentially sampled, representing 85% of the total. In the laboratory analysis, 3.26% of the symptomatic samples were found to be positive, while only 0.84% of the asymptomatic ones were positive. Altogether, 237 (2.91% of the total) samples were *Xf* positive with 221 of them belonging to *Prunus dulcis*, while the others corresponded to *R. officinalis* (1), *R. alaternus* (1), *P. armeniaca* (1), *P. myrtifolia* (5), *Phagnalon saxatile* (3), *Helycrhysum italicum* (3), *Calicotome spinosa* (1), and *Scabiosa atropurpurea* (1).

(a) 0.01 km² and 1 km² grid distribution

(b) Samples distribution
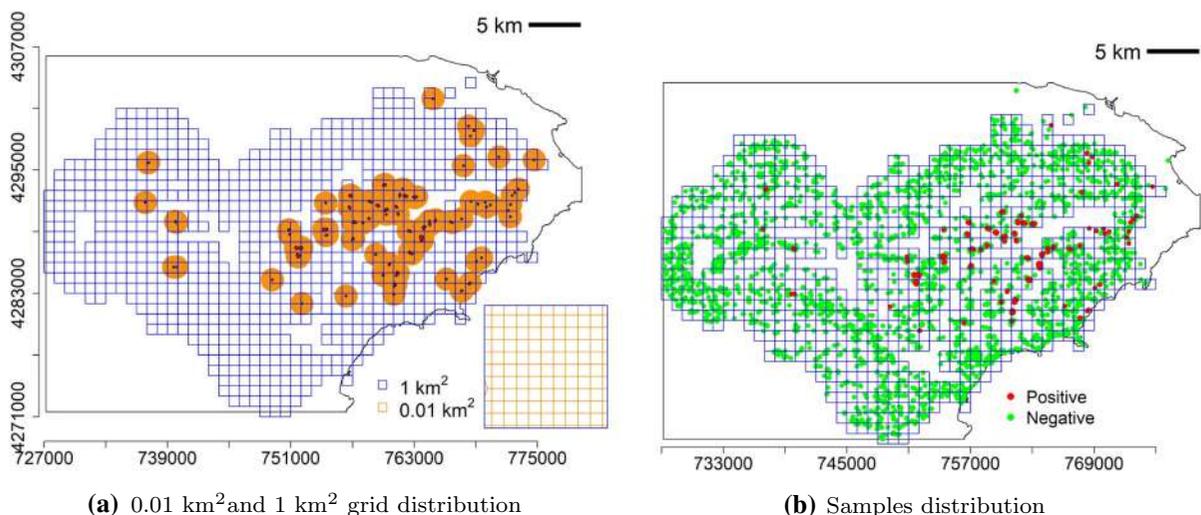
Fig. 1 Distribution of **a** 0.01 km² (orange cells) and 1 km² grid (blue cells), and **b** samples in the 1 km² grid positive and negative for *Xylella fastidiosa* in the demarcated area in Alicante, Spain, in 2018. The orange circles represent the 1st kilometer radius around each infested zone, which are split into 0.01 km² grid cells. Empty spaces (non-sampled grid cells) within the grid represent inaccessible sites and those without host plants

Table 1 Data from the official delimiting survey campaign in the demarcated area in Alicante, Spain, for *Xylella fastidiosa* in 2018 categorized by different grid resolutions ($j = \{1, 0.25, 0.01\}$ km^2)

Grid resolution (km^2)	Number of cells (%)			Sampling intensity (samples/cell)	
	C_j	$C_{j,s}$	$C_{j,+}$	Median	Maximum
$j = 1$	833	833 (100%)	71 (8.52%)	5	109
$j = 0.25$	3,332	1,644 (49.34%)	89 (2.67%)	1	101
$j = 0.01$	83,100	3,340 (4.02%)	161 (0.19%)	1	59

C_j denotes the number of cells covering the demarcated area depending on the grid resolution (j); $C_{j,s}$ is the number of cells per grid (j) in which at least one sample was taken; and $C_{j,+}$ is the number of cells per grid (j) in which at least one sample was detected as positive for *X. fastidiosa*. Sampling intensity (samples/cell) is described by the median and the maximum values

Table 2 Absolute and relative frequency distribution of the presence of symptoms and positive detection of *Xylella fastidiosa* in the database of the 2018 official survey in Alicante categorized by plant species

| Plant species	Asymptomatic (Positive/Total) (%)	Symptomatic (Positive/Total) (%)	Positive/Total (%)
Olea europea	0/260	0/1,804	0/2,064
Prunus dulcis	10/122	211/1,507	221/1,629
Ficus carica	0/203	0/342	0/545
Rosmarinus officinalis	0/17	1/482	1/499
Vitis spp.	0/103	0/213	0/316
Other	0/480	15/2,609	15/3,089
	10/1,185 (0.84%)	227/6,957 (3.26%)	237/8,142 (2.91%)

Sequential adaptive delimiting survey strategy

We propose a new delimiting strategy that seeks to improve the efficiency of the official strategy while preserving its delimiting efficacy. Optimum inspection intensity (i.e., number of cells inspected for each grid resolution) and sampling intensity (i.e., number of samples taken and tested in each cell) were estimated using the reference database.

The proposal was based on a sequential adaptive scheme that considers 1 km^2, 0.25 km^2 and 0.01 km^2 grid resolutions to be combined in a three-phase survey design or 1 km^2 and 0.01 km^2 grid resolutions in a two-phase survey design. Assuming an initial survey resolution of 1 km^2, the sequential approach allows the phases of the survey to be scheduled in sequential time frames. Additionally, the adaptive approach allows the inspection and sampling intensities to be tailored for each phase depending on the results obtained in the previous coarser spatial

resolution. The phase sequences of the two designs are illustrated with an example in Fig. 2.

Our proposal considers inspection and sampling of all 1 km^2 cells in the demarcated area, and inspection and sampling are performed at a finer spatial resolution only in those cells in which *Xf* was detected. Inspection intensities for 0.25 km^2 ($C_{0.25}$) and 0.01 km^2 ($C_{0.01}$) grid resolutions and sampling intensities for each phase (n_1, $n_{0.25}$, $n_{0.01}$ for 1, 0.25, 0.01 km^2 grid resolutions) are solved by adopting the optimization algorithm which is described as follows and illustrated in Fig. 3. The algorithm was designed to minimize the sampling intensity while identifying all the positive cells for *Xf*. For phase 1, samples aggregated per grid cell are fed into the algorithm, whereas for the subsequent phases, a subset was generated from the grid cells selected. Note that in the reference database both symptomatic and asymptomatic plants were sampled (Table 2) to account for the long incubation period that characterizes *Xf* (EFSA 2020a, b).
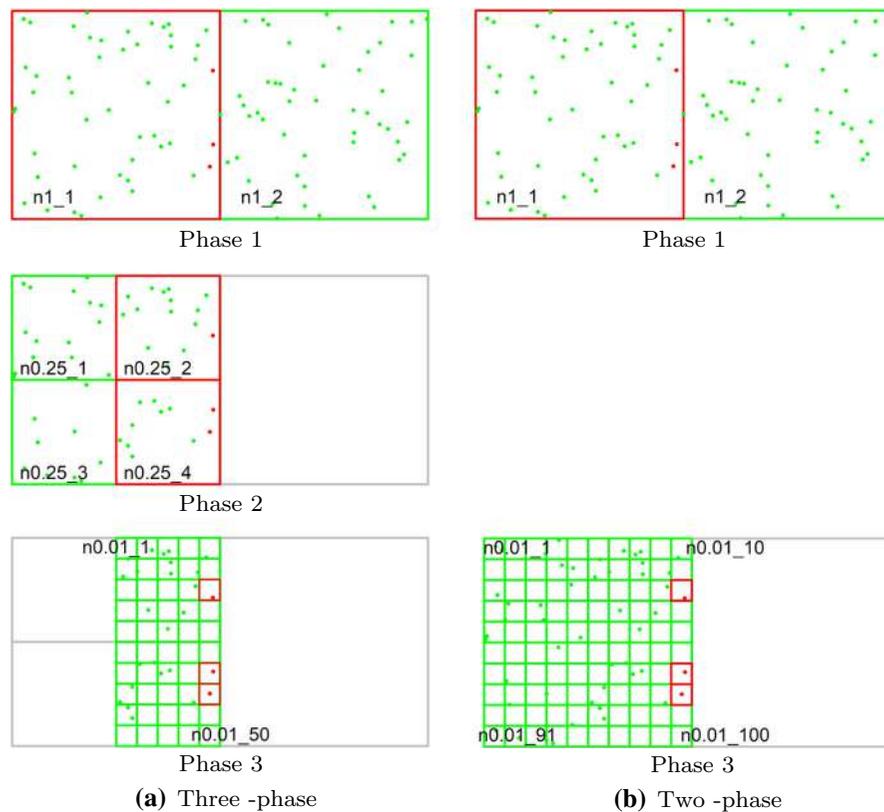
Fig. 2 Example of the sequential adaptive survey strategy for **a** the three-phase and **b** the two-phase designs for 1 km^2 (row 1), 0.25 km^2 (row 2), and 0.01 km^2 (row 3) grid resolutions. Red and green denote positive and negative samples (dots) or cells (squares), respectively, for *Xylella fastidiosa*. Gray cells represent those not inspected or sampled for a specific grid resolution. n_j for $j = \{1, 0.25, 0.01\}$ denotes sampling intensity (samples/cell) for the different grids

Phase 1.	For all 1 km^2 cells:
Step 1.	Initialize n_1 as 1, where n_1 denotes the sampling intensity for the 1 km^2 grid resolution.
Step 2.	Run 100 replicates ($R_1 = 100$) of a random sampling under the restriction imposed by the value n_1. That is, n_1 corresponds to the maximum sampling intensity allowed per cell.
Step 3.	Stopping rule: if 50% of the replicates (R_1) have identified all the positive cells for *Xf* in the 1 km^2 grid resolution ($C_{1,+}$), the current value of n_1 is the optimum sampling intensity. If not, go back to *step 1* increasing n_1 by 1 and continue the sequence.

Phase 2.	Only for 1 km^2 cells identified as positive for *Xf* in *phase 1*. (Phase 2 is not considered in the two-phase design):
Step 1.	Initialize $n_{0.25}$ as 1, where $n_{0.25}$ denotes the sampling intensity for the 0.25 km^2 grid resolution. That is, $n_{0.25}$ corresponds to the maximum sampling intensity allowed per cell.
Step 2.	Run 100 replicates ($R_{0.25} = 100$) of a random sampling under the restriction imposed by the value $n_{0.25}$.
Step 3.	Stopping rule: if 50% of the replicates ($R_{0.25}$) have identified all the positive cells for *Xf* in the 0.25 km^2 grid resolution ($C_{0.25,+}$), the current value of $n_{0.25}$ is the optimum sampling intensity. If not, go back to *step 1* increasing n_1 by 1 and continue the sequence.
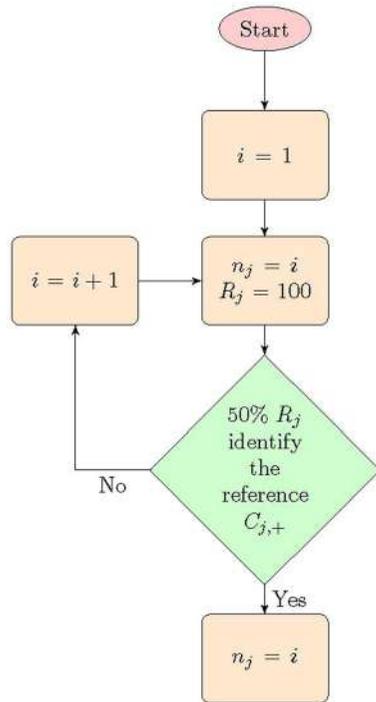
Fig. 3 Optimization algorithm to estimate optimum sampling intensity n_j for $j = \{1, 0.25, 0.01\}$ in the three-phase design and $j = \{1, 0.01\}$ for the two-phase design. R_j denotes the number of replicates and $C_{j,+}$ the number of positive cells for *Xylella fastidiosa* in each j grid resolution

Phase 3. Only for 0.25 km² cells identified as positive for *Xf* in *phase 2* (three-phase design) or 1 km² cells identified as positive in *phase 1* (two-phase design):

Step 1. Initialize $n_{0.01}$ as 1, where $n_{0.01}$ denotes the sampling intensity for the 0.01 km² grid resolution.

Step 2. Run 100 replicates ($R_{0.01} = 100$) of a random sampling under the restriction imposed by the value $n_{0.01}$. That is, $n_{0.01}$ corresponds to the maximum sampling intensity allowed per cell.

Step 3. Stopping rule: if 50% of the replicates ($R_{0.01}$) have identified all the positive cells for *Xf* in the 0.01 km² grid resolution ($C_{0.01,+}$), the current value of $n_{0.01}$ is the optimum sampling intensity. If not, go back to *step 1* increasing $n_{0.01}$ by 1, and continue the sequence.

Efficiency was assessed in the survey area (i.e., all cells covered by host plants) by means of the survey efforts (i.e., the total number of samples) that were computed for the three-phase and two-phase designs as:

$$N_{(three-phase)} = [n_1 \times C_1] + [n_{0.25} \times C_{1,+} \times 4] + [n_{0.01} \times C_{0.25,+} \times 25],$$
$$= [n_1 \times C_1] + [n_{0.25} \times C_{0.25}] + [n_{0.01} \times C_{0.01}],$$
$$N_{(two-phase)} = [n_1 \times C_1] + [n_{0.01} \times C_{1,+} \times 100],$$
$$= [n_1 \times C_1] + [n_{0.01} \times C_{0.01}],$$

where $C_1, C_{0.25}, C_{0.01}$ are the inspection intensities and $n_1, n_{0.25}, n_{0.01}$ are the optimum sampling intensities for the 1 km², 0.25 km², and 0.01 km² grid resolutions, respectively. Note that the inspection intensities for 0.25 km² and 0.01 km² are dependent on the number of cells identified as positive in the previous resolution. The equivalences of grid cells for the three-phase design are 1 km² ≡ 4 cells of 0.25 km² and 0.25 km² ≡ 25 cells of 0.01 km² . Likewise, for the two-phase design it is 1 km² ≡ 100 cells of 0.01 km².

For the official strategy, survey efforts were computed considering the pre-established inspection intensities and the optimum sampling intensities (n_1 and $n_{0.01}$) estimated for the two-phase design. Note that in the official strategy the sampling intensity is based on visual observation of symptoms, thus no minimum sampling intensity value was referred to.

Modeling the spatial distribution of *Xf* incidence

Bayesian hierarchical spatial model

Based on previous work (Cendoya et al. 2020), the spatial variation of *Xf* incidence was modeled by means of a Bayesian hierarchical spatial model. Incidence (i.e., the proportion of plants positive for *Xf*) is a magnitude commonly used in plant pathology to characterize the disease status or to evaluate the efficacy of a control program (Madden et al. 2007). This analysis was performed using the reference database georeferenced to the grid of 1 km² cells described above. The inference process was addressed under the integrated nested Laplace approximation

(INLA) proposed and implemented through the R-INLA package (Rue et al. 2009).

The model was formulated as

$$Y_i \sim \text{Binomial } (m_i, \pi_i) \quad i = 1, \ldots, n,$$
$$logit(\pi_i) = \boldsymbol{X}_i \boldsymbol{\beta} + v_i + u_i,$$
$$v_i | v_j \sim \text{N} \left(\frac{1}{k_i} \sum_{i \sim j} v_j, \frac{1}{\tau_v k_i} \right) i \neq j,$$
$$u_i \sim \text{N} (0, \tau_u) \quad i = 1, \ldots, n,$$
$$\beta_j \sim \text{N} (\mu = 0, \tau = 0.001) \quad j = 0, \ldots, M,$$
$$log(\tau_v) \sim \text{logGamma } (1, 5 \cdot 10^{-5}),$$
$$log(\tau_u) \sim \text{logGamma } (1, 5 \cdot 10^{-5}).$$

(1)

in which the number of samples positive for *Xf* in each grid cell (Y_i) was assumed to follow a Binomial distribution with π_i and m_i denoting the probability of a sample being positive for *Xf* and the total number of samples in cell i, respectively. The linear predictor was defined by a vector of covariates and its corresponding vector of coefficients, $\boldsymbol{X}_i$ and $\boldsymbol{\beta}$, and by a spatial and an independent random effect associated to each cell i, v_i, and u_i, respectively. Note that a non-informative scenario was considered to set prior specification with normal distributions centered at zero and a small precision for regression coefficients and log-gamma priors with a wide mean and variance for precisions of the spatial and independent random effects.

Three bioclimatic covariates were initially included and evaluated: annual mean temperature (°C) (coded as *bio*1), annual range temperature (°C) (coded as *bio*7), and precipitation of the wettest month (mm) (coded as *bio*13). These were acquired from the WorldClim v.2 database (Fick and Hijmans 2017) as spatially gridded climatic data (30" arc min resolution) and transformed to the UTM system with the raster package in R (Hijmans et al. 2015).

This model formulation makes it possible to take into account similarities among neighboring cells and also to quantify intra-cell behaviour (unstructured random effects). Following Cendoya et al. (2020), the neighborhood criterion was established at a maximum distance of 2.5 km among all the cells to ensure that all of them had at least one neighbor. Prior to including this spatial effect in the model, Moran's I test was used to check for the existence of spatial autocorrelation in the *Xf* incidence distribution (Dormann et al. 2007).

With the aim of selecting the most parsimonious model with the best explanatory and predictive abilities, all possible model component combinations, 2^5 (with 5 denoting the number of components of the linear predictor including random effects), were assessed. For this purpose, two selection model criteria were used: the Watanabe-Akaike information criteria (WAIC) (Watanabe 2010; Gelman et al. 2014), and the logarithmic conditional predictive ordinate (LCPO). While WAIC evaluates goodness of fit and model complexity, LCPO addresses predictive ability. The models with the lowest WAIC and LCPO values were chosen.

Effect of sampling intensity on Xf incidence estimates

To evaluate the influence of sampling intensity on the estimates of *Xf* incidence, different data subsets were built from the reference database. Specifically, four data subsets were obtained by limiting the maximum sampling intensity per 1 km² cell to: (i) the optimum sampling intensity calculated with the optimization algorithm for the 1 km² grid resolution (Sect. "Sequential adaptive delimiting survey strategy"); (ii) the third quartile of the sampling intensity observed in the reference database; and (iii) two arbitrary sampling intensities inbetween those values.

The data subset generation process consisted in running a simple random sampling for each cell, in which the sampling intensity was set according to the thresholds indicated above. Each data subset was replicated 100 times ($R = 100$) to ensure the inclusion of a wide range of sampling configurations. The spatial distribution of *Xf* incidence was estimated for each of the data subsets under the model specification selected based on WAIC and LCPO criteria.

The estimates obtained for each data subset were compared with those obtained using the reference database. As each data subset consisted of 100 replicas ($R = 100$), outputs were averaged to provide a single result for each data subset condition. The comparison was focused on assessing the stability of the estimates of: i) the regression coefficients of the bioclimatic covariates and standard deviations of the spatial and independent effects (i.e., the linear predictor parameters), ii) the incidence per cell, iii) the uncertainty associated with the incidence estimate, and iv) the spatial and independent effects per cell. Note that the

spatial and independent random effects were characterized by their corresponding standard deviation: $\sigma_v = 1/\sqrt{\tau_v}$ and $\sigma_u = 1/\sqrt{\tau_u}$.

The parameter estimates were described by their mean (i.e., point estimate) and their variance (i.e., uncertainty measure). Their corresponding stability was evaluated by means of the following discrepancy measures:

Bias	Difference between the average of the mean values of the replicas and the mean value of the reference inference process, $(\sum_{r=1}^{R} \bar{\theta}_{(r)}/R) - \theta$, where R is the number of replicas ($R = 100$ in our study), $\bar{\theta}_{(r)}$ is the mean of the relevant parameter corresponding to replica r, and θ is the mean of the relevant parameter corresponding to the reference model.
Standard error (SE)	Square root of the average of the variances of the replicas, $\sqrt{\sum_{r=1}^{R} s^2_{(r)}/R}$, where $s^2_{(r)}$ is the variance associated with each replica r parameter.
Standard deviation (SD)	Standard deviation of the set $\{\bar{\theta}_{(1)}, \ldots, \bar{\theta}_{(R)}\}$ that includes the mean of the parameter estimate of all replicas.

Note that $\{\bar{\theta}_{(1)}, \ldots, \bar{\theta}_{(R)}\}$ represents the vector of means for a generic regression parameter (denoted as θ) corresponding to the replica r, $r = 1, \ldots, 100$, of the inferential process.

The stability of the derived quantities (the incidence per cell and its corresponding standard deviation, spatial and independent effects) was assessed by means of the bias, computed as the difference between the average of the means (for the derived quantities) of the replicas and the mean of the reference inference process.

Results

Comparison between the official and the sequential adaptive delimiting survey strategies

Inspection intensity

Inspection intensities estimated for the two sequential adaptive designs (three-phase and two-phase) and the pre-established official strategy are described in Table 3. Both sequential adaptive designs required a higher inspection intensity (833 cells) at 1 km^2 than with the official strategy (656 cells). The sequential adaptive approaches were designed to cover the full extent of the demarcated area with inspections organized on a 1 km^2 grid in the initial phase. In contrast, the official strategy considers inspections in a single phase in which the first kilometer radius of the buffer zone must be inspected directly at a resolution of 0.01 km^2 which means an inspection intensity of 17,700 cells. For both alternative designs, the inspection intensity at a 0.01 km^2 grid resolution was lower (2,225 cells for the three-phase strategy and 7,100 cells for the two-phase strategy), with the three-phase alternative having the lowest value due to the intermediate 0.25 km^2 grid resolution.

Sampling intensity

For the three-phase design, the optimum sampling intensity was calculated in 51, 45 and 14 samples/cell for $j = \{1, 0.25, 0.01\}$ km^2 grid resolutions. For the two-phase design, it was calculated in 51 and 15 samples/cell for $j = \{1, 0.01\}$ km^2 grids (Table 4). These optimum values (under R_j=50) ensured the detection of all the positive cells (the $C^*_{j,+}$ values matched the $C_{j,+}$ values in Table 1). Sampling intensity was reduced by increasing the grid resolution but the translation of these estimates per hectare (ha) e.g., for the three-phase design, means that 0.5 samples/ha (1 sample/2ha), 2 samples/ha and 14 samples/ha are required for 1, 0.25, 0.01 km^2 grid resolutions, respectively.

Survey effort

The overall efficiency was compared among the sequential adaptive approaches and the official

Table 3 Inspection intensity for the three-phase, two-phase, and the official delimiting survey strategies in the demarcated area for *Xylella fastidiosa* in Alicante, Spain

Delimiting strategy		Grid resolution (km^2)	C_j
Three-phase		$j = 1$	833
		$j = 0.25$	284
		$j = 0.01$	2,225
Two-phase		$j = 1$	833
		$j = 0.25$	–
		$j = 0.01$	7,100
Official		$j = 1$	656
		$j = 0.25$	–
		$j = 0.01$	17,700

C_j denotes the inspection intensity as the number of cells to be inspected for each grid resolution $j = \{1, 0.25, 0.01\}$ km^2

strategy assessing the survey efforts (Table 5). Both sequential adaptive designs substantially reduced the inspection intensity for the 0.01 km^2 grid resolution. This result strongly influenced the estimation of the survey efforts, which were 86,413, 148,983 and 298,956 for the three-phase, two-phase and the official delimiting strategies, respectively. Compared with the official strategy, the sequential adaptive strategies reduced the survey effort by 71.10% in the case of the two-phase design and 50.17% in the case of the three-phase strategy.

Modeling the spatial distribution of *Xf* incidence

Bayesian hierarchical spatial model

The model selected to estimate the spatial distribution of *Xf* incidence was:

$$Y_i \sim \text{Binomial } (m_i, \pi_i) \quad i = 1, \ldots, n$$
$$\text{logit } (\pi_i) = \beta_0 + v_i, \tag{2}$$

in which only the spatial effect is considered (see Table S1 in Supplementary material for model ranking based on WAIC and LCPO values). A numeric descriptive of its corresponding parameter estimates is shown in Table 6. The full model [Eq. (1)] is also

Table 4 Optimum sampling intensity (n_j) summarized by the median value for each grid resolution $j = \{1, 0.25, 0.01\}$ in the three-phase and two-phase sequential adaptive survey designs applied to the demarcated area for *Xylella fastidiosa* in Alicante, Spain

	Grid resolution (km²)	R_1	$R_{0.25}$	$R_{0.01}$	$C_{j,+}^*$	n_j
Three-phase	$j = 1$	**50**	–	–	71	**51**
		25	–	–	70	46
		15	–	–	69	40
		5	–	–	68	36
	$j = 0.25$	**50**	**50**	–	89	**45**
			25	–	89	41
			15	–	88	37
			5	–	87	33
	$j = 0.01$	**50**	**50**	**50**	161	**14**
				25	160	13
				15	160	13
				5	160	12
Two-phase	$j = 1$	**50**	–	–	71	**51**
		25	–	–	70	46
		15	–	–	69	40
		5	–	–	68	36
	$j = 0.01$	**50**	–	**50**	161	**15**
			–	25	160	13
			–	15	159	12
			–	5	159	12

R_j denotes the condition established in the optimization algorithm for each grid resolution and $C_{j,+}^*$ is the number of positive cells for *Xf* identified by the algorithm under the constraint imposed

R_j=50 ensures that 50% of the replicates generated had identified all the positive cells for *Xf* for the j grid resolution. Given the sequential adaptive nature of the strategy, the condition R_j established for a particular phase (j) affected the subsequent one

In bold selected stopping rule ($R_j = 50$) for each j resolution and its corresponding optimum sampling intensity outputs (n_j)

described to illustrate why climatic variables were finally not included. Mean values of the estimates of the climatic covariates were very close to zero and showed probabilities of being greater than zero of around 0.50, thus having low explanatory capacity for incidence estimates.

The mean values of the estimates associated with the spatial effect and the incidence (0–1) as well as the standard deviation assessed for the incidence are described graphically per cell as displayed in Fig. 4. The estimate of the spatial effect was determined by the spatial dependence structure defined by a neighborhood relationship of a distance of 2.5 km among cells. Incidence estimates varied from 0.002 to 0.215 and the standard deviation from 0.003 to 0.191 (Fig. 4). The spatial effect ranged from -1.780 to 4.371, with positive values indicating higher *Xf* incidence estimates (Fig. 4). The cells with the highest

incidence and spatial effect values were concentrated in the central and eastern parts of the demarcated area. Those cells also reported the highest standard deviation values.

Effect of sampling intensity on Xf incidence estimates

The effect of the sampling intensity on *Xf* incidence estimates (Eq. 2) was assessed by generating different data subsets from the reference database by limiting the sampling intensity with the following threshold values: 9 (DS_9), 23 (DS_{23}), 37 (DS_{37}), and 51 (DS_{51}) samples/cell for the 1 km² grid resolution. Note that 51 samples/cell corresponded to the optimum sampling intensity estimated for the 1 km² grid resolution by the optimization algorithm (Sect. "Sequential adaptive delimiting survey strategy"). The value of 9 samples/cell was consistent with the third quartile value of

Table 5 Survey efforts in the three-phase and two-phase sequential adaptive designs and the official strategy in the demarcated area for *Xylella fastidiosa* in Alicante, Spain

	Grid resolution (km^2)	C_j	n_j	N_j
Three-phase	$j = 1$	833	51	42,483
	$j = 0.25$	284	45	12,780
	$j = 0.01$	2,225	14	31,150
				86,413 (28.90%)[a]
Two-phase	$j = 1$	833	51	42,483
	$j = 0.01$	7,100	15	106,500
				148,983 (49.83%)[a]
Official	$j = 1$	656	51	33,456
	$j = 0.01$	17,700	15	265,500
				298,956 (100%)[a]

For each grid resolution j with $j = \{1, 0.25, 0.01\}$, C_j denotes the inspection intensity (number of cells to be inspected), n_j is the sampling intensity (number samples/cell), and $N_j = C_j \times n_j$ is the sampling effort in terms of the total number of samples to be taken. $\sum_j N_j$ (in bold) denotes total survey effort

C_j values for $j = \{0.25, 0.01\}$ are the outputs obtained by the optimization algorithm for the sequential adaptive designs. C_j values for $j = \{1, 0.01\}$ are the thresholds indicated for the official strategy

n_j values for $j = \{1, 0.25, 0.01\}$ are the outputs obtained by the optimization algorithm for the sequential adaptive designs. n_j values for $j = \{1, 0.01\}$ are the outputs obtained for the two-phase design. In the official strategy, sampling is subject to the observation of symptoms and thresholds are established in that regard

[a] Percentage relative to the survey effort of the official strategy

Table 6 Marginal posterior distributions of parameters and hyperparameters for the model of *Xylella fastidiosa* incidence distribution in the demarcated area in Alicante, Spain, with mean, standard deviation (**SD**), median (**$Q_{0.5}$**), and 95% credible interval (**95% CI**)

	Mean	SD	$Q_{0.5}$	95% CI	$P(\cdot) > 0$
$\beta_0 + v$					
β_0	− 5.829	0.325	− 5.807	[− 6.526,− 5.253]	0
σ_v	1.762	0.202	1.748	[1.401,2.194]	–
$\beta_0 + bio1 + bio7 + bio13 + v$					
β_0	− 7.273	13.897	− 7.205	[− 34.859,19.906]	0.298
$bio1$	0.060	0.312	0.055	[− 0.540,0.687]	0.570
$bio7$	0.040	0.343	0.040	[− 0.634,0.717]	0.545
$bio13$	− 0.008	0.100	− 0.005	[− 0.214,0.181]	0.477
σ_v	1.798	0.208	1.791	[1.653,2.228]	–

The full model including climatic covariates is also indicated

*bio*1: annual mean temperature (°C)

*bio*7: annual range temperature (°C)

*bio*13: precipitation of the wettest month (mm)

v: spatial effect

the reference database sampling intensity value. The values of 23 and 37 samples/cell were chosen arbitrarily in the range between 9 and 51.

The different data subsets implied a reduction in the number of samples compared to the reference database. Furthermore, subsetting also implied a change in the distribution of samples and overall
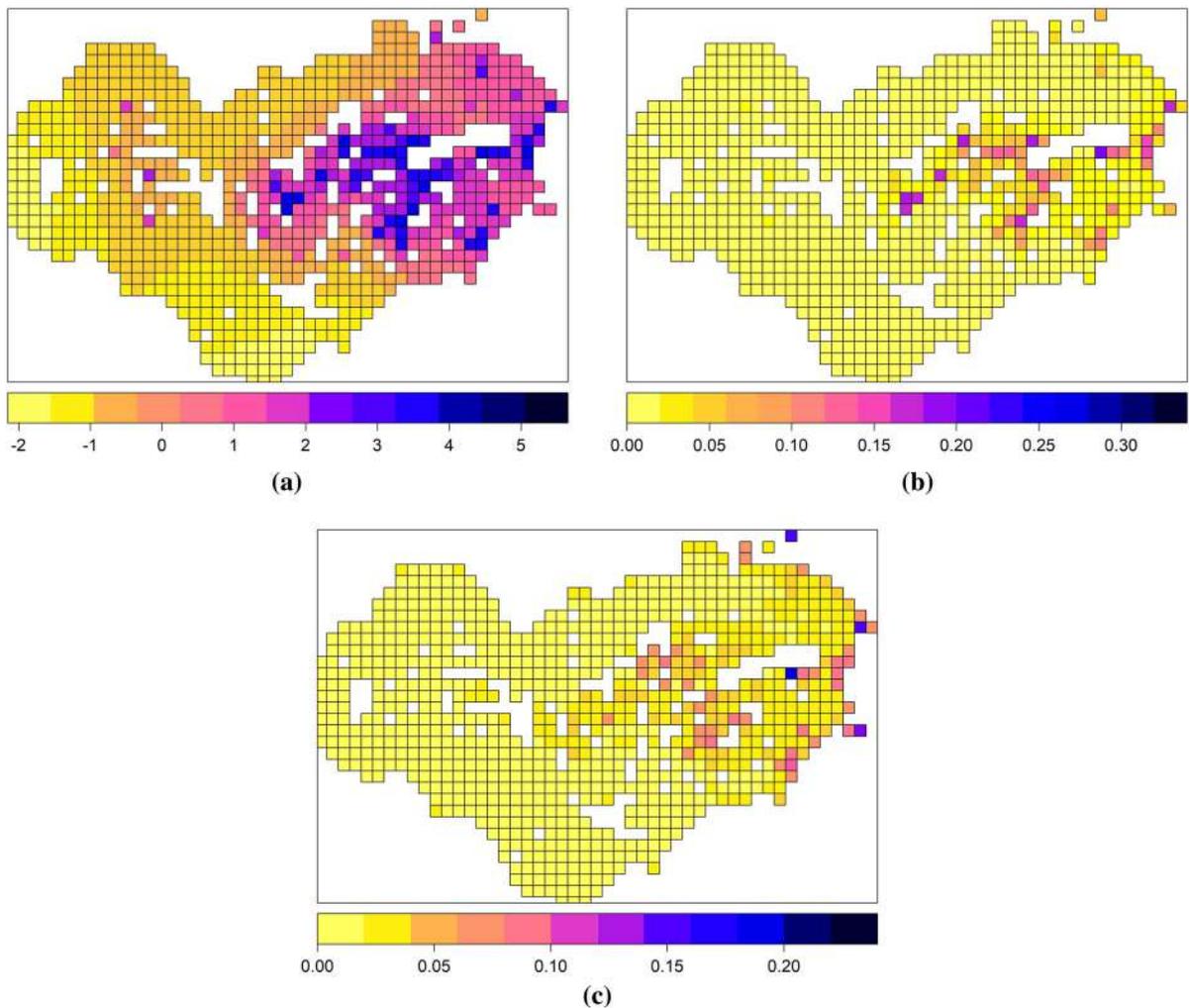
Fig. 4 Geographical representation of the model of *Xylella fastidiosa* incidence distribution in the demarcated area in Alicante, Spain, with **a** the mean value of the spatial effect, and **b** the mean value of the incidence (0–1), and **c** the standard deviation of the incidence

incidence (Table S2). Figures S1 and S2 display the changes in the distribution of the sampling intensity (samples/cell), number of positive samples per cell, cells with *Xf* presence, and incidence per cell. Note that the quantities obtained with the different data subsets were summarized by the replicate showing the median behavior in relation to the total number of positive samples.

The bias, SE, and SD were the discrepancy measures estimated to assess changes in model parameter estimates (Table S3). In relation to the β_0 estimate stability, bias (absolute values), SD, and SE showed the highest values in DS_9, which was the most restrictive database (Bias=0.364, SE=0.476 and

SD=0.138). Regarding the stability of the σ_v estimates, the bias did not show a clear trend and all data subsets had similar values. The SE and SD also exhibited the highest values associated with DS_9 (SE=0.281 and SD=0.181). In general, DS_{37} and DS_{51} presented the most stable estimates.

Changes in the estimates of *Xf* incidence and the spatial effect were assessed graphically (Fig. 5) by comparing the inference outcomes related to the replicas (averaged) in relation to the reference outputs (Fig. 4). In general cells in the central and eastern parts of the demarcated area, with the highest sampling intensities (Fig. S1) exhibited higher bias above all for those data subsets in which more restrictive sampling
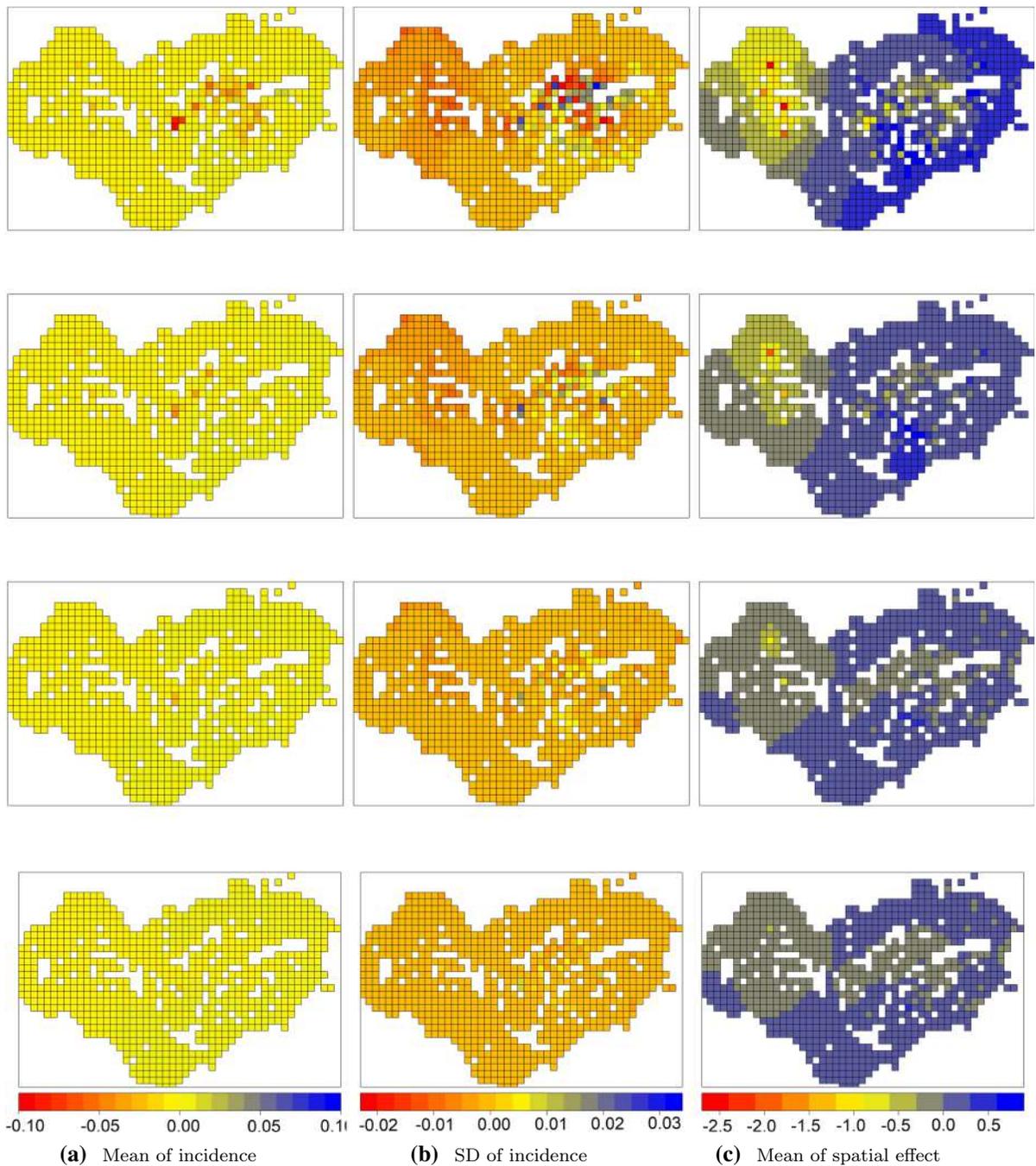
Fig. 5 Bias for the mean value and standard deviation (SD) of the incidence and mean of the spatial effect for the data subsets DS_9 (row 1), DS_{23} (row 2), DS_{37} (row 3), DS_{51} (row 4) relative to the model fitted to the reference database of the demarcated area for *Xylella fastidiosa* in Alicante, Spain

intensity conditions were applied. The data subsets DS_{37} and DS_{51} exhibited the most robust inferences.

Discussion

We have presented a generic sequential adaptive survey strategy operationally deployed in the

"demarcated area" of Xf in Alicante. Our results show that sequencing and adapting inspection and sampling to increasing spatial resolutions allows accurate delimitation of the infested zones while reducing the overall survey efforts, thus improving the efficiency of the surveillance program. On the one hand, with the sequential adaptive approach, the aggregated behaviour of the disease is exploited to enforce the survey efforts towards sites in which the pathogen is more likely to be found, in line with the proposals of Gottwald et al. (2014), Parnell et al. (2017) and EFSA (2020a, b). On the other hand, we ensure that part of them are distributed in such a way as to cover all the survey area in order to avoid deploying all the resources in the areas adjacent to the current outbreaks.

The sequential adaptive strategy and the optimization algorithm developed in our study can be used to improve the surveillance programs for other plant pests and allow it to be transferred to other geographical areas. Nevertheless, the operating conditions set out for the optimization algorithm such as survey resolution, sampling scheme and stopping criteria should be adjusted to the extent and the specific epidemiological characteristics of the survey area, including among other factors the distribution and abundance of host plants and vectors when the latter play a relevant role in the epidemiology of the pathogen. One important aspect of our work is that sampling intensity was estimated by means of an optimization-based principle that maximizes Xf detection for each spatial resolution. The vector distribution has not been considered in our proposal because the existing trapping methods have a relatively low efficiency and it is difficult to link the finding of an infected vector to a particular infected plant in the survey area (EFSA 2020b).

The sequential adaptive strategy resulted in a substantial reduction in the inspection intensities compared to the official one. This was noted particularly at the finest spatial resolution ($0.01~km^2$), in which 2,225 and 7,100 cells were estimated to be inspected for the three-phase/two-phase designs, respectively, compared to the 17,700 cells of the official strategy. Moreover, given its sequential scheme, it can lead to an optimization of the workflow both in the field and in the laboratory. The official strategy implied carrying out parallel inspections in two predefined surveillance areas each with a different spatial resolution: i) the first kilometer radius of the buffer zone using a grid resolution of $0.01~km^2$ cells, and ii) the outer buffer zone up to a radius of 5 km using a grid resolution of $1~km^2$.

Conversely, our strategy defines an increasing spatial resolution in the whole demarcated area, from $1~km^2$ up to $0.01~km^2$, based on the information obtained in the previous inspection/sampling phase. That is, it allows delimitation of the spatial extent of the pathogen at coarser spatial resolutions while delimiting infested areas for implementing control measures at finer resolutions. Eradication measures previously established in the Decision (EU 2015a) and now enforced by the new Implementing Regulation (EU) 2020/1201 (EU 2020) imply the removal of all host plants in the infested zones and the application of measures for vector control, in some cases also including insecticides. Therefore, increasing the accuracy in delimiting the infested zones will allow for a more targeted disease control, thereby minimizing the socioeconomic and environmental impacts associated with the removal of plants and the use of chemical insecticides (EFSA 2019b).

Optimum sampling intensity was estimated at 51, 45, 14 samples/cell for 1, 0.25 and $0.01~km^2$ grid resolutions in the three-phase and at 51, 15 samples/cell for 1 and $0.01~km^2$ grid resolutions in the two-phase design. These results contribute to the ongoing work aimed at developing risk-based survey approaches to quarantine pests and pathogens (EU 2019b, 2020) in which statistically based sample size calculations are specified (EFSA 2020a, b). Specifically, the new Implementing Regulation (EU) 2020/1201 (EU 2020) for Xf has recently enforced the use of statistically sound risk-based surveys (EFSA 2020a), where a different confidence level and design prevalence are assumed in the buffer zone. In the specific case of Xf, the Decision only considered statistically-based sampling within the infested zone itself, whereas sampling in the buffer zone was based solely on the observation of symptoms. Considering the long incubation period that characterizes Xf, this approach based on symptoms may negatively affect the accuracy of the whole surveillance program.

The modeling outputs evidence the known spatial aggregation of the disease in the demarcated area in

Alicante (Cendoya et al. 2020) and, more importantly, the adequacy of the proposed sequential adaptive survey strategy. The model selection process also highlighted the marginal influence of climatic covariates in that regard in the study area. In contrast to work conducted in other areas (Godefroid et al. 2019; Martinetti and Soubeyrand 2019), the low explanatory power of climatic covariates prevents us from establishing a relationship with the distribution of *Xf* in the study area. Nevertheless, *Xf* subsp. *multiplex* is more widely distributed in the EU territory than subspecies *fastidiosa* and *pauca*, and it is known to have suitable climatic conditions in the vast majority of the territory (EFSA 2019b). As this may be the general situation in most outbreaks, the limited extent of the study area might also have played an important role in this respect. However, both the spatial aggregation and the scant relevance of the climate should be considered in the planning of future surveillance actions.

Additionally, the optimum sampling intensity calculated for a resolution of 1 km^2 was then evaluated by assessing its effect on the estimate of *Xf* incidence by incorporating into our study the dual focus on applying advanced survey designs and modeling techniques to maximize the quality of the information collected and the rigor of inference (Johnson et al. 2013; Pacifici et al. 2016). Thus, the estimated sampling intensity values could be used as a benchmark to be explored in other *Xf* outbreaks and compared with other inspection/sampling size calculation methods.

The estimations of the survey efforts as the total number of samples to be collected and tested were 86,413, 148,983 and 298,956 for the three-phase, two-phase, and official delimiting strategies, respectively. Our study does not translate survey efforts into economic terms due to the complexity of estimating the overall costs. It is, however, possible to estimate unit cost per sample associated to the laboratory testing, whereas the inspection and sampling costs are more difficult to calculate, as they depend more on logistic planning. Nevertheless, the reduction in the number of samples estimated for both sequential adaptive designs in relation to the official strategy, 71.10% and 50.17%, will certainly have a strong impact on the survey costs and consequently on the feasibility of the implementation of the surveillance program.

Planning delimiting surveys entails reaching a compromise between the available resources and the extent of the area to be covered. After detecting an outbreak, the actual distribution of the disease is often unknown by plant health authorities, which further reduces the efficacy of control efforts, such as eradication or containment. Our methodology addresses a key question that usually arises when planning any surveillance program: "What are the best locations to deploy our inspection/sampling resources?". Our generic delimiting survey strategy deals with these challenges and demonstrates that sequencing and adapting inspection and sampling to different spatial resolutions allows identification of the optimal inspection/sampling sites for an accurate as well as efficient delimitation of the infested zones.

In sum, our sequential adaptive survey strategy was able to efficiently delimit the distribution of *Xf* in the demarcated area in Alicante. This survey strategy may assist plant health authorities in optimizing survey resources and implementing more targeted disease control and regulatory actions. A more focused allocation of management efforts could improve the efficiency and efficacy of control programs, thus reducing the associated side effects.

Declarations

following the procedures established by the current regulation in the EU.

References

Arias-Giraldo LF, Giampetruzzi A, Metsis M, Marco-Noales E, Imperial J, Velasco-Amo MP, Román-Écija M, Landa BB (2020) Complete circularized genome data of two Spanish strains of *Xylella fastidiosa* (IVIA5235 and IVIA5901) using hybrid assembly approaches. Phytopathology 110(5):969–972

Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data. Chapman and Hall/CRC, London

Bodino N, Cavalieri V, Dongiovanni C, Simonetto A, Saladini MA, Plazio E, Gilioli G, Molinatto G, Saponari M, Bosco D (2020) Dispersal of *Philaenus spumarius* (Hemiptera: Aphrophoridae), a vector of *Xylella fastidiosa*, in Olive Grove and Meadow agroecosystems. Environ Entomol. https://doi.org/10.1093/ee/nvaa140

Brown JA, Salehi M, Moradi M, Panahbehagh B, Smith DR (2013) Adaptive survey designs for sampling rare and clustered populations. Math Comput Simul 93:108–116

Büyüktahtakın IE, Haight RG (2018) A review of operations research models in invasive species management: state of the art, challenges, and future directions. Ann Oper Res 271(2):357–403

Cendoya M, Martínez-Minaya J, Dalmau V, Ferrer A, Saponari M, Conesa D, López-Quílez A, Vicent A (2020) Spatial Bayesian modeling applied to the surveys of *Xylella fastidiosa* in Alicante (Spain) and Apulia (Italy). Front Plant Sci 11:1204

Chaudhuri A, Stenger H (2005) Survey sampling: theory and methods. CRC Press, London

Cornara D, Cavalieri V, Dongiovanni C, Altamura G, Palmisano F, Bosco D, Porcelli F, Almeida RPP, Saponari M (2017) Transmission of *Xylella fastidiosa* by naturally infected *Philaenus spumarius* (Hemiptera, Aphrophoridae) to different host plants. J Appl Entomol 141(1–2):80–87

DG-SANTE (2018) Final report of an audit carried out in Spain from 27 February 2018 to 9 March 2018 in order to evaluate the situation and official controls for *Xylella fastidiosa*. https://ec.europa.eu/food/audits-analysis/audit_reports/details.cfm?rep_id=4056

DG-SANTE (2019) Final report of an audit carried out in Spain from 1 April 2019 to 9 April 2019 in order to evaluate the situation and official controls for *Xylella fastidiosa*. https://ec.europa.eu/food/audits-analysis/audit_reports/details.cfm?rep_id=4177

Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger JJ, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD et al (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30(5):609–628

Edwards TC, Cutler DR, Zimmermann NE, Geiser L, Alegria J (2005) Model-based stratifications for enhancing the detection of rare ecological events. Ecology 86(5):1081–1090

EFSA (2018) Updated pest categorisation of *Xylella fastidiosa*. EFSA J 16(7)

EFSA (2019a) Pest survey card on *Xylella fastidiosa*. EFSA J 16(6):1667E

EFSA (2019b) Update of the scientific opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory. EFSA J 17(5):5665

EFSA (2020a) General guidelines for statistically sound and risk-based surveys of plant pests. EFSA J 17(9):1919E

EFSA (2020b) Guidelines for statistically sound and risk-based surveys of *Xylella fastidiosa*. EFSA J 17(6):1873E

EFSA (2020c) Scientific report on the update of the *Xylella* spp. host plant database–systematic literature search up to 30 june 2019. EFSA J 18(4):6114

Epanchin-Niell RS, Liebhold AM (2015) Benefits of invasion prevention: effect of time lags, spread rates, and damage persistence. Ecol Econ 116:146–153

EPPO (2019) Eppo global database: *Xylella fastidiosa* (XYLEFA). https://gd.eppo.int/taxon/XYLEFA/distribution. Accessed 22 July 2020

EU (2015) Commission Implementing Decision (EU) 2015/2417 of 1 December 2015 amending Implementing Decision (EU) 2015/789 as regards measures to prevent the introduction into and the spread within the Union of *Xylella fastidiosa* (Wells et al.). OJ L333:143–147

EU (2015) Commission Implementing Decision (EU) 2015/789 of 18 May 2015 as regards measures to prevent the introduction into and the spread within the Union of *Xylella fastidiosa* (Wells et al.). OJ L125:36–53

EU (2016a) Commission Implementing Decision (EU) 2016/764 of 12 May 2016 amending Implementing Decision (EU) 2015/789 as regards measures to prevent the introduction into and the spread within the Union of *Xylella fastidiosa* (Wells et al.). OJ L126:77–84

EU (2016b) Regulation (EU) 2016/2031 of the European Parliament of the Council of 26 October 2016 on protective measures against pests of plants, amending Regulations (EU) 228/2013,(EU) 652/2014 and (EU) 1143/2014 and repealing Council Directives 69/464/EEC, 74/647/EEC, 93/85/EEC, 98/57/EC, 2000/29/EC, 2006/91/EC and 2007/33/EC. OJ L317:4–104

EU (2017) Commission Implementing Decision (EU) 2017/2352 of 14 December 2017 amending Implementing Decision (EU) 2015/789 as regards measures to prevent the introduction into and the spread within the Union of *Xylella fastidiosa* (Wells et al.). OJ L336:31–44

EU (2018a) Commission Implementing Decision (EU) 2018/1511 of 9 October 2018 amending Implementing Decision (EU) 2015/789 as regards measures to prevent the introduction into and the spread within the Union of *Xylella fastidiosa* (Wells et al.). OJ L255:16–17

EU (2018b) Commission Implementing Decision (EU) 2018/927 of 27 June 2018 amending Implementing Decision (EU) 2015/789 as regards measures to prevent the introduction into and the spread within the Union of *Xylella fastidiosa* (Wells et al.). OJ L164:49–50

EU (2019a) Commission Delegated Regulation (EU) 2019/1702 of 1 August 2019 supplementing Regulation (EU) 2016/2031 of the European Parliament and of the Council by establishing the list of priority pests. OJ L260:8–10

EU (2019b) Commission Implementing Regulation (EU) 2019/2072 of 28 November 2019 establishing uniform conditions for the implementation of Regulation (EU) 2016/2031 of the European Parliament and the Council, as regards protective measures against pests of plants, and repealing Commission Regulation (EC) No 690/2008 and amending Commission Implementing Regulation (EU) 2018/2019. OJ L319:1–279

EU (2020) Commission Implementing Regulation (EU) 2020/1201 of 14 August 2020 as regards measures to prevent the introduction into and the spread within the Union of *Xylella fastidiosa* (Wells et al.). OJ L269:2–39

Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol 37(12):4302–4315

Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. Stat Comput 24(6):997–1016

Godefroid M, Cruaud A, Streito JC, Rasplus JY, Rossi JP (2019) *Xylella fastidiosa*: climate suitability of European continent. Sci Rep 9(1):8844

Gottwald T, Luo W, McRoberts N (2014) Risk-based residential HLB/ACP survey for California, Texas and Arizona. J Citrus Pathol 1(1):1–11

GVA (2020) Situación de *Xylella fastidiosa* en la Comunitat Valenciana. Servicio de Sanidad Vegetal, DG Agricultura, Ganadería y Pesca, Generalitat Valenciana. http://www.agroambient.gva.es/documents/163214705/163847802/20200608+Situacion+de+Xf+en+la+C+Valenciana+web.pdf/8632e264-3147-457f-9f4e-9c19c09062e6. Accessed 8 June 2020

Hauser CE, Giljohann KM, Rigby M, Herbert K, Curran I, Pascoe C, Williams NSG, Cousens RD, Moore JL (2016) Practicable methods for delimiting a plant invasion. Divers Distrib 22(2):136–147

Hijmans RJ, Etten JV, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, Perpinán O, Bevan A et al (2015) Package "raster". R package

Johnson DS, Conn PB, Hooten MB, Ray JC, Pond BA (2013) Spatial occupancy models for large data sets. Ecology 94(4):801–808

Landa BB, Castillo AI, Giampetruzzi A, Kahn A, Román-Écija M, Velasco-Amo MP, Navas-Cortés JA, Marco-Noales E, Barbé S, Moralejo E, Coletta-Filho HD, Saldarelli P, Saponari M, Almeida RPP (2020) Emergence of a plant pathogen in Europe associated with multiple intercontinental introductions. Appl Environ Microbiol 86(3):e01521-19

Latimer A, Andrew M, Wu S, Gelfand AE Jr, Silander JA (2006) Building statistical models to analyze species distributions. Ecol Appl 16(1):33–50

Madden LV, Hughes G, van den Bosch F (2007) The study of plant disease epidemics. APS Press, St. Paul

Martinetti D, Soubeyrand S (2019) Identifying lookouts for epidemio-surveillance: application to the emergence of *Xylella fastidiosa* in France. Phytopathology 109(2):265–276

Martínez-Minaya J, Cameletti M, Conesa D, Pennino MG (2018) Species distribution modeling: a statistical review with focus in spatio-temporal issues. Stoch Environ Res Risk A 32(11):3227–3244

Mastin AJ, Gottwald TR, van den Bosch F, Cunniffe NJ, Parnell S (2020) Optimising risk-based surveillance for early detection of invasive plant pathogens. PLoS Biol 18(10):e3000863

Moore AL, McCarthy MA (2016) Optimizing ecological survey effort over space and time. Methods Ecol Evol 7(8):891–899

Pacifici K, Reich BJ, Dorazio RM, Conroy MJ (2016) Occupancy estimation for rare species using a spatially-adaptive sampling design. Methods Ecol Evol 7(3):285–293

Parnell S, van den Bosch F, Gottwald T, Gilligan CA (2017) Surveillance to inform control of emerging plant diseases: an epidemiological perspective. Annu Rev Phytopathol 55:591–610

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J Roy Stat Soc B Met 71(2):319–392

Saponari M, Loconsole G, Cornara D, Yokomi RK, Stradis AD, Boscia D, Bosco D, Martelli GP, Krugner R, Porcelli F (2014) Infectivity and transmission of *Xylella fastidiosa* by *Philaenus spumarius* (Hemiptera: Aphrophoridae) in Apulia, Italy. J Econ Entomol 107(4):1316–1319

Saponari M, D'Attoma G, Kubaa RA, Loconsole G, Altamura G, Zicca S, Rizzo D, Boscia D (2019) A new variant of *Xylella fastidiosa* subspecies *multiplex* detected in different host plants in the recently emerged outbreak in the region of Tuscany, Italy. Eur J Plant Pathol 154:1195–1200

Schaad NW, Postnikova E, Lacy G, Fatmi M, Chang CJ (2004) *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp *piercei*, subsp. nov., *X. fastidiosa* subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov. Syst Appl Microbiol 27(3):290–300

Schneider K, Van der Werf W, Cendoya M, Mourits M, Navas-Cortés JA, Vicent A, Lansink AO (2020) Impact of *Xylella fastidiosa* subspecies *pauca* in European olives. Proc Natl Acad Sci USA 117(17):9250–9259

Vicent A, Blasco J (2017) When prevention fails. Towards more efficient strategies for plant disease eradication. New Phytol 214(3):905–908

Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 11(Dec):3571–3594

Wells JM, Raju BC, Hung HY, Weisburg WG, Mandelco-Paul L, Brenner DJ (1987) *Xylella fastidiosa* gen. nov., sp. nov:

Gram-negative, xylem limited, fastidious plant bacteria related to *Xanthomonas* spp. Int J Syst Bacteriol 37:136–143

White SM, Bullock JM, Hooftman DAP, Chapman DS (2017) Modelling the spread and control of *Xylella fastidiosa* in the early stages of invasion in Apulia, Italy. Biol Invasions 19(6):1825–1837

White SM, Bullock JM, Cavers S, Chapman DS (2019) Using modelling to investigate the effectiveness of national surveillance monitoring aimed at detecting a *Xylella fastidiosa* outbreak in Scotland. Plant Health Centre, Invergowrie, p 22

Yemshanov D, Haight RG, Koch FH, Lu B, Venette R, Fournier RE, Turgeon JJ (2017) Robust surveillance and control of invasive species using a scenario optimization approach. Ecol Econ 133:86–98

Yemshanov D, Haight RG, Koch FH, Venette R, Studens K, Fournier RE, Swystun T, Turgeon JJ (2017) A safety rule approach to surveillance and eradication of biological invasions. PLoS ONE 12(7):e0181482