UNIVERSITAT DE VALÈNCIA

Programa de Doctorat en Estadística i Optimització

Spatial models for the analysis of plant disease epidemics and their control

by

Martina Cendoya Martínez

PhD Thesis
in Statistics and Optimization

Supervised by
David Valentín Conesa Guillén,
and Antonio Vicent Civera

in the
Faculty of Mathematics
Department of Statistics and Operations Research

November 2023

Agradecimientos

Esta Tesis Doctoral representa el fin de una etapa, que en ocasiones ha sido difícil, pero también gratificante, y que no habría sido posible sin el apoyo y la colaboración de todas las personas que han estado presentes tanto en lo profesional como en lo personal durante este proceso. Por ello, quiero dedicar esta sección a todas ellas.

Quiero empezar destacando el papel fundamental de mis directores de tesis, Antonio Vicent (Toni) y David Conesa, que más allá de la supervisión académica, han sido los guías en este camino. Agradezco la confianza depositada en mí al ofrecerme esta oportunidad, así como por haberme empujado a enfrentar nuevos retos que, de otra manera, ni siquiera me hubiese planteado. Sus consejos y paciencia han sido fundamentales en el desarrollo de este trabajo y en mi crecimiento profesional y personal.

También quiero agradecer a Joaquín Martínez Minaya su infinita ayuda, especialmente al comienzo de este camino. Aprecio enormemente su generosidad al estar siempre dispuesto a ofrecer su ayuda y por sus valiosos consejos.

A Elena Lázaro, con quien he compartido buenos momentos en el IVIA. Agradezco su apoyo diario, que ha hecho más llevadero cada día. Con el inicio de una nueva etapa para ambas, le deseo todo lo mejor.

Asimismo, quiero reconocer la contribución de Antonio López, quien ha participado en gran parte de los trabajos que se presentan, por sus buenos consejos y la experiencia científica aportada.

También quiero agradecer a Chiara Rosace, con la que he formado un gran equipo, por transmitir su positividad y alegría. Espero que continuemos aprendiendo la una de la otra en los futuros retos que nos esperan.

En estas líneas quiero hacer mención especial a la Sociedad Española de Fitopatología (SEF) y a la Sociedad Española de Bioestadística (SEB), porque me han brindado la oportunidad de conocer a grandes profesionales y de

difundir mis aportaciones en un entorno acogedor. Igualmente, quiero destacar las Jornadas de Estudiantes de la SEB con las que tanto he aprendido y me han permitido compartir experiencias con multitud de compañeros/as.

Quiero agradecer a mi familia, por ser la constante entre tantos cambios. En especial, a mi madre, Tere, a mi padre, Jose María, a Ramón y a Pilar, por escucharme, apoyarme y darme todas las facilidades posibles que han hecho que pueda llegar hasta aquí.

Finalmente, quiero dedicar un agradecimiento especial a Fernando, por recordarme que puedo conseguir lo que me proponga. Su apoyo y actitud positiva ante las dificultades que han surgido, especialmente en el último periodo, me han dado la fuerza necesaria para continuar avanzando.

A efectos de elaboración de esta Tesis, me gustaría destacar la financiación recibida por el Instituto Valenciano de Investigaciones Agrarias (IVIA) mediante una beca parcialmente financiada por el Fondo Social Europeo (FSE). Asimismo, agradezco al Servicio de Sanidad Vegetal haber puesto a nuestra disposición los datos para poder llevar a cabo los estudios que se presentan en esta Tesis.

"Although I am sure there are some who feel that the emphasis on mathematics and statistics obscures the understanding of the biology of epidemics, I would make the opposite claim, and declare emphatically that mathematical and statistical modelling are foundations for understanding epidemics."

Laurence V. Madden

UNIVERSITAT DE VALÈNCIA

Resumen

Facultad de Ciencias Matemáticas
Departamento de Estadística e Investigación Operativa

Programa de Doctorado en Estadística y Optimización

El crecimiento exponencial del comercio global y de los viajes internacionales en las últimas décadas ha tenido un impacto importante en la introducción y propagación de plagas y patógenos de plantas en todo el mundo. Además, el cambio climático ha intensificado esta problemática propiciando condiciones climáticas favorables para el establecimiento de estas plagas y patógenos en nuevas áreas geográficas. Como resultado, las plagas y patógenos emergentes representan una amenaza creciente para la seguridad alimentaria y la estabilidad socioeconómica a escala global. Esta problemática plantea un desafío importante para la agricultura, especialmente en un escenario en el que es necesario incrementar la producción de alimentos para satisfacer las demandas de una población mundial en aumento. Por lo tanto, comprender los factores que intervienen en la entrada, establecimiento y dispersión de los patógenos vegetales, así como analizar la dinámica de las enfermedades que causan, es crucial para prevenir y gestionar eficazmente nuevos brotes y minimizar su impacto.

En la última década, la bacteria fitopatógena *Xylella fastidiosa* ha provocado una crisis fitosanitaria en la Unión Europea, afectando, entre otros, a dos de los cultivos mediterráneos más importantes, olivo y almendro. Una de las complejidades en el estudio de este patógeno es su amplia diversidad genética y su capacidad para infectar a un amplio rango de vegetales hospedantes. A pesar de su presencia durante décadas en el continente Americano y los numerosos estudios realizados al respecto, aún sigue habiendo muchas incógnitas en torno a su epidemiología, así como en cuanto a las estrategias de vigilancia, control y formas de prevención. Este patógeno ha sido el caso de estudio principal de esta Tesis, en la cual abordamos diversas cuestiones que han ido surgiendo de forma dinámica a lo largo de su estudio. Es importante destacar que durante el transcurso de la misma, se han ido

incorporando tanto los datos disponibles en cada momento sobre la distribución del patógeno como los nuevos conocimientos proporcionados por los avances científicos.

El objetivo general de esta Tesis es proporcionar distintas herramientas de modelización espacial aplicables al estudio de epidemias de enfermedades de plantas, que permitan analizar su distribución espacial y los factores relacionados, así como servir de referencia para su control. En concreto abordamos diversos aspectos sobre *X. fastidiosa* mediante métodos estadísticos avanzados novedosos en el ámbito de la fitopatología, con especial atención en el análisis de su distribución espacial. En este contexto, se plantean los siguientes objetivos principales: i) el análisis de los efectos climáticos y espaciales sobre la distribución de *X. fastidiosa*; ii) analizar el efecto de la incorporación de barreras de dispersión sobre la distribución del patógeno; iii) la implementación de un modelo de dispersión espacialmente explícito basado en individuos; y iv) la evaluación de la eficacia y eficiencia de distintos planes de gestión de brotes incluyendo estrategias de vigilancia y control.

La modelización espacial se enfoca desde la perspectiva de los modelos espaciales jerárquicos bayesianos, particularmente usando la aproximación de Laplace anidada integrada (INLA) para realizar inferencia y predicción. Por medio de esta metodología se modelizan distintos tipos de datos espaciales, en concreto datos de área y datos geoestadísticos, adaptando los modelos a cada uno de ellos. Como extensión de estos modelos, se emplea un modelo no estacionario para la incorporación de barreras de dispersión. Por otra parte, se propone un modelo de simulación de la dispersión de la enfermedad basado en individuos, donde una función de correlación establece su relación espacial. Este modelo permite establecer la base para la implementación y evaluación de las estrategias de vigilancia y control.

En cuanto a los resultados y conclusiones más relevantes de esta Tesis, los modelos espaciales jerárquicos bayesianos son eficaces para abordar los diferentes tipos de datos espaciales de los brotes del patógeno. Destaca el fuerte efecto de la componente espacial sobre la distribución del patógeno en los brotes analizados, a diferencia de la escasa relevancia de las variables climáticas, poniendo en evidencia la importancia de la consideración de la estructura espacial en la modelización. La inclusión de barreras de dispersión mediante un modelo no estacionario reflejan la gran influencia que estas estructuras presentan sobre la probabilidad de presencia del patógeno. La dispersión de la enfermedad a escala individual permite obtener una perspectiva detallada sobre su propagación, y el algoritmo propuesto permite manejar de forma eficiente grandes poblaciones. La evaluación de los distintos planes de gestión de brotes destaca la importancia de la vigilancia temprana, así

como la necesidad de adaptar las medidas de control a la situación epidemiológica específica de cada brote. Las metodologías propuestas y aplicadas a *X. fastidiosa* son flexibles y pueden ser aplicables a otras enfermedades de plantas.

Los contenidos de esta Tesis se presentan como se describe a continuación. El Capítulo 1 proporciona un contexto epidemiológico completo sobre *X. fastidiosa*, patógeno vegetal objeto de estudio central de esta Tesis. Se presentan las características fundamentales de la bacteria, sus principales huéspedes y los mecanismos de dispersión, lo que proporciona una visión completa de su biología y ecología. Se incluye también una visión global de la situación actual a nivel mundial, especificando su distribución geográfica y enfermedades principales que afectan a cada región. Para facilitar la comprensión de los conceptos abordados, se detalla el marco legislativo que regula su control en la Unión Europea, lo que contribuye a contextualizar la problemática. Además, este capítulo ofrece una revisión general de los modelos tanto espacialmente implícitos como explícitos que se han aplicado previamente a este patógeno. Esto incluye modelos de distribución de especies, los cuales estudian las posibles asociaciones entre la distribución geográfica del patógeno y factores climáticos relevantes. También se describen otros trabajos relacionados con diversas enfermedades causadas por este patógeno que se centran en el estudio de los patrones espaciales, así como modelos que describen las dinámicas de la enfermedad.

En el Capítulo 2 se presentan las bases de la metodología y los modelos utilizados para abordar los objetivos planteados en esta Tesis. En primer lugar, se introducen los modelos jerárquicos bayesianos, así como una descripción general de la metodología INLA. A continuación, nos centramos en la modelización espacial, comenzando por la descripción de los distintos tipos de datos espaciales, sus características generales y propósito de estudio. En mayor profundidad, detallamos cómo modelizar datos de área y geoestadísticos. En este último caso, también detallamos el enfoque de ecuaciones diferenciales parciales estocásticas (SPDE), el cual proporciona una herramienta rápida y precisa para la inferencia y la predicción. Estos últimos modelos asumen estacionariedad e isotropía de la componente espacial, de forma que la dependencia espacial entre dos localizaciones únicamente depende de la distancia euclidiana. Sin embargo, estos supuestos no son correctos cuando en el área de estudio existen barreras que obstaculizan la dispersión del patógeno, pudiendo dar lugar a resultados erróneos o predicciones poco precisas. Por ello, el enfoque SPDE es adaptado a los modelos no estacionarios y la inclusión de barreras. En la última sección de este capítulo se expone una visión global de los modelos

de dispersión de enfermedades, destacando la relevancia de los modelos basados en individuos (IBMs) en el estudio de la propagación de enfermedades en plantas. Se exploran aspectos clave a considerar al modelizar la dispersión de enfermedades, incluyendo la escala de estudio, la heterogeneidad espacial y las interacciones entre los huéspedes.

Después de establecer el contexto del problema y presentar la metodología general en los primeros capítulos, los Capítulos 3 al 6 se centran en abordar los objetivos específicos planteados. Cada uno de estos capítulos sigue una estructura similar a los artículos de investigación, incluyendo así una introducción donde se plantea el problema en cuestión, una sección que detalla la metodología específica empleada, un análisis de los resultados obtenidos y, por último, una discusión detallada sobre estos resultados.

El Capítulo 3 se centra en el estudio de los efectos climáticos y espaciales sobre la distribución de *X. fastidiosa* en dos regiones de Europa: Alicante (España) y Lecce (Italia). Estos brotes difieren en cuanto a sus características específicas desde el punto de vista epidemiológico, así como en cuanto al tipo de datos espaciales provenientes de los muestreos oficiales. En Lecce, *X. fastidiosa* subsp. *pauca* es la causante de la enfermedad conocida como el síndrome del decaimiento súbito del olivo. Los datos proporcionados de esta región contienen las coordenadas geográficas de las muestras recogidas para la detección del patógeno, considerándose como datos geoestadísticos. En Alicante, la enfermedad del chamuscado de la hoja del almendro es causada por *X. fastidiosa* subsp. *multiplex*. En este caso, los datos de los muestreos se presentan agregados en una cuadrícula, representando datos en una red fija de localizaciones. Los datos de presencia/ausencia de *X. fastidiosa* de las encuestas oficiales se analizan utilizando modelos jerárquicos bayesianos a través de la metodología INLA. Debido a los distintos tipos de datos espaciales en cada región, la modelización es adaptada a cada uno de ellos. Para ello, en el caso de los datos en una red fija de localizaciones, el efecto espacial se introduce mediante una estructura condicional autorregresiva. La modelización de datos geoestadísticos mediante la metodología INLA implica el uso de la aproximación SPDE. Junto con las covariables climáticas de la base de datos WorldClim v.2, se incluye una variable categórica (categorías de Purcell) basada en umbrales de temperatura invernal mínima establecida para el riesgo de la enfermedad de Pierce en la vid, causada por *X. fastidiosa* subsp. *fastidiosa*. Para evitar problemas de multicolinealidad, se realiza una preselección de las covariables climáticas, así como un análisis de componentes principales. Una vez definidos los modelos y las covariables a incluir, se

realiza el ajuste de los modelos con todas las posibles combinaciones de las covariables junto con el efecto espacial. Posteriormente, se evalúan los modelos y se hace una selección en función de diversos criterios. Para esta selección de modelos se emplea el criterio de información de Watanabe Akaike (WAIC) y la ordenada predictiva log-condicional (LCPO). El modelo seleccionado en el caso de Alicante es el que incluye únicamente el efecto espacial. En el caso de Lecce, además del efecto espacial, se mantienen dos covariables climáticas en el modelo seleccionado. Sin embargo, esto puede deberse a que la distribución heterogénea del muestreo tiene un efecto de confusión con las covariables climáticas. Se concluye que en ambos casos la estructura espacial tiene una fuerte influencia en los modelos, pero no las covariables climáticas, lo que sugiere que la distribución del patógeno se define principalmente por las relaciones espaciales entre las localizaciones geográficas. Este efecto espacial estaría determinado por los mecanismos de dispersión propios de *X. fastidiosa*, principalmente a través de los insectos vectores, aunque habría que tener en cuenta también el movimiento de material propagativo infectado. Estos resultados destacan la importancia de considerar la componente espacial en la modelización para posteriormente poder diseñar estrategias de vigilancia y control efectivas.

Los modelos espaciales empleados en el Capítulo 3 son herramientas muy útiles para comprender y predecir la distribución de los patógenos, pero a menudo asumen estacionariedad e isotropía que pueden no reflejar adecuadamente la realidad. Estos supuestos implican que la dependencia espacial es invariante a la dirección y uniforme sobre el área de estudio. Sin embargo, estos supuestos son erróneos cuando hay barreras que impiden u obstaculizan la dispersión del patógeno. A pesar de esto, el problema de la no estacionariedad ha sido muy poco explorado en el contexto de la sanidad vegetal. El Capítulo 4 se centra en abordar esta problemática y evaluar la influencia de las barreras sobre la distribución de *X. fastidiosa* en el área demarcada de Alicante. Para llevar a cabo este análisis, se utiliza una base de datos actualizada de la ocurrencia del patógeno en la región. Se emplean modelos espaciales jerárquicos bayesianos que permiten integrar de manera efectiva la componente espacial. Los datos georreferenciados de presencia/ausencia del patógeno se consideran como datos geoestadísticos, de forma que la estimación y predicción se realiza con la metodología INLA mediante el enfoque SPDE. En este caso no se consideran covariables climáticas, ya que el interés principal se centra en el efecto de los distintos tipos de barreras sobre la componente espacial. Se plantea un modelo estacionario y tres modelos que consideran la no estacionariedad mediante la inclusión de distintas barreras, físicas y de control. El modelo estacionario representa

un escenario en el que no se han aplicado intervenciones de control ni se consideran accidentes geográficos. Los modelos no estacionarios incluyen un modelo con las montañas como barreras físicas, y dos modelos con una barrera perimetral continua y discontinua que representan intervenciones de control hipotéticas. Se destaca el resultado de la media posterior del rango espacial del modelo estacionario, el cual indica la distancia a partir de la cual dos observaciones pueden considerarse no correlacionadas espacialmente. En base a ello, se compara con el tamaño de la zona tampón que determina la legislación y se concluye que los resultados obtenidos pueden servir de referencia para definir la zona tampón en el área demarcada. Como parte de la discusión, los resultados obtenidos se contrastan con las hipótesis sugeridas en otros trabajos, interpretando que en este caso el efecto espacial estaría principalmente determinado por la capacidad de dispersión de los insectos vectores. Los resultados del modelo estacionario se comparan con los modelos que incorporan las distintas barreras. Debido a la reducida extensión de las montañas como barreras, su incorporación no tiene gran influencia. A diferencia de esto, la incorporación de las barreras de control perimetrales reducen considerablemente la probabilidad de presencia del patógeno en el área externa de la barrera. Las diferencias entre los modelos de barrera continua y discontinua muestran que en las zonas de baja intensidad de muestreo, la falta de medidas de control incrementa la probabilidad de presencia del patógeno. La metodología implementada asume que las barreras son impermeables, por lo que se interpreta como áreas sin plantas hospedantes y a través de las cuales los vectores infectados o el material propagativo no puede pasar. Asumir estas propiedades sobre las medidas de control puede no ser demasiado realista, por lo que en este caso, la barrera perimetral discontinua estaría proporcionando una perspectiva más flexible respecto a la barrera continua. Sin embargo, se pone de manifiesto la necesidad de poder integrar barreras con distintos grados de permeabilidad para poder ofrecer un escenario más realista en el ámbito de la sanidad vegetal. Finalmente, se concluye que los resultados obtenidos pueden ayudar a las autoridades a priorizar las áreas para la vigilancia y el control de la enfermedad.

En el ámbito de la epidemiología de enfermedades en plantas, conocer la extensión de la enfermedad en el área de estudio, los posibles estados de la enfermedad, los mecanismos de dispersión y alcance de los mismos, son algunos de los aspectos fundamentales a la hora de estudiar la dispersión de la enfermedad y diseñar planes de gestión efectivos. Asimismo, como se ilustra en los capítulos anteriormente descritos, para el estudio de la dispersión de enfermedades es esencial considerar la estructura espacial de la población en cuestión. El Capítulo 5 se dedica a la

presentación de un modelo para la dispersión de enfermedades en plantas. Considerando la importancia de conocer el estado de la enfermedad en cada uno de los individuos, se plantea un modelo espacialmente explícito basado en individuos, capaz de tener en cuenta la heterogeneidad espacial. En este enfoque, el proceso de infección de un individuo susceptible (no infectado) está determinado por la fuerza de infección, la cual depende de la tasa de transmisión de la enfermedad y de la correlación espacial existente entre el individuo susceptible y los individuos infecciosos. Esta relación espacial se introduce mediante la función de correlación Matérn, que permite establecer la dependencia espacial entre dos individuos en función de la distancia. La simulación de la dispersión de la enfermedad se lleva a cabo empleando como caso de estudio la enfermedad del chamuscado de la hoja del almendro, causada por *X. fastidiosa*, en el área afectada de Alicante. Debido a la falta de datos espacio-temporales para poder estimar apropiadamente los parámetros del modelo, la dispersión de la enfermedad se simuló bajo distintos valores de los parámetros partiendo de la información disponible en la literatura. Además, el valor de referencia para la tasa de transmisión de la enfermedad se aproxima mediante la optimización de un modelo compartimental con los datos disponibles de los muestreos oficiales. Para estudiar el efecto debido a las características de la introducción inicial de los individuos infectados, se consideran diversos escenarios: introducción aleatoria, introducción agregada en 5 focos e introducción agregada en un único foco. En base al modelo propuesto, se explora el efecto de los diferentes parámetros en la propagación de la enfermedad, la variabilidad dependiendo de la introducción inicial, y la variabilidad intrínseca del modelo. Para ello, se realizan múltiples simulaciones de la dispersión de la enfermedad con cada una de las combinaciones de parámetros y tipo de introducción inicial. Dadas todas las combinaciones propuestas, para la visualización de los resultados se ha desarrollado un aplicación en `Shiny` que permite visualizar todos los resultados obtenidos de forma interactiva. La gran variabilidad obtenida en función del rango, así como del tipo de introducción inicial considerada, pone de manifiesto que son dos elementos con gran influencia sobre la dispersión de la enfermedad. Además, el modelo basado en individuos evidencia la influencia que ejerce la distribución espacial de los individuos en la propagación de la enfermedad. En particular, se destaca que las áreas sin la presencia de árboles actúan como barreras en el proceso de dispersión, siempre y cuando su extensión supere el valor del rango de la función Matérn. Cabe destacar que este modelo basado en individuos puede ser aplicado a otras enfermedades vegetales, adaptando los valores de los parámetros a sus características epidemiológicas

particulares. Además, este modelo de dispersión puede ser una herramienta útil para establecer áreas de mayor riesgo, evaluar la vigilancia epidemiológica y optimizar estrategias de control.

La regulación de brotes de patógenos y plagas de cuarentena en la Unión Europea se rige por la legislación, sin embargo, los planes de gestión de brotes han sido cuantificados de manera muy limitada en términos de su efectividad y eficiencia. En el Capítulo 6 se plantea evaluar la eficacia y eficiencia de planes de gestión de brotes para enfermedades emergentes de plantas a través de un enfoque epidemiológico. Como caso de estudio nos centramos en el área afectada en Alicante por la enfermedad del chamuscado de la hoja del almendro. El inicio del brote se data en función de estudios filogenéticos, los cuales sugieren la primera aparición de este brote décadas atrás. Esto, por tanto, ofrece un escenario poco habitual donde la enfermedad se ha estado dispersando durante un largo periodo de tiempo sin ningún tipo de intervención. La extensión de la enfermedad en el área de estudio se delimita a partir de los datos de vigilancia oficiales. En base a esto, haciendo uso del modelo de dispersión basado en individuos detallado en el Capítulo 5, donde los individuos son los almendros, se plantea un escenario de base donde la enfermedad se dispersa a lo largo 30 años sin ningún tipo de intervención de vigilancia o control. Los planes de gestión de brotes a evaluar contemplan distintas estrategias de vigilancia y medidas de control, incluyendo los establecidos por la legislación. Las estrategias de vigilancia se definen siguiendo las "Directrices para las prospecciones estadísticamente sólidas y basadas en el riesgo de la Autoridad Europea de Seguridad Alimentaria (EFSA)". De esta forma, el tamaño de muestra representativo en cada caso se calcula en base a la distribución hipergeométrica, la cual depende del tamaño poblacional y de tres parámetros interrelacionados: el nivel de confianza, la sensibilidad del método y la prevalencia del diseño. Se consideran dos enfoques para la vigilancia: el enfoque en un paso, donde el tamaño de muestra se calcula en función del tamaño de la población; y el enfoque en dos pasos, donde tras dividir el área de estudio en celdas, en el primer paso se calcula el número de muestras por celda y en el segundo paso, el número celdas a inspeccionar. Las medidas de control incluyen distintos tamaños para la zona tampón, distintos tamaños de radio de erradicación y el efecto de las medidas de control sobre los vectores y la reducción del inóculo mediante la reducción de la tasa de transmisión de la enfermedad en la zona tampón. Se evalúa la efectividad y eficiencia de los distintos planes de gestión de brotes resultantes de las combinaciones de las estrategias de vigilancia y control en relación al escenario de referencia sin intervenciones, definiendo la efectividad en función de la reducción de individuos infectados, y la eficiencia en función de

los árboles susceptibles no erradicados. Además, se cuantifica la intensidad de inspección, es decir, las celdas inspeccionadas al año, y el esfuerzo de muestreo, es decir, el número de muestras recogidas por año. Como resultado de la simulación de la dispersión de la enfermedad en el escenario de referencia sin intervenciones, prácticamente todo el área de estudio se infecta tras 30 años, con una extensión que puede asemejarse a la proporcionada por los datos de los muestreos oficiales. La poca variabilidad observada resultante de las múltiples simulaciones realizadas con cada plan de gestión de brotes indica que los resultados son consistentes. Se observa gran efectividad en todos los planes de gestión de brotes implementados al ser capaces de reducir considerablemente el número de árboles infectados. Sin embargo, la eliminación por completo de los árboles infectados únicamente ocurre en planes de gestión con el enfoque de vigilancia en dos pasos. Por otra parte, se observan grandes diferencias en cuanto a la eficiencia, en términos de árboles susceptibles no erradicados, de los distintos planes de gestión de brotes. Se evidencia que el aumento del nivel de confianza para el cálculo del tamaño de muestra, el enfoque de vigilancia en dos pasos y la reducción de la tasa de transmisión de la enfermedad implican una mayor eficiencia en los planes de gestión de brotes. A pesar de que el enfoque de vigilancia en dos pasos destaca por su eficacia y eficiencia, se debe considerar que supone un elevado esfuerzo de muestreo. Por otra parte, este enfoque permitiría calcular el coste tanto de la intensidad de inspección como del esfuerzo de muestreo, permitiendo un ajuste de los parámetros en función del mismo. Sin embargo, con el enfoque en un paso únicamente podría calcularse el coste que implica el esfuerzo de muestreo. Los resultados obtenidos con los distintos tamaños de zona tampón y radios de erradicación se comparan con otros estudios que también emplean estas medidas de control. En este trabajo, no se observa mayor influencia sobre la eficacia y eficiencia de los planes de gestión de brotes al variar el tamaño de la zona tampón y del radio de erradicación. Sin embargo, estas medidas de control ayudan a reducir la propagación de la enfermedad. A pesar de esto, destaca el efecto sobre la eficacia y eficiencia de las estrategias de vigilancia es destacable. De esta forma, se pone en evidencia el papel fundamental de la vigilancia y la detección temprana sobre los planes de gestión de brotes. Por otra parte, aunque el enfoque de vigilancia en dos pasos se describe en las guías de EFSA, es la primera vez que es evaluado en estos términos y que se cuantifica el esfuerzo de muestreo que puede suponer. Además, como aporte novedoso, el proceso utilizado para el enfoque de vigilancia en dos pasos, que normalmente parte del supuesto de que hay un número uniforme de individuos por celda, se ajusta a una situación en la que el número de individuos por celda puede variar.

En el Capítulo 7 se exponen en conjunto las conclusiones más relevantes derivadas de la metodología y resultados de los trabajos presentados en esta Tesis. Se destaca la importancia de las herramientas de modelización espacial en la comprensión y gestión de las epidemias de enfermedades en plantas, así como las implicaciones prácticas de los resultados obtenidos. Siguiendo a este último capítulo, se proporciona un listado completo de la bibliografía utilizada en la elaboración de esta Tesis. Por último, se incluyen tres anexos con figuras y tablas suplementarias a los trabajos expuestos.

Contents

List of Figures

List of Tables

Introduction

The exponential growth of global trade and international travel in recent decades has had a major impact on the introduction and spread of pests and pathogens affecting important crops around the world. In addition, climate change has provided favorable climatic conditions for the establishment of these pests and pathogens in new geographic areas. emerging pests and pathogens threaten global food security and socioeconomic stability. This presents a major challenge for agriculture in a context where there is a need to increase food production to meet the needs of a growing world population. Therefore, understanding the factors involved in the entry, establishment and spread of pathogens and disease dynamics, is crucial to effectively prevent and manage new outbreaks and minimize their impact.

In the last decade, the plant pathogenic bacterium *Xylella fastidiosa* has caused a phytosanitary crisis in the European Union, affecting, among others, two of the most important Mediterranean crops, olive and almond trees. One of the complexities in the study of this pathogen is its wide genetic diversity and its ability to infect an extensive range of host plants. Despite its presence for decades in the American continent, and the numerous studies carried out, there are still many unknowns regarding its epidemiology, as well as surveillance, control and prevention strategies. This pathogen has been the main case study of this Thesis, in which we address different issues that have emerged dynamically throughout its study. It is noteworthy that

during the course of this Thesis, both the available data on the distribution of the pathogen and new knowledge from scientific advancements have been incorporated.

The general objective of this PhD Thesis is to provide different spatial modeling tools applicable to the study of plant disease epidemics, to analyze their spatial distribution and related factors, as well as to serve as a reference for their control. In particular, we address several issues on *X. fastidiosa* through advanced statistical methods that are novel in the field of plant pathology, with special focus to the analysis of its spatial distribution. In this context, the following main objectives are proposed: i) the analysis of climatic and spatial effects on the distribution of *X. fastidiosa*; ii) the effects of the incorporation of dispersal barriers on the distribution of the pathogen; iii) the implementation of a spatially explicit individual-based spread model; and iv) the evaluation of the effectiveness and efficiency of different outbreak management plans, including surveillance and control strategies.

Spatial modeling is approached from the perspective of Bayesian hierarchical spatial models, particularly using the integrated nested Laplace approximation (INLA) for inference and prediction. Through this methodology, lattice data and geostatistical data are modeled, adapting the models to each type of spatial data. As an extension of these models, a non-stationary model is used for the incorporation of dispersal barriers. Moreover, a simulation model of disease spread at individual level is proposed, where a correlation function establishes their spatial relationship. This model allows establishing the basis for the implementation and evaluation of surveillance and control strategies.

Regarding the most relevant results and conclusions of this Thesis, Bayesian hierarchical spatial models are effective in dealing with different types of spatial data of pathogen outbreaks. The strong effect of the spatial component on pathogen distribution in the analyzed outbreaks, in contrast to the low relevance of climatic variables, highlights the importance of taking into account the spatial structure in the models. The inclusion of dispersal barriers using a non-stationary model illustrates the strong influence of the incorporation of these structures on the probability of pathogen presence. Disease spread at the individual scale provides detailed on its propagation,

and the proposed algorithm allows for efficient management of large populations. The evaluation of different outbreak management plans highlights the importance of early surveillance, as well as the need to adapt control measures to the specific epidemiological situation of each outbreak. The methodologies proposed and applied to *X. fastidiosa* are flexible and can be extended to other plant diseases.

The contents of this Thesis are presented as outlined as follows. Chapter 1 provides a complete epidemiological background on *X. fastidiosa*, plant pathogen that is the main subject of study of this Thesis. The principal characteristics of the bacterium, its main hosts and dispersal mechanisms are presented, providing a complete overview of its biology and ecology. An overview of the current situation worldwide is also included, specifying its geographical distribution and the main diseases affecting each region. To facilitate the understanding of the concepts addressed, the legislative framework that regulates their control in the European Union is detailed, which helps to contextualize the problem. In addition, this chapter provides a general review of both spatially implicit and explicit models that have been previously applied to this pathogen. This includes species distribution models, which study possible associations between the geographic distribution of the pathogen and relevant climatic factors. Other works, related to different diseases caused by this pathogen that focus on the study of spatial patterns, as well as models describing disease dynamics, are also described.

Chapter 2 presents the basis of the methodology and the models used to address the objectives of this Thesis. First, Bayesian hierarchical models are introduced, as well as a general description of the INLA methodology. Next, we focus on spatial modeling, starting with a description of the different types of spatial data, their general characteristics and purpose of study. In more depth, we detail how to model areal and geostatistical data. In the latter case, we also detail the stochastic partial differential equations (SPDE) approach, which provides a fast and accurate tool for inference and prediction. Geostatistical models assume stationarity and isotropy of the spatial component, so that the spatial dependence between two locations depends only on Euclidean distance. However, these assumptions are not correct when there are barriers in the study area that hinder the spread of the pathogen, which can lead to erroneous results or inaccurate predictions.

Therefore, the SPDE approach is adapted to non-stationary models and the inclusion of barriers. The last section of this chapter provides an overview of disease spread models, highlighting the relevance of individual-based models (IBMs) in the study of disease spread in plants. Key aspects to consider when modeling disease spread are explored, including scale of study, spatial heterogeneity, and interactions between hosts.

After establishing the context of the problem and presenting the general methodology in the first chapters, Chapters 3 to 6 focus on addressing the specific objectives. Each of these chapters follows a structure similar to research papers, thus including an introduction where the particular problem is presented, a section detailing the specific methodology employed, an analysis of the results obtained and, finally, a detailed discussion of these results.

Chapter 3 focuses on the study of climatic and spatial effects on the distribution of *X. fastidiosa* in two regions of Europe: Alicante (Spain) and Lecce (Italy). These outbreaks differ in their epidemiological specificities as well as in the type of spatial data from official surveys. In Lecce, *X. fastidiosa* subsp. *pauca* is the cause of the disease known as olive quick decline syndrome. The data provided for this region contain the geographical coordinates of the samples collected for pathogen detection and are considered as geostatistical data. In Alicante, almond leaf scorch disease is caused by *X. fastidiosa* subsp. *multiplex*. In this case, sampling data are presented aggregated in a grid, representing lattice data. The presence/absence data of *X. fastidiosa* from official surveys are analyzed using Bayesian hierarchical models through the INLA methodology. Due to the different spatial data types in each region, the modeling is adapted to each of them. For this purpose, in the case of lattice data, the spatial effect is introduced through a conditional autoregressive structure. The modeling of geostatistical data using the INLA methodology involves the use of the SPDE approach. Along with the climatic covariates from the WorldClim v.2 database, a categorical variable (Purcell categories) is included. These categories are based on minimum winter temperature thresholds established for the risk of Pierce's disease in grapevines, caused by *X. fastidiosa* subsp. Fastidiosa. To avoid multicollinearity issues, a preselection of the climatic covariates is performed, as well as a principal component analysis. Once

the models and the covariates to be included have been defined, the models are fitted with all the possible combinations of covariates together with the spatial effect. Subsequently, the models are evaluated and a selection is made based on different criteria. For this model selection, the Watanabe Akaike information criterion (WAIC) and the log-conditional predicted ordinate (LCPO) are used. The model selected in the case of Alicante is the one that includes only the spatial effect. In the case of Lecce, in addition to the spatial effect, two climatic covariates are retained in the selected model. However, this may be because the heterogeneous sampling distribution has a confounding effect with climatic covariates. It is concluded that in both cases the spatial structure has a strong influence on the models, but not the climatic covariates, suggesting that pathogen distribution is mainly defined by the spatial relationships between geographic locations. This spatial effect would be determined by the dispersal mechanisms of *X. fastidiosa*, mainly through insect vectors, although the movement of infected propagative material should also be taken into account. These results highlight the importance of considering the spatial component in modeling in order to be able to design effective surveillance and control strategies.

The spatial models employed in Chapter 3 are very useful tools for understanding and predicting pathogen distribution, but they often assume stationarity and isotropy that may not adequately reflect reality. These assumptions imply that spatial dependence is directionally invariant and uniform over the study area. However, these assumptions are erroneous when there are barriers that prevent or hinder pathogen spread. Despite this, the issue of non-stationarity has been little explored in the context of plant health. Chapter 4 of this Thesis focuses on addressing this issue and evaluating the influence of barriers on the distribution of *X. fastidiosa* in the demarcated area of Alicante. To perform this analysis, an updated database of the occurrence of the pathogen in the region is used. Bayesian hierarchical spatial models are used to effectively integrate the spatial component. The georeferenced pathogen presence/absence data are considered as geostatistical data, so that inferential and prediction processes are performed with the INLA methodology using the SPDE approach. In this case, climatic covariates are not considered, since the main interest is focused on the effect of different types of barriers on the spatial component. A stationary model

and three models that consider non-stationarity through the inclusion of different physical and control barriers are proposed. The stationary model represents a scenario in which no control interventions have been applied and no geographical features are considered. The non-stationary models include a model with mountains as physical barriers, and two models with a continuous and discontinuous perimeter barrier representing hypothetical control interventions. The result of the posterior mean of the spatial range of the stationary model is highlighted, which indicates the distance from which two observations can be considered spatially uncorrelated. Based on this, it is compared with the size of the buffer zone determined by the legislation and it is concluded that the results obtained can serve as a reference to define the buffer zone in the demarcated area. As part of the discussion, the results obtained are contrasted with the hypotheses suggested in other works, interpreting that in this case the spatial effect would be mainly determined by the dispersal capacity of the insect vectors. The results of the stationary model are compared with the models that incorporate the different barriers. Due to the reduced extension of the mountains as barriers, their incorporation does not have a great influence. In contrast to this, the incorporation of perimeter control barriers considerably reduces the probability of pathogen presence in the area outside the barrier. The differences between the continuous and discontinuous barrier models show that in areas of low sampling intensity, the lack of control measures increases the probability of pathogen presence. The implemented methodology assumes that the barriers are impermeable, thus interpreted as areas without host plants and through which infected vectors or propagating plant material cannot pass. These assumptions about control measures may not be too realistic, so in this case, the discontinuous perimeter barrier would be providing a more flexible perspective with respect to the continuous barrier. However, the need to be able to integrate barriers with different degrees of permeability is highlighted in order to provide a more realistic scenario in the field of plant health. Finally, it is concluded that the results obtained can help authorities to prioritize areas for disease surveillance and control.

In plant disease epidemiology, knowledge of the extent of the disease in the study area, the possible disease states, the dispersal mechanisms and their range are some of the fundamental aspects when studying disease

spread and for the design of effective management plans. Likewise, as illustrated in the chapters described above, for the study of disease spread it is essential to consider the spatial structure of the population in question. Chapter 5 of this Thesis is dedicated to present a model for plants disease spread. Considering the importance of knowing the disease status of each individual, a spatially explicit individual-based model based, able to take into account spatial heterogeneity, is proposed. In this approach, the infection process of a susceptible (uninfected) individual is determined by the force of infection, which depends on the disease transmission rate and the spatial correlation between the susceptible individual and the infectious individuals. This spatial relationship is introduced by the Matérn correlation function, which makes it possible to establish the spatial dependence between two individuals as a function of distance. The simulation of disease spread is performed using as a case study the almond leaf scorch disease, caused by *X. fastidiosa*, in the affected area of Alicante. Due to the lack of spatio-temporal data to properly estimate the model parameters, the spread of the disease was simulated under different parameter values based on the information available in the literature. In addition, the reference value for the disease transmission rate is approximated by optimizing a compartmental model with the available data from official surveys. To study the effect due to the characteristics of the initial introduction of infected individuals, several scenarios are considered: random introduction, aggregated introduction in 5 foci and aggregated introduction in a single focus. Based on the proposed model, the effect of the different parameters on the disease spread, the variability depending on the initial introduction, and the intrinsic variability of the model are explored. For this purpose, multiple simulations of disease spread are performed with each of the combinations of parameters and type of initial introduction. Given all the combinations proposed, a Shiny application has been developed for the visualization of the results, which allows all the results obtained to be displayed interactively. The great variability obtained depending on the range, as well as the type of initial introduction considered, shows that these are two elements with a great influence on the disease spread. In addition, the individual-based model shows the influence of the spatial distribution of individuals on the disease spread. In particular, it is highlighted that areas without the presence of trees act as barriers in the spread process, as long as their extension exceeds

the value of the range of the Matérn function. It should be noted that this individual-based model can be applied to other plant diseases, adapting the parameter values to their particular epidemiological characteristics. In addition, this spread model can be a useful tool to establish areas of higher risk, evaluate epidemiological surveillance and optimize control strategies.

The regulation of outbreaks of quarantine pathogens and pests in the European Union is regulated by legislation. However, outbreak management plans have been quantified in a limited extent in terms of their effectiveness and efficiency. In Chapter 6 we propose to evaluate the effectiveness and efficiency of outbreak management plans for emerging plant diseases through an epidemiological approach. As a case study, we focus on the area affected in Alicante by almond leaf scorch disease. The onset of the outbreak is dated based on phylogenetic studies, which suggests the first occurrence of this outbreak decades ago. This, therefore, provides an unusual scenario where the disease has been spreading over a long period of time without any intervention. The extent of the disease in the study area is delimited from official survey data. Based on this, using the individual-based spread model detailed in Chapter 5, where the individuals are the almond trees, a baseline scenario is proposed where the disease spreads over 30 years without any type of surveillance or control intervention. The outbreak management plans to be evaluated contemplate different surveillance strategies and control measures, including those established by legislation. Surveillance strategies are defined following the "Guidelines for statistically sound and risk-based surveys of the European Food Safety Authority (EFSA)"'. Thus, the representative sample size in each case is calculated based on the hypergeometric distribution, which depends on the population size and three interrelated parameters: the confidence level, the method sensitivity and the design prevalence. Two approaches to surveillance are considered: the one-step approach, where the sample size is calculated based on the population size; and the two-step approach, where after dividing the study area into cells, the number of samples per cell is calculated in the first step and the number of cells to be inspected in the second step. The control measures include different sizes for the buffer zone, different sizes of eradication radius, and the effect of the control measures on the vectors and the reduction of inoculum by reducing the disease transmission rate in the

buffer zone. The effectiveness and efficiency of the different outbreak management plans resulting from the combinations of surveillance and control strategies are evaluated in relation to the baseline scenario without interventions, defining effectiveness as a function of the reduction of infected individuals, and efficiency as a function of susceptible trees not eradicated. In addition, the inspection intensity, i.e., cells inspected per year, and the survey effort, i.e., number of samples collected per year, are quantified. As a result of the simulation of disease spread in the baseline scenario without interventions, almost the entire study area is infected after 30 years, with an extension that approximates to that provided by official survey data. The low variability observed resulting from the multiple simulations performed with each outbreak management plan indicates that the results are consistent. All implemented outbreak management plans have been observed to be highly effective in reducing the number of infected trees considerably. However, complete elimination of infected trees only occurs in management plans with the two-step surveillance approach. On the other hand, large differences in the efficiency, in terms of susceptible trees not eradicated, of the different outbreak management plans are observed. It is evident that the increased confidence level for the sample size calculation, the two-step surveillance approach and the reduction of the disease transmission rate imply higher efficiency in the outbreak management plans. Although the two-step surveillance approach outperforms in effectiveness and efficiency, it should be considered that it involves a high survey effort. On the other hand, this approach would allow calculating the cost of both inspection intensity and survey effort, enabling parameter adjustment based on it. However, with the one-step approach, only the cost of the sampling effort could be calculated. The results obtained with different buffer zone sizes and eradication radii are compared with other studies that also employ these control measures. In this work, these measures do not have a major influence on the effectiveness and efficiency of outbreak management plans, however, they contribute to reduce the disease spread. Despite the role of these control measures, the effect on the effectiveness and efficiency of surveillance strategies is noteworthy. Thus, one of the most notable conclusions of this work is the key role of surveillance. Moreover, although the two-step surveillance approach is described in EFSA guidelines, it is the first time that it is evaluated in these terms and that the survey effort involved is quantified. In

addition, as a novel contribution, the process used for the two-step surveillance approach, which usually assumes a uniform number of individuals per cell, is adjusted to a situation where the number of individuals per cell may vary.

Chapter 7 presents the most relevant conclusions derived from the methodology and results of the work presented in this Thesis. The importance of spatial modeling tools in the understanding and management of plant disease epidemics is highlighted, as well as the practical implications of the results obtained. Following this last chapter, a complete list of the bibliography used in the development of this Thesis is provided. Finally, three annexes with supplementary figures and tables are included.

Chapter 1

Diseases caused by *Xylella fastidiosa*. Epidemiology and spatial analysis

This chapter aims to present the epidemiological context of *X. fastidiosa* and the current global situation to provide the background and motivation for the study objectives addressed. The regulatory framework in the European Union is described to facilitate the understanding of the concepts discussed. A general review of the spatially implicit and explicit models that have been previously applied to this pathogen is also presented.

1.1 Epidemiology of *Xylella fastidiosa* diseases

Xylella fastidiosa (Wells et al., 1987) is a plant-pathogenic bacterium of the Xanthomonadaceae family. It colonizes the xylem tissues of a wide range of plant species: the latest update of the host plants database includes 690 species belonging to 306 genera and 88 different families (EFSA, 2023). However, not all of these plants develop disease symptoms, and both infection and disease severity depend on the interaction between the host

plant, the pathogen aggressiveness and the environment. It is only in some combinations of host plant and bacterial strain that infections can result in some of the most destructive diseases, such as Pierce's disease of grapevine (PD), olive quick decline syndrome (OQDS), and almond leaf scorch disease (ALSD), threatening several crops of great economic importance.

A total of six subspecies of *X. fastidiosa* have been defined to date, although based on genetic differences, host ranges, and associated diseases the majority of the bacterial strains fall into four main subspecies. *X. fastidiosa* subsp. *fastidiosa* causes, among others, PD and ALSD; *X. fastidiosa* subsp. *pauca* has been identified as the cause of OQDS, citrus variegated chlorosis (CVC) and coffee leaf scorch (CLS). *X. fastidiosa* subsp. *multiplex* has the widest host range, being found in more than 200 plant species. It is associated with leaf scorch diseases in numerous tree species, including ALSD (EFSA, 2023). *X. fastidiosa* subsp. *sandyi* has been associated with oleander leaf scorch (Schaad et al., 2004; Janse and Obradovic, 2010). Nevertheless, the Committee on the Taxonomy of Plant Pathogenic Bacteria of the International Society of Plant Pathology (ISPP) only considers as valid subspecies names *fastidiosa* and *multiplex* (Bull et al., 2012). In addition to these, the subsp. *tashke* (Randall et al., 2009) and subsp. *morus* (Nunney et al., 2014b) have been proposed. Indeed, recently the *X. fastidiosa* strain isolated from *Pyrus pyrifolia* in Taiwan has been proposed to form a distinct species in the genus *Xylella*: the *X. taiwanensis* sp. nov (Su et al., 2016).

This pathogen was first described as the cause of PD in California, US. For decades, the geographic distribution of *X. fastidiosa* was restricted to the American continent (Janse and Obradovic, 2010), but in 2013 it was first reported in Europe as causing OQDS in southern Italy, with thousands of hectares affected and millions of trees killed (Martelli, 2016). Two years later, *X. fastidiosa* subsp. *multiplex* was reported in Corsica and Provence-Alpes-Côte d'Azur (PACA) in France, where *X. fastidiosa* subsp. *pauca* was also detected. Three subspecies of *X. fastidiosa* (i.e., *multiplex*, *pauca*, and *fastidiosa*) were later reported in the Balearic Islands. Outbreaks associated with *X. fastidiosa* subsp. *multiplex* were detected in Alicante and Madrid, Spain, Porto District, Portugal, and Tuscany and Lazio, Italy (DG SANTE, 2023). In 2019, the presence of *X. fastidiosa* subsp. *fastidiosa* was officially

reported in Israel (EPPO, 2019a). In Lisbon and Algarve, Portugal, *X. fastidiosa* was detected in 2021, subspecies not yet confirmed. The outbreaks in Madrid and Algarve are considered closed, i.e., the infestation has been officially eradicated (DG SANTE, 2023).

Xylem sap-feeding insects represent the only natural means of spreading *X. fastidiosa* between plants. The bacterium colonize the foregut of the insects that act as vectors. The human-assisted spread of *X. fastidiosa*, however, is mainly through the movement of infected plants and grafting. Therefore, two habitats are distinguished in the life cycle of *X. fastidiosa*, one is the foregut of the insect vectors and the second is the xylem tissue of the host plant (Almeida and Nunney, 2015). All confirmed vectors of *X. fastidiosa* are hemiptera belonging to two groups of insects: sharpshooters (Cicadellidae family, Cicadellinae subfamily) (Almeida and Purcell, 2003) and spittlebugs (Aphrophoridae, Cercopidae and Clastopteridae families) (Cornara et al., 2018). The transmission efficiency of *X. fastidiosa* by vectors depends on several factors, such as the ecology and abundance of the insects present in the area as well as the population density of the bacterium in the host plants (Almeida et al., 2005). Vectors acquire the bacteria by feeding on the xylem of infected plants, but lose infectivity with molting. Adult vectors can inoculate healthy plants immediately after acquisition and throughout their whole lifetime, although the bacterium is not transmitted to the progeny (Almeida and Purcell, 2006). The pathogen can be introduced and further spread into new areas with infected plant material for planting or grafting (EFSA, 2015). Human-assisted movement of infected insect vectors on plants or on their own as 'hitch-hikers' in vehicles can also disseminate *X. fastidiosa*, though information on these means of spread is limited (EFSA, 2015).

In the American continent, species of sharpshooters are the main vectors of *X. fastidiosa*. Most of the available information on these vectors is based on studies conducted with *Homalodisca vitripennis* (Germar) (Hemiptera: Cicadellidae), also known as *Homalodisca coagulata* (Say). It is an invasive species and currently considered the main vector of PD in California, US. In Europe, sharpshooters are not abundant or widespread. Different species of spittlebugs have been found to play, or potentially play, a major role in the transmission of *X. fastidiosa* (Cornara et al., 2018). In particular,

Philaenus spumarius L. (Hemiptera: Aphrophoridae) has been considered the main vector of *X. fastidiosa* subsp. *pauca* in Italy (Cavalieri et al., 2019). This species of spittlebug is widely distributed and is characterized by its color polymorphism and the secretion of foam by nymphs. In Alicante, of the potentially vector insects analyzed, *X. fastidiosa* subsp. *multiplex* has been detected in three species: *P. spumarius*, *Neophilaenus campestris* Fallen (Hemiptera: Aphrophoridae) and *Neophilaenus lineatus* L. (Hemiptera: Aphrophoridae) (GVA, 2023).

In Alicante, *X. fastidiosa* subsp. *multiplex* ST6 was identified, where of the 5,540 positives confirmed in 21 different plant species until June 2023, 89,4 % of them were found in almond trees (GVA, 2023). Studies comparing genetic sequences of *X. fastidiosa* strains from Alicante and elsewhere managed to trace back to the origin of this introduction, which was likely initiated with infected plant material from California, USA (Landa et al., 2020). Phylogenetic studies have also suggested an approximate date of introduction in Alicante as far back as the mid-1990s (Moralejo et al., 2020). Starting in 2017, surveys have been conducted in Alicante to delimit the infested zone and substantiate disease freedom in the surrounding buffer zone. Eradication measures have been progressively applied over recent years, including vector control and the removal of host plants in the infested zone. With a demarcated area of 139,459 ha and 187,400 almond trees already destroyed (GVA, 2023), this outbreak represents one of the largest plant disease eradication campaigns ever attempted in Europe.

1.2 Regulatory framework in the European Union

Despite the implementation of phytosanitary measures for risk reduction and border controls, new introductions of pests and diseases are likely to occur. Between 1999 and 2019, a mean of 10 new introductions of diseases and pests were reported in the European Union (EU), with up to 25 new introductions per year (Rosace et al., 2023). In response to this situation, plant health has acquired unprecedented prominence in the policy agenda of the EU. The EU modernized the rules on plant health and in 2019 adopted Regulation (EU) 2016/2031 on protective measures against plant pests and

pathogens (EU, 2016) (hereinafter "EU Plant Health Law"). This new EU Plant Health Law addresses the epidemiological surveillance of the EU territory, adjusting the intensity of the measures according to pest and pathogen categorization and prioritization (Montanari and Traon, 2017). They are the so-called Union quarantine "pests" by Implementing Regulation (EU) 2019/2072, which are subject to strict phytosanitary measures and 20 of them are classified as priority "pests" by Regulation (EU) 2019/1702 with enhanced requirements. To preserve the EU territory from its introduction from areas where the presence of the bacterium is known, *X. fastidiosa* is regulated as a quarantine organism and it is also included in the list of the priority "pests" for the EU.

Eradication or containment of new outbreaks is only feasible in situations where their extent and prevalence are still limited. Once the pest or disease spreads due to delayed and incomplete detection of outbreaks, their management is much more costly and often not feasible (Václavík et al., 2010; Brown et al., 2020). In this regard, the EU Plant Health Law establishes enhanced risk-based surveillance of the EU territory for early detection of new introductions. Additionally, contingency and action plans are also foreseen for outbreak management in the case of priority "pests".

Epidemiological surveillance of the territory is targeted to specific locations and hosts, following a risk-based structured design to provide a statistical basis for the interpretation of the collected data (Brown et al., 2020). The European Food and Safety Agency (EFSA) provides scientific advice to EU Member States in different fields, including epidemiological surveillance. The EFSA Plant Pest Survey Toolkit (EFSA, 2023) provides the methodological framework and tools for survey planning, specifically to prepare and design surveys under a statistical risk-based approach (EFSA, 2020a,b,c,d). Based on these EFSA guidelines, new requirements for enhanced statistically-based surveillance have been introduced for quarantine pests and pathogens in the EU. Detection surveys are conducted to detect outbreaks in their early stages or to substantiate pest freedom. Delimiting surveys are carried out once a new outbreak is detected, to delineate the infested zone. Monitoring surveys are implemented to quantify the prevalence of the pest or pathogen in infested zones (EFSA, 2020a; Brown et al., 2020; Lázaro et al., 2021).

After the first detection of *X. fastidiosa* in the EU, emergency phytosanitary measures were laid down under Decision (EU) 2015/789 in all EU Member States with the aim of preventing further introduction and spread within the EU territory. In southern Apulia, Italy, two years after the first detection of *X. fastidiosa*, the demarcated area was declared a containment zone, as the pathogen was widely established and eradication was considered not feasible. Since then, the demarcated area has been updated several times as the disease spread north. Commission Implementing Regulation (EU) 2020/1201 provides for containment zones intensive surveillance and immediate removal of at least infected plants the last 5 km of the infected zone adjacent to the buffer zone, as well as around places of high cultural and social value.

In Alicante, since the first report in 2017 of *X. fastidiosa*, eradication measures were adopted in accordance with the regulations, currently Commission Implementing Regulation (EU) 2020/1201. Following this Regulation, a demarcated zone has been established in the territory, consisting of an infested zone surrounded by a buffer zone. The infested zone is delimited by a minimum radius of 50 m around the infected plants. The buffer zone is then defined with a radius of at least 2.5 km. Annual surveys are conducted to update the demarcated area (GVA, 2021). According to the *X. fastidiosa* Regulation, surveillance strategies should follow the EFSA guidelines for statistically sound and risk-based surveys (EFSA, 2020a,d). Eradication measures include the removal of all host plants in the infested zone and control of vector populations in the whole demarcated area by means of chemical, biological or cultural treatments.

1.3 Spatial models for *Xylella fastidiosa*

A number of studies have been conducted on the epidemiology of diseases caused by *X. fastidiosa*, through different methodologies and with different specific objectives. Nevertheless, in general the main aim has been to contribute to better understand epidemic dynamics and improve disease management.

Previous studies have focused on the study of the relationship between the distribution of *X. fastidiosa* and climatic variables using different approaches, generally included in the so-called species distribution models (SDMs), which will be addressed in more detail in Chapter 2. The Maxent model was used to evaluate the potential distribution of *X. fastidiosa* in Italy based on climatic variables (Bosso et al., 2016b). Also, this approach was applied to predict the potential distribution of *X. fastidiosa* in the Mediterranean basin including climate change scenarios (Bosso et al., 2016a). The Maxent model has also been used to study the distribution of *X. fastidiosa* in the Balearic Islands (Hernández and García, 2018), and on a global scale (Hernández and García, 2019). It has also been used to analyze the potential distribution of *X. fastidiosa* subsp. *multiplex* and the vector *P. spumarius* in southern Spain, including the regions of Alicante, Murcia and Andalusia (Fernández, 2020). Godefroid et al. (2019) analyzed the potential effect of climate change on different subspecies of *X. fastidiosa*. The SDMs Bioclim and Domain were fitted using presence data of the subspecies *multiplex* and *pauca* to estimate their potential geographic distribution under current and future climate conditions. Furthermore, the severity of PD and its relationships with climatic variables were modeled by means of ordinal regression, using the PD risk maps proposed by Purcell (Anas et al., 2008). These data, adapted and updated, were also used to model, through GLMs, the distribution of PD and bacterial leaf scorch of shade trees (BLS) in order to predict the potential extent and severity as a function of climatic conditions under future climate change Godefroid et al. (2022). EFSA (2019) used SDM ensemble modeling to asses the potential establishment of *X. fastidiosa* in the EU.

The analysis of spatial patterns and the incorporation of spatial structure have also been used for modeling diseases caused by *X. fastidiosa*, through different approaches. A commonly used method to analyze the spatial aggregation of infected plants between rows and fields is the ordinary run analysis, where the null hypothesis that infected plants are randomly distributed indicating that the pathogen does not spread from plant to plant (Madden et al., 1982). Da Rocha et al. (2010) used this approach for the analysis of the spatial aggregation of CLS in Brazil, complemented by the analysis of the spatial distribution of disease incidence using the modified

Taylor's law (Madden and Hughes, 1995). This approach was also used in the study of the spatial aggregation of PD Tubajika et al. (2004) and ALSD (Groves et al., 2005) in San Joaquin Valley, California. In addition, both studies complemented the spatial analysis with geostatistical methods. The anisotropy and spatial dependence of PD in grapevines over different years (Tubajika et al., 2004), and the spatial dependence of ALSD in different almond cultivars (Groves et al., 2005) was analyzed by fitting semivariogram models. Park et al. (2011) also used semivariograms to characterize the spatial distribution of PD in vine blocks in San Joaquin Valley, including spatial analysis with distance indices (SADIE) to test the significance of clustering in the distribution of PD. Geostatistical analysis, using semivariograms and kriging, was used to analyze the distribution of CVC in Bahia, Brazil (Laranjeira et al., 2008). In a recent study on unmanned aerial vehicles (UAVs) image data of olive trees affected by *X. fastidiosa* subsp. *pauca* in Italy, a geostatistical analysis incorporating a linear coregionalization model and multi-factor kriging was performed (Belmonte et al., 2023). Other methods for the analysis of spatial patterns, such as binomial dispersion index, Ripley's K function, average minimum distance, and autologistic models fitted for different neighboring structures, including temporal and spatial covariates, were used to study the spatial aggregation patterns of plum leaf scald in Sao Paulo State, Brazil (Ferreira et al., 2016).

Several modeling approaches have been employed to analyze the spread of *X. fastidiosa* with the incorporation of a spatially explicit structure. In Apulia, Italy, different studies on *X. fastidiosa* subsp. *pauca* spread in olive trees have been developed. White et al. (2017) used a spatially explicit spread model, with deterministic and stochastic kernels, which represented short-distance and long-distance dispersion, respectively. This study also incorporated control strategies such as buffer and eradication zones. Aniţa et al. (2021) proposed a spatially structured model based on an integro-partial differential system, including the spatial dynamics of susceptible and infectious vectors and the spatial dynamics of susceptible and infected olive trees, also including control strategies. Gilioli et al. (2023) proposed a spatially explicit mechanistic model by which the spatio-temporal dynamics of the disease and the vector were simulated based on the data available in this

region. This model considered the growth of the bacterium in the host, the dynamics of the vectors, their dispersal and the acquisition of the pathogen.

For Corsica (France), Abboud et al. (2019) used a mechanistic-statistical approach based on parabolic partial differential equations, considering spatial heterogeneity and a Bayesian inference approach with the aim of dating and locating the origin of the introduction of *X. fastidiosa*. For this region as well, Martin et al. (2021) modeled pathogen and vector dynamics of the pathogen using a spatially explicit compartmental model composed of a Susceptible–Exposed–Infected (SEI) sub-model for hosts and a Susceptible–Infected (SI) sub-model for vectors.

Regarding the models related to the spread of *X. fastidiosa* in Spain, Strona et al. (2020) modeled the spread through spatial networks connecting olive orchards in Andalusia based on geographic proximity and a Susceptible-Infected-Removed (SIR) model. Camino et al. (2021) combined remote sensing information from almond trees in Alicante with a stochastic spatial spread model to improve detection of infected trees in the early stages of disease caused by *X. fastidiosa* subsp. *multiplex*. Moreover, Monte Carlo methods have been implemented to study the risk of infection risk due to proximity in almond trees in California (Sisterson et al., 2012), and to characterize the spatio-temporal pattern of transmission of pecan bacterial leaf scorch disease (Li et al., 2011). On a global scale, from *X. fastidiosa* occurrence data and potential insect vectors, Yoon and Lee (2023) analyzed dispersal patterns based on the distances between occurrence points.

Chapter 2

Contemporary approaches in spatial modeling

Epidemiology can be defined as "the study of the spread of diseases, in space and time, with the objective to trace factors that are responsible for, or contribute to, their occurrence" (Diekmann and Heesterbeek, 2000). It involves the three components of the classic disease triangle: host, pathogen, and environment. Thus, in epidemiological models, these three components are implicitly interrelated and often include epidemic dynamics and rates of change as well as the intensity or magnitude of the disease (Madden et al., 2007). These models are useful for understanding how diseases spread and affect populations. The development of new methods and computational advances has made epidemic modeling a useful tool for estimating the dynamics of plant diseases, assessing risks, predicting their spread, and optimizing plant health surveillance programs and control strategies. To understand, analyze, and control plant diseases, it is important to consider several aspects, including environmental conditions, transmission pathways, human interventions in the spread of pathogens, and the distribution and density of host plants (Cunniffe et al., 2015). The increased availability of geo-referenced information and the development of geographic information systems (GIS) and software for this type of data have encouraged research on the spatial characteristics of diseases.

This chapter is devoted to present the basis used in this Thesis for plant disease modeling and spatial data analysis. In particular, the basis of disease dynamics modeling are presented. The analysis of spatial data using Bayesian approach is also presented, which deals relatively easily with the specification of complex models. Furthermore, the integrated nested Laplace approximation (INLA) to perform Bayesian inference is also introduced, along with specific aspects of spatial modeling under this approach.

2.1 Disease dynamics modeling

Disease spread models are motivated by the need to better understand the dynamics of disease transmission, predict the disease spread in a population, and assess the impact. These models also serve as tools for exploring ways in which to address disease transmission, including the evaluation of strategies to improve surveillance and control.

As in the epidemiology of human diseases, compartmental models are widely used for the modeling of plant disease epidemics (Jeger et al., 2000). The formulation of these models was initially developed by Kermack and McKendrick (1927) in the field of medical epidemiology. However, Van der Plank (1963) is recognized for the development of theory underlying this type of models in plant disease epidemics, with no apparent influence from what was developed in medical epidemiology (Jeger et al., 2000; Madden et al., 2007). Compartmental models assume that the population can be classified into different compartments according to disease status. The classical compartmental model comprises three compartments representing susceptible, infected, and removal individuals (*SIR* model), but there are many variants with additional compartments. This approach involves explicitly representing disease status categories and deriving linked equations for each of these. The *SIR* model can be represented schematically as shown in Figure 2.1, where the transition from compartment S to I is determined by the transmission rate (β), and the transition from I to R by the removal rate (γ). This model can be described mathematically by a set of ordinary

differential equations (ODEs) as:

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS, \\ \frac{dI}{dt} &= \beta IS - \gamma I, \\ \frac{dR}{dt} &= \gamma I. \end{aligned} \tag{2.1}$$

FIGURE 2.1: Schematic representations for the *SIR* (susceptible, infected, and removed) model. β transmission rate and γ removal rate.

Compartmental models analyze changes in compartments in the global population but without specifying the identities of individuals moving from one compartment to another. In contrast, individual-based models (IBMs) work from the perspective of the characteristics of each individual, as well as their interactions with other individuals. Thus, the disease dynamics (or event under study) emerge from the interactions between independent individuals and their environment (DeAngelis and Grimm, 2014). Therefore, IBMs are useful for simulating phenomena characterized by stochasticity and/or heterogeneity (Willem et al., 2017). Nevertheless, the main limitation of these models is the high computational cost associated with a relatively large population size or the simulation of long time periods (Keeling and Rohani, 2008).

As in compartmental models, IBMs for disease spread modeling start from the classification of individuals according to disease status. Therefore, the different states to be considered and the interactions between them also determine the complexity of the model. In addition to the possible disease states to be considered, there are several aspects to be taken into account when modeling disease dynamics, as the scale of the units of interest, the

spatial heterogeneity, or the interaction between individuals, as discussed in the following.

Despite the name individual-based model, these can be applied at different scales, which will determine the units of interest for the simulation of disease spread (Kirkeby et al., 2021). The scale often depends on the information available or the objective of the study. For instance, the units of interest can be plants, plant parts or crop plots. IBMs can also be used with data aggregated in areas or cells of a grid. In this case, the interactions between different areas can be established by means of the neighborhood relationship. Although the use of a larger scale can produce a bias in the dynamics, it can greatly reduce the computational time (Keeling and Rohani, 2008).

Often, disease spread models ignore spatial heterogeneity in the distribution of individuals, assuming that the population is homogeneously distributed. When there is spatial heterogeneity, groups of hosts can be found at considerable distances from the source of infection, resulting in epidemiological isolation. Consequently, the risk of infection within these isolated groups decreases (Keeling and Rohani, 2008). Addressing this issue is especially critical in disease modeling. IBMs provide an alternative to address this problem, as they can explicitly account for the spatial distribution of individuals and capture the disease dynamics accounting for spatial heterogeneity.

In the study of disease dynamics it is important to establish the interaction behavior between hosts, which in the case of plant diseases, is through the associated transmission way. This type of interaction usually decreases with distance, and it is common to use a transmission (or dispersal) kernel to integrate the reduction of transmission as a function of distance. A transmission kernel is a probability distribution function based on distance. There are different functions for transmission kernels, Nathan et al. (2012) reviewed them in detail. One of the characteristics of transmission kernel functions is their shape and tail thickness, and the choice of the suitable kernel function will depend on the transmission distance of the disease, i.e., whether the disease is spread over a short or long distance. Because of this, one of the challenges is to estimate the shape and parameters of the kernel,

which will depend on the particular disease (Keeling and Rohani, 2008). In Chapter 5 of this thesis, we propose as an alternative the incorporation of a spatial correlation function to take into account the interaction between individuals. Specifically, the Matérn correlation function defined in equation (2.5) is used.

Following Keeling and Rohani (2008) in an IBMs for plant diseases spread, the probability of a susceptible individual i becoming infected can be defined through the force of infection φ_i as:

$$\begin{aligned} P_i &= 1 - \exp(-\varphi_i), \\ \varphi_i &= \beta_i \sum_j \rho_{ij}, \end{aligned} \tag{2.2}$$

where φ_i depends on the transmission rate β and the interaction between the susceptible individual i and infected individuals j.

2.2 Spatial analysis in plant diseases

The spatial patterns of plant diseases are mainly determined by the relationship between the disease status of neighboring plants, which is influenced by the dispersal process and environmental factors (Madden et al., 2007). Thus, analysis of spatial patterns can be useful for developing hypotheses about underlying processes or obtaining quantitative information about disease dynamics (Ristaino and Gumpertz, 2003). Additionally, the study of spatial patterns can be informative for the design of surveillance methods or management strategies. The probability of an observation unit being diseased depends on the pathogen, the hosts, and the environment (Madden et al., 2007). Both individual and aggregated data sources can be used to study spatial patterns in plant diseases by assessing the spatial distribution within sampling units and spatial dependence between sampling units (Maanen and Xu, 2003).

Often, one of the objectives of studying plant disease epidemics is to predict the distribution of the pathogen or disease in a specific area based on environmental data. These models are usually referred to as species

distribution models (SDMs). These models are usually built from data on the presence or abundance of a particular species, with the objective of studying how environmental conditions are linked to the phenomenon. SDMs are widely used to associate the geographic settings of species with biotic and abiotic factors, develop risk maps for the potential establishment of pathogens, define favorable areas for the expansion of populations, and predict the distribution of species in space and time under climate change scenarios, among others (Martínez-Minaya et al., 2018a). SDMs can be developed with different methodologies, such as generalized linear models (GLM), generalized additive models (GAM), neural networks, maximum entropy models (e.g. Maxent) and climate envelope models (e.g. Bioclim). The literature available on the applications and methodologies of SDMs is quite extensive, with some reviews such as Guisan and Zimmermann (2000), Elith and Leathwick (2009) or Martínez-Minaya et al. (2018a) that compile and describe the different modeling approaches. However, when the spatial dependence of data is ignored, the degree of uncertainty can be largely underestimated, generating imprecise parameter estimations and providing relatively low predictive capacity (Latimer et al., 2006; Martínez-Minaya et al., 2018b). Advances in computational methods have made it possible to implement more complex models, and hence, a more straightforward incorporation of spatial dependencies in SDMs (Blangiardo and Cameletti, 2015). Among these advances, Bayesian hierarchical models allow random effects and complex dependency structures to be incorporated easily taking into account all non-observed uncertainties (Banerjee et al., 2004; Blangiardo and Cameletti, 2015).

In spatial statistics, data are realizations of a stochastic process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^d\}$, where Z is the observed process and $\boldsymbol{s}$ is the location of the observation. Three types of spatial data are traditionally distinguished, depending on the problem to be studied and the characteristics of the data: areal data, geostatistical data, and point patterns (Cressie, 1993). In *areal data*, or *lattice data*, the domain $\mathcal{D}$ is fixed and divided into a finite number of spatial units. Areal data refer to irregular divisions, such as administrative regions, and lattice data to regular spatial units. The interest in the analysis of these data is usually in smoothing or mapping a result over the spatial domain $\mathcal{D}$. In *geostatistical data*, or *point-referenced*, the domain

$\mathcal{D}$ is continuous and fixed, $\boldsymbol{s}$ varies continuously over $\mathcal{D}$, and $Z(\boldsymbol{s})$ can be observed at any point of $\mathcal{D}$. The interest in this context usually lies in the prediction of unobserved locations in $\mathcal{D}$. Unlike previous ones, in *point patterns*, the spatial domain $\mathcal{D}$ is a set of points in $\mathbb{R}^d$ where some events occur, and $Z(\boldsymbol{s})$ represents the occurrence of an event. If additional information on the covariates is provided, this is referred to as a marked point pattern process. In this case, the interest is usually to study whether the observed event presents a clustering pattern or not. In this Thesis we will focus on lattice data and geostatistics, where specific modeling methods for both types of data are used in Chapter 3 and Chapter 4.

2.2.1 Areal data

Areal or lattice data are widely used in disease mapping to estimate the disease risk over the areas. Commonly, the spatial neighborhood relationship between the n areas is provided by the sparse adjacency matrix W, with non-zero entries at $w_{i,j}$, $i, j \in \{1, ..., n\}$, when the areas i and j are neighbors (denoted as $i \sim j$) and 0 otherwise. The spatial correlated random effects are can be modeled using a multivariate Gaussian distribution with zero mean and precision matrix τQ, where τ is a precision hyperparameter, and Q is based on the adjacency matrix W.

There are different ways to define spatially correlated random effects through adjacency matrix. One of the common ways of modeling the spatial random effects $\boldsymbol{u} = \{u_1, ..., u_n\}$ is by an intrinsic autoregressive conditional model (ICAR), known as the Besag model (Besag et al., 1991), where the conditional distribution for u_i is specified as:

$$u_i | \boldsymbol{u_{-i}}, \tau_u \sim \mathcal{N}\left(\frac{1}{k_i}\sum_{i\sim j} u_i, \frac{1}{\tau_u k_i}\right), \; i \neq j, \tag{2.3}$$

where τ_u is the precision of the random effect and k_i the number of neighbors of area i.

As an extension of this model that only accounts for similarities between areas, Besag et al. (1991) proposed the addition of an unstructured spatial

random effect to account for independent area-specific noise, known as the Besag, York and Mollié (BYM) model (Besag et al., 1991). In this model, the spatial effect is the sum of an structured spatial ($\boldsymbol{u}$) and unstructured ($\boldsymbol{v}$) components, so that $\boldsymbol{b} = \boldsymbol{u} + \boldsymbol{v}$, with $\boldsymbol{u} \sim ICAR(\tau_u)$ and $\boldsymbol{v} \sim \mathcal{N}(0, \tau_v)$. However, the structured and unstructured components are not identifiable and therefore cannot be seen independently. To deal with this problem, different variants of the BYM model have been proposed, such as those described by Leroux et al. (2000) or Dean et al. (2001). A reparameterization of the BYM model (BYM2) has been proposed that accounts for scaling and addresses the identifiability issue (Riebler et al., 2016; Simpson et al., 2017). Here, the spatial random effect is reparameterized as:

$$\boldsymbol{b} = \frac{1}{\sqrt{\tau_b}} \left(\sqrt{1-\phi}\boldsymbol{v} + \sqrt{\phi}\boldsymbol{u}_* \right), \tag{2.4}$$

where $\boldsymbol{u}_*$ is the scaled ICAR model, and $\boldsymbol{v}$ is the vector of unstructured component. This model includes a single precision parameter $\tau_b > 0$ wich controls for the marginal variance of the weighted sum of $\boldsymbol{u}_*$ and $\boldsymbol{v}$, and a mixing parameter $0 \leq \phi \leq 1$ for the proportion of the marginal variance explained by the spatial structured effect. Therefore, $\phi = 0$ will indicate only the effect of the unstructured component, and only a spatially structured model when $\phi = 1$.

2.2.2 Geostatistical data

Geostatistical or point-referenced data are observations of a continuous spatial process that varies within a fixed domain. This spatial process is referred to as a Gaussian field (GF) if the vector $u(\boldsymbol{s})$ in the locations $\boldsymbol{s}$ follows a multivariate Normal distribution with mean $\boldsymbol{\mu}$ and spatially structured covariance matrix $\boldsymbol{\Sigma}$. The covariance matrix $\boldsymbol{\Sigma}$ is determined by a covariance function C as $\Sigma_{ij} = \text{Cov}(u(\boldsymbol{s}_i), u(\boldsymbol{s}_j)) = C(u(\boldsymbol{s}_i), u(\boldsymbol{s}_j))$, which represents the dependence between the random variables at two points $\boldsymbol{s}_i, \boldsymbol{s}_j$.

This random process is usually assumed to be weakly stationary, referred to as second-order stationarity, which requires that the process has a constant mean, i.e, $\mu(\boldsymbol{s}_i) = \mu$ for each location i, and that the covariances depend only on the differences between locations, i.e., $\text{Cov}(u(\boldsymbol{s}_i), u(\boldsymbol{s}_j)) =$

$C(\boldsymbol{s}_i - \boldsymbol{s}_j)$. Moreover, it is usual to assume that the process is isotropic, this implies that the covariance function depends only on the Euclidean distance between the locations $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$, and not on the direction.

Several covariance functions have been proposed when the process is stationary and isotropic (Banerjee et al., 2004). One of the most widely used in many fields is the Matérn covariance function due to its flexibility (Guttorp and Gneiting, 2006). The Matérn covariance function for two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$ at Euclidean distance $\|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ is defined as:

$$C(\boldsymbol{s}_i, \boldsymbol{s}_j) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa\|\boldsymbol{s}_i - \boldsymbol{s}_j\|)^\nu K_\nu(\kappa\|\boldsymbol{s}_i - \boldsymbol{s}_j\|), \tag{2.5}$$

where σ^2 is the marginal variance of the spatial field, Γ is the Gamma function and K_ν is the modified Bessel function of second kind and order $\nu > 0$, which is the smoothness parameter, where for $\nu = 1/2$ the Matérn covariance function is equivalent to the exponential covariance function. κ is a scale parameter related to the range parameter r, the distance at which the correlation between two locations is approximately zero, where $r = \sqrt{8\nu}/\kappa$.

However, the so-called "big n problem" is a recognized disadvantage of the modeling approach that utilizes the spatial covariance function, which arises from the computational costs required for performing algebra operations with dense covariance matrices (Banerjee et al., 2004). Specifically, dense matrix operations scale cubically with the size of the matrix, which is determined by the number of locations. Consequently, as the number of locations increases, these operations become computationally expensive. To address this issue, a computationally effective alternative is the stochastic partial differential equation (SPDE) approach proposed by Lindgren et al. (2011), which is introduced in section 2.4.

2.3 The Bayesian framework and INLA

Bayesian inference has experienced a resurgence in recent years due to computational advances, being a tool that allows to deal with simple models up to the incorporation of very complex structures. The Bayesian approach

involves treating both the observed data and parameters as random variables, where uncertainty is expressed in terms of probability distributions. An advantage of Bayesian models is that, in addition to the information provided by the data, they allow the incorporation of additional or "prior" information about the parameter' (Clark and Gelfand, 2006). such as knowledge from experts or previous studies. This can be especially useful when the information provided by the data is scarce and expert knowledge or information from previous studies is available, as well as to update the model based on new sources of information. Furthermore, the hierarchical structure of this approach makes it possible to combine data from multiple sources (Banerjee et al., 2004).

In the Bayesian paradigm, parameters are random variables whose randomness accounts for the uncertainty about them. The probability distribution, i.e, likelihood, $p(\boldsymbol{y}|\boldsymbol{x})$ comes from the observations $\boldsymbol{y} = (y_1, ..., y_n)$ given a vector of unknown parameters $\boldsymbol{x}$. The inference on $\boldsymbol{x}$ is based on the posterior distribution through Bayes' theorem as:

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}, \boldsymbol{x})}{p(\boldsymbol{y})} = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{\int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}},$$

where the denominator $\int p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$ defines the marginal distribution of the data $\boldsymbol{y}$. Since it does not dependent on $\boldsymbol{x}$, it can be fixed as a constant in the posterior distribution of the parameters. Thus, the posterior distribution is often expressed as:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}).$$

To interpret the results on the posterior distribution of the parameters, it is usually summarized by different statistics, such as the posterior mean, median, or mode, which are point estimates. In addition, posterior distributions can also be summarized by intervals instead of point estimates, which in the Bayesian context are called credible intervals and represent the probability that the parameter lies within that interval. This allows to make probabilistic statements and to account for the underlying uncertainty.

Structures such as spatial autocorrelation can be taken into account in

a regression model by incorporating random effects that capture the inherent spatial dependence of the data. The Bayesian approach using a hierarchical structure is very convenient for the incorporation of spatial or spatio-temporal effects correlated by hierarchical specification of prior distributions. The first state or level of this structure is given by the likelihood of the data $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta_1})$, where the observed data $\boldsymbol{y}$ are conditional on the parameters $\boldsymbol{x}$ and hyperparameters $\boldsymbol{\theta_1}$; in the second state the parameters are modeled through a prior probability distribution $p(\boldsymbol{x}|\boldsymbol{\theta_2})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ is a vector of hyperparameters; and a third state defines the prior distributions of the hyperparameters, i.e., hyperpriors, $p(\boldsymbol{\theta})$.

The main interest is to obtain $p(\boldsymbol{x}|\boldsymbol{y})$, however this usually involves multidimensional integration, which makes it a complex process. Markov chain Monte Carlo (MCMC) methods have traditionally been used to solve this issue. Moreover, the software developed for these methods, like `WinBUGS` (Lunn et al., 2000), `JAGS` (Plummer, 2003), `Stan` (Carpenter et al., 2017), `BayesX` (Umlauf et al., 2015) or `NIMBLE` (de Valpine et al., 2017), have made their use easier and have had a great impact in several fields. These computational methods are a class of sampling methods rely on produce samples from a Markov chain, whose stationary distribution is the same as the posterior. Although they enable Bayesian inference for complex models, they can be computationally demanding and may involve convergence issues (Clark and Gelfand, 2006).

The integrated nested Laplace approximation (INLA) (Rue et al., 2009) is a computationally less expensive alternative to the above. A large variety of models are available for implementation through the R package `R-INLA` (`www.r-inla.org`). The INLA methodology and its application with different kind of models are described in detail in the literature, some of these works are presented among others by Blangiardo and Cameletti (2015), Rue et al. (2017), Bakka et al. (2018), Krainski et al. (2019) and Gómez-Rubio (2020).

2.3.1 The integrated nested Laplace approximation (INLA)

The INLA methodology is designed to perform Bayesian inference with latent Gaussian models (LGM), which comprise a wide range of commonly used models. These models have the general form of a linear additive model, where the mean μ_i of the n observations $\boldsymbol{y} = (y_1, ..., y_n)$ is linked to a structured additive prediction η_i through a convenient link function $g(\mu_i) = \eta_i$, which includes fixed effects and random effects:

$$\eta_i = \beta_0 + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_f} f^{(k)}(u_{ki}) + \varepsilon_i;\ i = 1, ..., n,$$

where β_0 is the intercept, β_j, $j = 1, ..., n_\beta$ the coefficients of covariates $\boldsymbol{z}$, $f^{(k)}$ are unknown functions for n_f random effects of the covariates $\boldsymbol{u}$, which can be defined to model, for example, spatial or spatio-temporal dependence, and ε_i is an error term. This general model can be represented by a hierarchical structure with the following components:

Likelihood:	$y_i \vert \boldsymbol{x}, \boldsymbol{\theta_1} \sim p(y_i \vert x_i, \boldsymbol{\theta_1}), i = 1, ..., n,$
Latent Gaussian field:	$\boldsymbol{x} \vert \boldsymbol{\theta_2} \sim N(\mu(\boldsymbol{\theta_2}), \boldsymbol{Q}^{-1}(\boldsymbol{\theta_2})),$
Hyperpriors:	$\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) \sim p(\boldsymbol{\theta}),$

where the observations $\boldsymbol{y}$ depend on the latent field $\boldsymbol{x}$ and on a vector of hyperparameters $\boldsymbol{\theta_1}$. The vector of latent effects is represented as $\boldsymbol{x} = \{\boldsymbol{\eta}, \beta_0, \boldsymbol{\beta}, \boldsymbol{f}\}$. The distribution of $\boldsymbol{x}$ is assumed to be a Gaussian Markov random field (GMRF), with zero mean and precision matrix $\boldsymbol{Q}(\boldsymbol{\theta_2})$, where $\boldsymbol{\theta_2}$ is a vector of hyperparameters, and $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ is the vector of all hyperparameters.

The computational advantages of INLA are mainly based on three aspects: i) it is focused on obtaining the individual posterior marginal distributions $p(x_i|\boldsymbol{y})$ and $p(\theta_j|\boldsymbol{y})$, instead of estimating the multivariate joint posterior distributions; ii) the models can be expressed as GMRF, which offer computational advantages due to their properties (Rue and Held, 2005); iii) the approximation of the marginal distributions of the parameters through Laplace approximations.

The posterior marginals of each element x_i of the latent field $\boldsymbol{x}$, and for hyperparameters θ_j, can be defined as:

$$p(x_i|\boldsymbol{y}) = \int p(x_i|\boldsymbol{\theta}, \boldsymbol{y}) p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta},$$
$$p(\theta_j|\boldsymbol{y}) = \int p(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}_{-j}.$$

The posterior marginals $p(x_i|\boldsymbol{y})$ and $p(\theta_j|\boldsymbol{y})$ are approximated by nested formulation, through analytical approximations to the full conditionals $p(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ and $p(\boldsymbol{\theta}|\boldsymbol{y})$ and numerical integration. The joint posterior of the hyperparameters $p(\boldsymbol{\theta}|\boldsymbol{y})$ is approximated by the Gaussian approximation of the full posterior of the latent field $\tilde{p}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$, evaluated at the posterior mode $\boldsymbol{x}^*(\boldsymbol{\theta})$:

$$\tilde{p}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \left. \frac{p(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{p}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} \right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}.$$

Then, the posterior marginals of x_i and θ_j is obtained through a numerical integration as:

$$\tilde{p}(x_i|\boldsymbol{y}) = \sum_k \tilde{p}(x_i|\boldsymbol{\theta}_k, \boldsymbol{y}) \times \tilde{p}(\boldsymbol{\theta}_k|\boldsymbol{y}) \times \Delta_k,$$

where Δ_k are the weights corresponding to the vector of values $\boldsymbol{\theta}_k$.

The posterior marginals $p(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ can be approximated in three different ways, by a Gaussian, a Laplace or a simplified Laplace approximation (Rue et al., 2009). The Gaussian approximation is a fast method since the result is derived from $\tilde{p}_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$, but the results may not be accurate in terms of the posterior mean location and skewness. The Laplace approximation is more accurate but can be relatively costly. The simplified Laplace approximation is considered the most efficient, as it is computationally fast and corrects for the location and skewness of the Gaussian approximation.

2.4 Spatial models under INLA methodology

As mentioned above, INLA exploits the computational advantages provided by GMRFs. When working with areal data, neighborhood-based CAR models satisfy the Markov property, that is, x_i and x_j are conditionally independent given all the other components $\boldsymbol{x}_{-i,j}$. Thus, the corresponding element of the precision matrix $\boldsymbol{Q}$ will be zero if areas i and j are not neighbors, i.e., $x_i \perp x_j | \boldsymbol{x}_{-i,j} \iff Q_{ij} = 0$. This specification is a GMRF (Rue and Held, 2005), and results in a sparse matrix $\boldsymbol{Q}$, which yields great computational advantages.

To deal with continuous indexed GFs, Lindgren et al. (2011) proposed the stochastic partial differential equation (SPDE) approach, which consists of representing the GF using a discretely indexed spatial random process (i.e., a GMRF). This approach allows to handle the "big n problem", as it reduces the computational cost from a magnitude of order $O(n^3)$ to $O(n^{3/2})$.

2.4.1 SPDE approach

The SPDE approach allows to approximate a continuous spatial process in a computationally efficient way using a GMRF representation of the GF. In this approach, a GF $u(\boldsymbol{s})$ with a Matérn covariance function is a solution to the SPDE:

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau u(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \tag{2.6}$$

where $\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$ is the Laplacian operator, α controls the smoothness, τ controls the variance, and $\mathcal{W}(\boldsymbol{s})$ is a Gaussian spatial white noise process. This SPDE is related with the parameters to the Matérn covariance function (equation 2.5) through the smoothness parameter ν and the marginal variance σ^2 as:

$$\begin{aligned} \nu &= \alpha - d/2, \\ \sigma^2 &= \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}. \end{aligned} \tag{2.7}$$

A solution to the SPDE can be obtained using the finite element method through a basis function representation defined on a triangulation of the spatial domain $\mathcal{D}$, resulting in a mesh. This approximation can be represented as:

$$u(\boldsymbol{s}) = \sum_{k=1}^{m} \psi_k(\boldsymbol{s}) w_k, \tag{2.8}$$

where $k = 1, ..., m$ with m the number of vertices in the triangular mesh, ψ_k are the basis functions, and w_k are zero mean Gaussian distributed weights. Assigning a Gaussian distribution to the joint distribution of the weight vector, with zero mean and precision matrix $\boldsymbol{Q}$, approximates the solution $u(\boldsymbol{s})$ of the SPDE. Further details on this solution can be found at Bakka et al. (2018) and Krainski et al. (2019).

2.4.2 Non-stationarity

Spatial models will usually consider the spatial process to be stationary (i.e., the spatial effect is invariant to the map translation) and isotropic (i.e., the spatial effect is invariant to the map rotation), that is, the autocorrelation between two locations only depends on the Euclidean distance. However, assuming stationarity and isotropy of the spatial process is not always correct. Several approaches have been proposed to deal with non-stationary GFs. Sampson and Guttorp (1992) were pioneers in dealing with this issue using covariance functions based on the spatial deformation approximation. Since then, different approximations have been proposed. Some of these proposals include the Bayesian approach based on the method by Sampson and Guttorp (1992) (Damian et al., 2001, 2003; Schmidt and O'Hagan, 2003; Schmidt et al., 2011).

Under the SPDE approach presented by Lindgren et al. (2011) detailed in the previous section, the modeling has been extended to the non-stationary version (Lindgren et al., 2011; Bolin and Lindgren, 2011). In this approach the non-stationary model is defined by the spatial variation of the parameters $\kappa(\boldsymbol{s})$ and $\tau(\boldsymbol{s})$, rewriting the SPDE of equation (2.6) as:

$$(\kappa(\boldsymbol{s})^2 - \Delta)^{\alpha/2}(\tau(\boldsymbol{s})u(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^2. \tag{2.9}$$

A special case of the non-stationary approximation using the SPDE approach is the *barrier model* proposed by Bakka et al. (2019). In this approach the spatial process is interrupted by barriers, hindering the spatial dependence between locations that are at a close Euclidean distance. The underlying concept lies in approximating the range parameter r of the Matérn covariance function close to zero (i.e., κ close to ∞) in the areas covered by the barriers (referred to as barrier area), and a constant r value in the remaining spatial domain (referred to as normal area). This method provides a solution to non-stationary GFs with no increase in the computational cost and no additional complexity in the implementation compared to the stationary approach.

For a better interpretation of the SPDE, a reparametrization of equation (2.6) is proposed, and fixing $\alpha = 2$, so that:

$$u(\boldsymbol{s}) - \nabla \cdot \frac{r^2}{8} \nabla u(\boldsymbol{s}) = r\sqrt{\frac{\pi}{2}}\sigma_u \mathcal{W}(\boldsymbol{s}), \tag{2.10}$$

where $u(\boldsymbol{s})$, $\boldsymbol{s} \in \mathcal{D} \subseteq \mathbb{R}^2$ is the GF, r is the range, σ_u is the marginal standard deviation, $\nabla = \left(\frac{\partial}{\partial_x}, \frac{\partial}{\partial_y}\right)$ and $\mathcal{W}(\boldsymbol{s})$ denotes white noise.

In this case a system of two SPDEs is introduced, one for the barrier area and one for the normal area. The correlation in the barrier area is eliminated through a different Matérn field, with the same σ_u, but r close to zero. In particular, the barrier GF $u(\boldsymbol{s})$ is the solution to:

$$\begin{aligned} u(\boldsymbol{s}) - \nabla \cdot \frac{r^2}{8} \nabla u(\boldsymbol{s}) &= r\sqrt{\frac{\pi}{2}}\sigma_u \mathcal{W}(\boldsymbol{s}), \text{ for } \boldsymbol{s} \in \Omega_n, \\ u(\boldsymbol{s}) - \nabla \cdot \frac{r_b^2}{8} \nabla u(\boldsymbol{s}) &= r_b\sqrt{\frac{\pi}{2}}\sigma_u \mathcal{W}(\boldsymbol{s}), \text{ for } \boldsymbol{s} \in \Omega_b, \end{aligned} \tag{2.11}$$

where Ω_n is the normal area and Ω_b is the barrier area. r and r_b are the spatial ranges for the normal and barrier areas, respectively.

The resulting fields have very few boundary effects compared to other approximations, which makes it a useful alternative for many practical situations (Lindgren et al., 2022).

Chapter 3

Spatial Bayesian modeling for lattice and geostatistical data

After the background and the required tools have been presented, this chapter presents our first study on the distribution of *X. fastidiosa*, where the main objective is to analyze the effects of climatic covariates accounting for the spatial dependence in two outbreaks in the Mediterranean basin.

3.1 Introduction

The current geographical distribution of *X. fastidiosa* comprises areas with different climate types (EFSA, 2019). Although the highest prevalence of the pathogen occurs in tropical and subtropical climates, it is also found in regions that are much colder and/or drier. Despite the relatively wide temperature range where *X. fastidiosa* develops, it should be noted that its performance at low temperatures depends largely on the interaction between the subspecies and the host plant. This would explain the differences in the geographical range and prevalence of the different *X. fastidiosa* subspecies on their associated hosts. For instance, in California, US, the severity of

almond leaf scorch is much lower than that of Pierce's disease in grapevines, although they are caused by the same subspecies of *X. fastidiosa* (Purcell, 1997). Similarly, in North America, *X. fastidiosa* subsp. *multiplex* is present in areas that are colder than those where *X. fastidiosa* subsp. *fastidiosa* is prevalent (EFSA, 2015). It is not known whether the geographic distribution of the different *X. fastidiosa* subspecies is associated with their ability to thrive at low temperatures, to the distribution and abundance of their host plants and vectors, or simply because they have not reached their maximum geographical and/or environmental extent.

In this context, species distribution models (SDMs) can be a useful tool to study the geographic range of *X. fastidiosa* given that they link spatial occurrence data with multivariate environmental data that can be used to estimate the relationship between the species and its habitat, and subsequently predict spatial occurrence in unsampled locations or time-periods (Martínez-Minaya et al., 2018a). As reviewed in Chapter 1, SDMs have been used in previous works to describe the relationship between *X. fastidiosa* and climatic factors. However, all these previous studies with SDMs for *X. fastidiosa* have two characteristics in common: they used presence-only data or generated pseudo-absences such as Maxent, and did not include spatial autocorrelation. Models based on presence-only data may be useful when absence data are not available (Franklin, 2010). However, without absence data the accuracy of the models can be overestimated (Brotons et al., 2004). In addition, they did not explicitly include spatial dependence, which may lead to an overestimation of the model parameters and thus inaccurate results (Latimer et al., 2006; Martínez-Minaya et al., 2018b).

In this study, the geographic distribution of *X. fastidiosa* was analyzed in two affected regions in Europe: Alicante (Spain) and the province of Lecce in the Salento peninsula (Apulia, Italy). These two study regions were selected due to them having different, but relatively simple scenarios with regard to the prevalent subspecies of *X. fastidiosa* and the main hosts affected. In Alicante, only *X. fastidiosa* subsp. *multiplex* ST6 was identified, mainly affecting almonds (Landa et al., 2020). On the other hand, in Apulia, only *X. fastidiosa* subsp. *pauca* ST53 was identified, mainly affecting olives (Saponari et al., 2013). The presence/absence data of *X. fastidiosa* were analyzed in each study region with Bayesian hierarchical

models, which allowed us to include different spatial dependencies of each dataset. Furthermore, in Bayesian statistics observations and parameters are considered as random variables and so their uncertainty can be incorporated in a natural way via Bayesian hierarchical models (Banerjee et al., 2004). Computational advances have made it possible to approximate the posterior distribution of the parameters involving these complex models by means of integrated nested Laplace approximation (INLA) (Rue et al., 2009). The primary scope of this study was to determine the influence of climatic variables on the geographic distribution of the pathogen, as well as the spatial relationship between the positive locations in each region. Results in the form of risk maps will help Plant Health Authorities to optimize the official delimiting surveys for *X. fastidiosa* as well as to implement control strategies, such as eradication or containment, as established by the Decision (EU) 2015/789.

3.2 Materials and methods

3.2.1 Database

As the kind of spatial data gathered defines the final hierarchical spatial model used, it is necessary to describe in detail the two different databases available. A georeferenced dataset was provided by the Plant Health Authority (*Sanitat Vegetal*) of the *Generalitat Valenciana*, including the results of the official delimiting surveys for *X. fastidiosa* in 2017 in the demarcated area in the province of Alicante, Spain. Surveillance (i.e., inspection and sampling) in the demarcated area was based on the specifications established by Decision (EU) 2015/789, according to which, a visual examination of the plants specified as susceptible and, in case of suspicion of infection by the pathogen, the collection of samples and laboratory testing. A total of 3203 samples were considered, 206 of them were positive (i.e., presence) for *X. fastidiosa* and 2997 were negative (i.e., absence) based on real-time PCR (EPPO, 2019b) (Figure 3.1a). Only *X. fastidiosa* subsp. *multiplex* ST6 was detected. Samples were taken from 57 different plant species, but all the positives were in almond (*Prunus dulcis*). Nevertheless, all sampled plant species were included in the analysis as they were considered susceptible to

X. fastidiosa (EFSA, 2020). The total number of samples was presented aggregated on the 1x1 km spatial grid used by *Sanitat Vegetal*. Non-sampled grid cells were removed, since most of them corresponded to mountain peaks with difficult access and/or the absence of host plants, thus the study area had an extension of 638 km^2.

Data on the distribution of *X. fastidiosa* in Apulia (Italy) were obtained from the official surveillance program conducted by the Plant Health Authority of the *Regione Puglia* from 2013 to 2018. Due to the different surveillance strategies each year, only data from the province of Lecce (2766 km^2) during the first sampling campaign (from November 2013 to December 2014) were considered for further analysis. Samples were first tested by enzyme-linked immunosorbent assay (ELISA), all samples yielding positive reactions were confirmed by a second round of assay using real-time PCR (EPPO, 2019b). The selected dataset included a total of 4205 samples, 224 of them were positive (i.e., presence) for *X. fastidiosa* and 3981 were negative (i.e., absence) (Figure 3.1b). Only *X. fastidiosa* subsp. *pauca* ST53 was detected. In this case, data were not presented aggregated on a grid structure as before. As a result, samples were considered as georeferenced observations in a continuous space.

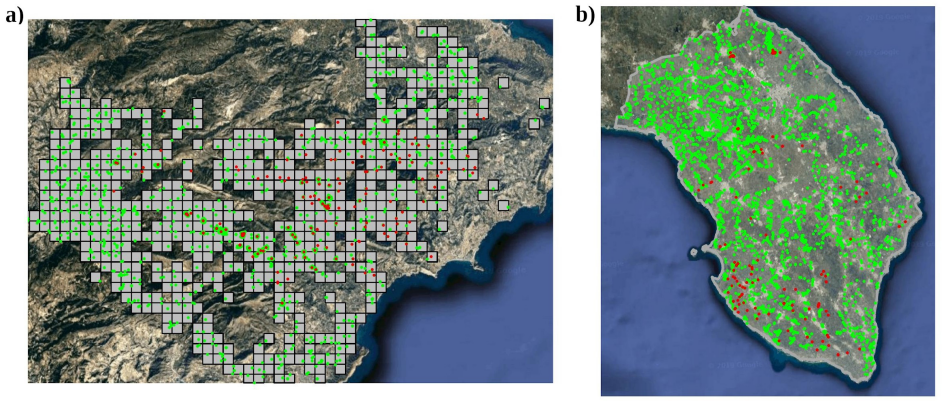

FIGURE 3.1: Presence (●) and absence (●) of *Xylella fastidiosa*. (**a**) Sampled grid cells (1 km^2) in the demarcated area in Alicante, Spain, in 2017; (**b**) sampling in the province of Lecce, Italy, during the 2013-2014 campaign.

Climatic data for the demarcated area in Alicante and the province of Lecce were obtained from the WorldClim v.2 database with a resolution of 30" arc sec (Fick and Hijmans, 2017). This database contains monthly average data of temperature and precipitation from 1970 to 2000 and 19 derived bioclimatic variables. In addition, accumulated degree-days (T_{base} 15 ℃) during the vegetative growth period of the main host species in each region were also considered, i.e., from February to October for almond in Alicante (Pou, 2004) and from April to October in olive in Lecce (Rallo and Cuevas, 2017). The UTM coordinate system was used in all the spatial datasets with the raster package for R software (Hijmans and van Etten, 2012; R Core Team, 2022).

For Alicante, the climatic variables considered relevant by Martinetti and Soubeyrand (2019) for explaining the presence of *X. fastidiosa* in areas of France where subsp. *multiplex* is prevalent were also included in the analysis. These variables were the average minimum temperature in winter from December to March (*tmin*), the average precipitation during the dry season from July and August (*precd*) and the solar radiation (*srad*).

Purcell's classification, based on the minimum winter temperature (*bio6*), was proposed in the USA to estimate the risk of PD (Anas et al., 2008). This classification consists of four categories: severe (> 4.5 ℃), moderate (1.7 to 4.5 ℃), occasional (-1.1 to 1.7 ℃), and negligible (<-1.1 ℃). A categorical variable considering these four categories based on the minimum temperature of the coldest month (*bio6*) in WorldClim v.2 was included in the analysis in both study regions.

Due to the nature of the climatic variables, high linear correlations were found among most of them (Appendix A, Figure A.1). In order to minimize potential problems of multicollinearity, a variable selection was performed based on the Pearson's correlation coefficient excluding pairs of variables with $|\rho| > 0.7$ (Dormann et al., 2013). Alternatively, principal component analysis (PCA) was conducted to reduce the number of variables and obtain new uncorrelated variables. Since the climatic variables had different metrics, PCA was performed based on the correlation matrix. The correlation of each variable with the principal components (PCs) was expressed by a rotation with the Varimax method. The variable temperature annual

range (*bio7*) was excluded from PCA because it is a linear combination of the variables maximum temperature of the warmest month (*bio5*) and minimum temperature of the coldest month (*bio6*).

3.2.2 Models

Two different Bayesian hierarchical spatial models were used to analyze the variation of the presence of *X. fastidiosa* in the study areas. Bayesian hierarchical models allow the incorporation of sources of variability and non-observed uncertainty. Nevertheless, computational methods, such as Markov chain Monte Carlo (MCMC) and integrated nested Laplace approximation (INLA) are generally required in order to obtain posterior distributions of the parameters and hyperparameters. In particular, the INLA methodology (Rue et al., 2009) is designed for latent Gaussian models (LGMs), a large class of models including the hierarchical spatial models used here, and provides accurate results in shorter computing times compared with MCMC (Blangiardo and Cameletti, 2015).

LGMs can also be considered as a particular case of the structured additive regression models (STAR), (Fahrmeir et al., 2013), where the mean of the response variable Y_i is linked to a structured predictor that accounts the various effects in an additive way:

$$\eta_i = \beta_0 + \sum_{j=1}^{n_\beta} \beta_j z_{ji} + \sum_{k=1}^{n_f} f^{(k)}(u_{ki}) + \varepsilon_i;\ i = 1, ..., n, \tag{3.1}$$

where η_i enters the likelihood through a link function, β_0 is the intercept of the model, β_j are the coefficients of covariates $\boldsymbol{z}$, $f^{(k)}$ define random effects in terms of the vector of covariates $\boldsymbol{u}$, and ε_i represents unstructured terms. The models which we deal with in this work include only fixed effects and in some cases a structured spatial term. The prior knowledge of the additive predictor is expressed using Gaussian prior distributions. In this context, all the latent Gaussian variables can be seen as components of a vector known as the latent Gaussian field. This class of models have several applications due to their flexibility and can be fitted using the INLA methodology with the R package `R-INLA`.

In both case studies, logistic regressions were performed. In the case of Alicante, as we considered the total number of positives of the total number of samples in each grid cell, the response variable (y_i) was assumed to be a Binomial distribution, i.e., $y_i \sim \text{Binomial}(n_i, \pi_i)$, where π_i is the probability of a sample being positive in the grid cell i, and n_i is the total number of samples in that grid cell. On the other hand, the response variable y_i in Lecce was assumed to follow a Bernoulli distribution where 0 indicates the absence and 1 the presence of *X. fastidiosa* at location i, that is, $y_i \sim \text{Bernoulli}(\pi_i)$, π_i being the probability of presence at location i. But as the samples were not collected in the same way, different structures for the spatial random effect were employed in each case.

3.2.2.1 Model for Alicante

Data in Alicante come from a sampling carried out in a georeferenced regular lattice. A common way to deal with the spatial dependence amongst the grid cells is to consider an intrisic Gaussian Markov random field (GMRF) model, also known as the Besag model presented by Besag (1974). The main idea is to construct a model where each random effect u_i conditionally to its neighbor random effects has a Gaussian distribution with a mean equal to the average of the neighbors and a precision proportional to the number of neighbors:

$$u_i \mid \boldsymbol{u_{-i}}, \tau_u \sim \mathcal{N}\left(\frac{1}{k_i}\sum_{i\sim j} u_j, \frac{1}{\tau_u k_i}\right), \quad i \neq j, \tag{3.2}$$

where τ_u is the precision of the random effect and k_i the number of neighbors corresponding to the grid cell i, which has a set of neighbors $j \neq i$. As some grid cells did not have any adjacent ones, the neighborhood relation was established at a distance of 2.5 km, that is, two grid cells were considered neighbors if the distance between their centroids was $\leq$ 2.5 km. Due to the 1×1 km resolution of the data, 2.5 km was the minimum distance resulting in at least one neighbor for all grid cells. However, this kind of effect only accounts for similarities between grid cells, and it does not take into account the individual variability of each grid cell. Then, the Besag,

York and Mollié model (Besag et al., 1991) was used, simply adding an independent random effect to the model. In this way, $\eta_i = \beta_0 + \boldsymbol{X}_i\boldsymbol{\beta} + u_i + v_i$, where β_0 is the intercept, $\boldsymbol{\beta}$ represents the effect of the covariates $\boldsymbol{X}_i$, u_i is the spatial effect and v_i is the independent random Gaussian effect. Simpson et al. (2017) proposed a reparameterisation of this model that allows an straightforward incorporation of penalized complexity priors (PC-prior) (Riebler et al., 2016; Simpson et al., 2017) to the model hyperparameters, including a standardised spatial component $\boldsymbol{u^*}$:

$$\eta_i = \beta_0 + \boldsymbol{X}_i\boldsymbol{\beta} + \frac{1}{\sqrt{\tau}}\left(\sqrt{1-\phi}\, v_i + \sqrt{\phi}\, u_i^*\right), \tag{3.3}$$

where τ controls the marginal variance of $\boldsymbol{u^*}$ and $\boldsymbol{v}$. In addition, it incorporates the mixing parameter $0 \leq \phi \leq 1$, which measures the proportion of variance explained by $\boldsymbol{u^*}$, so values close to 1 would imply a strong weight of the spatial component, while $\phi = 0$ only that of the independent random effect.

Consequently, the complete model with covariates, random effects and all the necessary prior distributions for the parameters and hyperparameters involved had the following structure:

$$\begin{aligned}
y_i &\sim \text{Binomial}(n_i, \pi_i),\ i = 1, ..., 638, \\
\text{logit}(\pi_i) &= \beta_0 + \boldsymbol{X}_i\boldsymbol{\beta} + \frac{1}{\sqrt{\tau}}\left(\sqrt{1-\phi}\, v_i + \sqrt{\phi}\, u_i^*\right), \\
\text{P}(\beta_0) &\propto 1, \\
\beta_m &\sim \mathcal{N}(\mu = 0, \tau = 10^{-3}),\ m = 1, \ldots, N_\beta, \\
\tau &\sim \text{PC-prior}(0.5/0.31, 0.01), \\
\phi &\sim \text{PC-prior}(0.5, 2/3),
\end{aligned} \tag{3.4}$$

where n_i is the total number of samples for each cell i; π_i is the probability that a sample taken in cell i is positive; $\boldsymbol{\beta}$ is the vector of coefficients of covariates $\boldsymbol{X}_i$, to which a non-informative prior has been assigned in the form of a Normal distribution with mean 0 and precision of 10^{-3}. The scaled spatial effect u_i^* is specified by an intrinsic conditional autoregressive (ICAR) distribution, and v_i is the independent random effect, where $v_i \sim N(0, \boldsymbol{I})$.

Following Simpson et al. (2017), a PC-prior for the precision τ was defined as $\mathrm{P}(1/\sqrt{\tau} > 0.5/0.31) = 0.01$. On the other hand, a prior distribution was assigned for the mixing parameter ϕ that assumes that the independent random effect explaining more variability than the spatial component, where $\mathrm{P}(\phi < 0.5) = 2/3$.

3.2.2.2 Model for Lecce

Another Bayesian hierarchical spatial model, more specifically a Bayesian geostatistical model, was used to analyze the presence of *X. fastidiosa* in Lecce. As mentioned above, this was based on the fact that data were not presented aggregated on a grid structure as in the other dataset, therefore samples were considered as georeferenced observations in a continuous space. In this case, the spatial dependence is expressed via the spatial effect $\boldsymbol{u}$ (geostatistical term) that is assumed to follow a multivariate Gaussian distribution whose covariance matrix $\sigma^2_{\boldsymbol{u}} H(r)$ depends on the distance between locations, and the hyperparameters $\sigma^2_{\boldsymbol{u}}$ and r, the variance, and the range of the spatial effect, respectively. In order to fit and predict using this kind of model, where an indexed continuous Gaussian field (GF) is included in the formula, Lindgren et al. (2011) proposed an explicit link between GMRF and GF with a Matérn covariance structure via a weak solution to a stochastic partial differential equation (SPDE). With this approximation, the spatial term is reparametrized as $\boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{Q}^{-1}(\kappa, \tau))$, depending on two different parameters, κ and τ. More precisely, the range is approximately $r = \sqrt{8}/\kappa$ and the variance is $\sigma^2_{\boldsymbol{u}} = \frac{1}{4\pi\kappa^2\tau^2}$ (Lindgren et al., 2011). However, instead of using the default parametrization, Krainski et al. (2019) recommended using the one which is more intuitive to control the parameters through the marginal standard deviation and the range.

Lastly, we specified prior distributions for the parameters and hyperparameters. In particular, Normal vague priors with mean and precision were used for the regression coefficients. Following Fuglstad et al. (2019), PC-priors were used for the range and the standard deviation of the spatial field.

Taking all this into account, the final model with covariates, random effects, and all the necessary prior distributions for the parameters and hyperparameters involved had the following structure:

$$\begin{aligned} y_i &\sim \text{Bernoulli}(\pi_i),\ i = 1, ..., 4205, \\ \text{logit}(\pi_i) &= \boldsymbol{X}_i\boldsymbol{\beta} + u_i, \\ \\ \boldsymbol{u} &\sim \mathcal{N}(0, \boldsymbol{Q}^{-1}(r, \sigma_{\boldsymbol{u}})), \\ \text{P}(\beta_0) &\propto 1, \\ \beta_m &\sim \mathcal{N}(\mu = 0, \tau = 10^{-3}),\ m = 1, \ldots, N_\beta, \\ \\ r &\sim \text{PC-prior}(\mu_r, 0.5), \\ \sigma_{\boldsymbol{u}} &\sim \text{PC-prior}(1, 0.5), \end{aligned} \tag{3.5}$$

where π_i is the probability of the presence of the pathogen in the location i and $\boldsymbol{\beta}$ is the vector of the coefficients of the covariates $\boldsymbol{X}_i$ and $\boldsymbol{u}$ is the spatial effect. The PC-priors were defined as $\text{P}(r < \mu_r) = 0.5$ and $\text{P}(\sigma > 1) = 0.5$ for the range and standard deviation, respectively, where μ_r was chosen as 50% of the diameter of the geographic region under study (Fuglstad et al., 2019).

3.2.3 Model selection

Given the large number of models resulting from all the possible combinations of covariates, model selection was carried out. Pearson's correlation coefficients among covariates were previously calculated to assist in variable selection and minimize potential problems of multicollinearity. If the correlation between two variables was greater than 0.7, one of those covariates was taken out of the analysis (Dormann et al., 2013). With the resulting covariates, all possible 2^k (where k represents the number of components of the model: covariates and the random effects) models were evaluated and the best one was chosen according to information criteria (Heinze et al., 2018). In this work, we used the Watanabe Akaike information criterion (WAIC) (Gelman et al., 2014; Watanabe, 2010), which is the sum of two

components, one quantifying model fit and other evaluating model complexity. The predictive capacity of the models was evaluated by cross validation using the logarithmic conditional predictive ordinate (LCPO) (Pettit, 1990; Roos and Held, 2011). Models with the lowest values of WAIC and LCPO were selected. When several models presented similar information criteria, the parsimony criterion was applied and models with fewer covariates were selected.

3.3 Results

3.3.1 Alicante

All the 23 climatic variables included in the analysis showed high linear correlation (Appendix A, Figure A.1a). Nevertheless, annual mean temperature (*bio1*), temperature annual range (*bio7*), and precipitation of the wettest month (*bio13*) had $|\rho| < 0.7$. These covariates presented low variability in the study area, minimum and maximum values of *bio1* were 10.49 ℃ and 17.69 ℃, respectively. The covariate *bio7* varied between 24.70 ℃ and 30.93 ℃, and *bio13* between 46 mm and 67.31 mm. Furthermore, the distribution of *X. fastidiosa* in the study area was not coincident with those covariates (Appendix A, Figure A.2). Despite the relative climatic homogeneity, the four categories defined by Purcell based on the minimum winter temperature were represented in the study area. Nevertheless, *X. fastidiosa* was detected in all of them. Moreover, the proportion of positive samples in the area for each category was similar with 1.78% in the severe category (> 4.5 ℃), 8.28% in the moderate category (1.7 to 4.5 ℃), 6.03% in the occasional category (-1.1 to 1.7 ℃), and 2.29% in the negligible category (<-1.1 ℃) (Appendix A, Figure A.2d).

All model combinations (n = 32) with the three selected climatic covariates (*bio1, bio7* and *bio13*), Purcell's categories and the spatial effect were fitted. The model with the lowest WAIC value was the one including only the spatial effect, with a WAIC of 617.627 and a LCPO of 1.638 (Appendix A, Table A.1).

The posterior distribution of the spatial effect was in the linear scale, meaning that positive values imply a higher probability of the presence of *X. fastidiosa* and negative values were associated with a lower probability, and it was related to the data according to the neighborhood structure defined. The posterior mean of the spatial effect took positive values in areas where there was a high proportion of positives, while negative values were concentrated where *X. fastidiosa* was not detected (Figure 3.2a). The standard deviation of the posterior distribution of the spatial effect varied between 0.32 and 2.44, with lower uncertainty in areas where the spatial effect was greater, and the highest values in the isolated cells (Figure 3.2b). The posterior mean of the mixing parameter ϕ was 0.931, indicating that most of the variability was explained by the spatial effect (Table 3.1).

TABLE 3.1: Mean, standard deviation (sd), quantiles (Q) and mode for the parameters and hyperparameters (τ, ϕ) of the best model for the distribution of *Xylella fastidiosa* in the demarcated area in Alicante, Spain.

*Parameters and hyperparameters	Mean	sd	$Q_{0.025}$	$Q_{0.5}$	$Q_{0.975}$	Mode
β_0	-3.524	0.173	-3.884	-3.516	-3.208	-3.500
τ	0.936	0.264	0.529	0.898	1.557	0.827
ϕ	0.931	0.063	0.763	0.949	0.996	0.989

*β_0 is the intercept, τ is the precision of the spatial effect and ϕ the mixing parameter.

The mean of the predictive posterior distribution of the response variable was expressed in terms of probability (0-1), in our case with values ranging from 0.01 to 0.36. The highest probability of the presence of *X. fastidiosa* was in grid cells where the posterior distribution of the spatial effect was also highest. In grid cells where the pathogen was not detected nor in their neighboring cells, the probability of presence was close to zero (Figure 3.2c). The standard deviation of the predictive posterior distribution increased with the probability of presence, with the highest value of 0.22 (Figure 3.2d).

PCA was performed with the original covariates, resulting in 96.4% explained variance with the first three PCs (Appendix A, Table A.2). The covariates mean temperature of coldest quarter (*bio11*) and precipitation

seasonality (*bio15*) had a strong influence on PC1, with positive coefficients greater than 0.97. PC2 was mainly defined by the precipitation in the wettest month (*bio13*), which had a coefficient of 0.94. In PC2, the covariates related with temperature had negative coefficients, whereas other covariates of precipitation were positive. In PC3, the covariate with the highest weight (0.97) was the mean diurnal range (*bio2*). Nevertheless, when plotting the PCs there was no clear coincidence with the distribution of *X. fastidiosa* (Appendix A, Figure A.3). All model combinations were fitted (n = 32) considering the three PCs, Purcell's categories and the spatial effect. Again, the model including only the spatial effect was selected based on WAIC and LCPO values, and it was also found to be the most parsimonious (Appendix A, Table A.1).

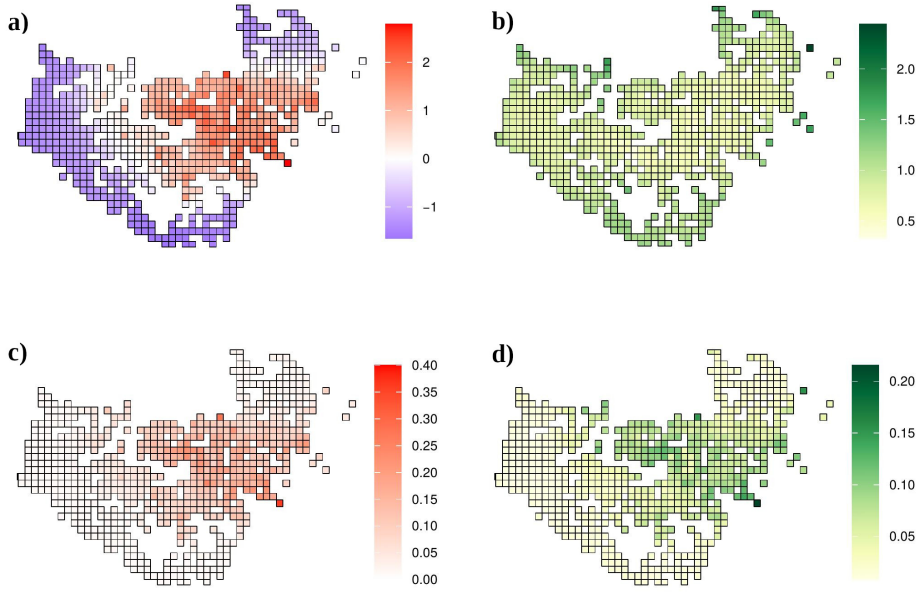

FIGURE 3.2: Model with the spatial effect. (**a**) Mean and (**b**) standard deviation of the posterior distribution of the spatial effect. (**c**) Mean and (**d**) standard deviation of the posterior predictive distribution of the probability of *Xylella fastidiosa* presence in the demarcated area in Alicante, Spain.

3.3.2 Province of Lecce

All the 20 climatic variables included in the analysis showed high linear correlation (Appendix A, Figure A.1b). Among them, annual mean temperature (*bio1*), mean diurnal range (*bio2*), mean temperature of the wettest quarter (*bio8*), mean temperature of the driest quarter (*bio9*), annual precipitation (*bio12*), and precipitation of the driest month (*bio14*) had $|\rho| < 0.7$. Most of the positives were observed in the southwestern area of Lecce, coinciding with low values of *bio2, bio9, bio12* and *bio14*, and high values of *bio1* and *bio8* (Appendix A, Figure A.4). Some positives were observed in the eastern part of the province, where *bio8* and *bio14* had the lowest values and *bio12* displayed the highest. Climatic variables had low variability, for instance, with *bio1* ranging from 16.07 ℃ to 17.23 ℃ and *bio12* from 506.6 mm to 679.7 mm. With regard to Purcell's categories based on minimum winter temperatures, only two (moderate and severe) were represented in the province of Lecce (Appendix A, Figure A.4g). Hence, this categorical variable was not further considered in the models.

All model combinations (n = 128) with the selected climatic covariates (*bio1, bio2, bio8, bio9, bio12* and *bio14*), and the spatial effect were fitted. The model selected was the one including the covariates *bio2* and *bio8* as well as the spatial effect. This model was chosen because it was the most parsimonious of those with the lowest values of WAIC and LCPO, meaning good model fit and predictive capacity with fewer covariates (Appendix A, Table A.3).

In the selected model, the posterior mean of the parameters of *bio2* and *bio8* was negative (Table 3.2). Regarding the spatial effect, positive values of the mean posterior distribution were associated with a higher probability of the presence of *X. fastidiosa*, while negative values were related to a lower probability. The mean of the posterior distribution of the spatial effect had higher values and lower variability in the areas where *X. fastidiosa* was first detected, on the southwestern coast (Figure 3.3). The posterior predictive distribution of the response variable showed that the model was strongly influenced by the spatial effect. Consequently, the probability of the presence of *X. fastidiosa* was much higher in the areas around positive findings, along with higher values of the spatial effect, but practically null in the areas

farther away from the positives (Figure 3.3). In addition, the standard deviation (uncertainty) of the predictive posterior distribution increased with the probability of presence.

TABLE 3.2: Mean, standard deviation (sd), quantiles (Q) and mode for the parameters and hyperparameters (r, $\sigma_{\boldsymbol{u}}$) of the best model for the distribution of *Xylella fastidiosa* in Lecce, Italy, based on mean diurnal range (*bio2*) and mean temperature of wettest quarter (*bio8*).

*Parameters and hyperparameters	Mean	sd	$Q_{0.025}$	$Q_{0.5}$	$Q_{0.975}$	Mode
β_0	17.465	11.509	-4.153	17.087	41.300	16.413
*bio*2	-1.149	0.744	-2.639	-1.144	0.313	-1.133
*bio*8	-1.216	0.706	-2.727	-1.177	0.064	-1.107
r	5609.527	1112.976	3858.475	5455.545	8202.005	5132.501
$\sigma_{\boldsymbol{u}}$	4.837	0.753	3.527	4.778	6.480	4.662

*β_0 is the intercept, *bio2* and *bio8* represent the parameters of the covariates mean diurnal range and mean temperature of wettest quarter, respectively. r and $\sigma_{\boldsymbol{u}}$ are the range and variance of the spatial effect, respectively.

The three PCs with climatic covariates explained an accumulated variance of 87.7% (Appendix A, Table A.2). PC1 was strongly influenced by two covariates: the mean diurnal range (*bio2*), with a positive coefficient of 0.96, and minimum temperature of the coldest month (*bio6*), with a negative coefficient of -0.96. The precipitation of the wettest month (*bio13*), the wettest quarter (*bio16*), and the coldest quarter (*bio19*) were the covariates that made a greater contribution in PC2, all of them with positive coefficients higher than 0.95. PC3 was mainly driven by the accumulated degree-days over 15 ℃ (*ADD*), with a coefficient of 0.97. The Ionian coast in Lecce, where most of the positives were located, showed the lowest values of PC1, although *X. fastidiosa* was also found in areas with high values for PC1. Similarly, the pathogen was detected in areas with both positive and negative values of PC2 and PC3 (Appendix A, Figure A.5).

All model combinations (n = 16) with the three PCs and the spatial effect were fitted. The best model based on the WAIC criteria was the one including the three PCs and the spatial effect. Considering that WAIC values were virtually equivalent, the model including PC2, PC3 and the

spatial effect was selected as the best one. Since the estimated parameters associated with PC2 and PC3 were negative in the model including the spatial effect (Table 3.3), higher values of these covariates were associated with a lower probability of the presence of *X. fastidiosa*. Nevertheless, the spatial component had the strongest effect on the model. Mean and standard deviation of the posterior distribution of the spatial effect were similar to those obtained with the model including climatic covariates, with higher values and lower variability in areas where the pathogen was detected (Figure 3.4). The predictive posterior distribution was also similar to that obtained with the model including climatic covariates. However, due to the effect of the PCs, in this case the highest probability of the presence of *X. fastidiosa* was concentrated on the Ionian coast, and was lower in other areas of the province. Moreover, the standard deviation (uncertainty) of the predictive posterior distribution was higher in the areas with a higher probability.

TABLE 3.3: Mean, standard deviation (sd), quantiles (Q) and mode for the parameters and hyperparameters (r, $\sigma_{\boldsymbol{u}}$) of the best model for the distribution of *Xylella fastidiosa* in Lecce, Italy, based on the second and third principal components (PC2 and PC3).

*Parameters and hyperparameters	Mean	sd	$Q_{0.025}$	$Q_{0.5}$	$Q_{0.975}$	Mode
β_0	-9.720	1.564	-13.316	-9.530	-7.203	-9.146
$PC2$	-1.139	0.543	-2.299	-1.110	-0.146	-1.056
$PC3$	-0.811	0.399	-1.628	-0.799	-0.056	-0.778
r	5609.527	1112.976	3858.475	5455.545	8202.005	5132.501
$\sigma_{\boldsymbol{u}}$	4.837	0.753	3.527	4.778	6.480	4.662

*β_0 is the intercept, *PC2* and *PC3* represent the parameters of the second and third principal components, respectively. r and $\sigma_{\boldsymbol{u}}$ are the range and variance of the spatial effect, respectively.

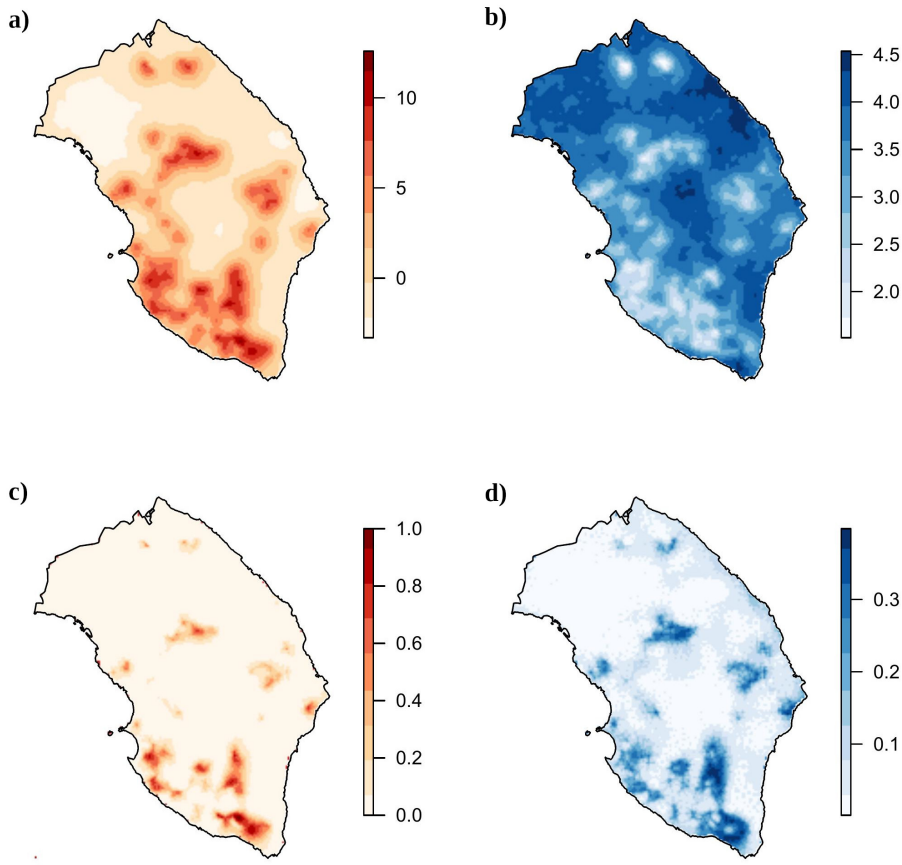

FIGURE 3.3: Model with the covariates mean diurnal range (*bio2*), temperature of the wettest quarter (*bio8*) and the spatial effect. (**a**) Mean and (**b**) standard deviation of the posterior distribution of the spatial effect, (**c**) mean, and (**d**) standard deviation of the posterior predictive distribution of the probability of *Xylella fastidiosa* presence in Lecce, Italy.

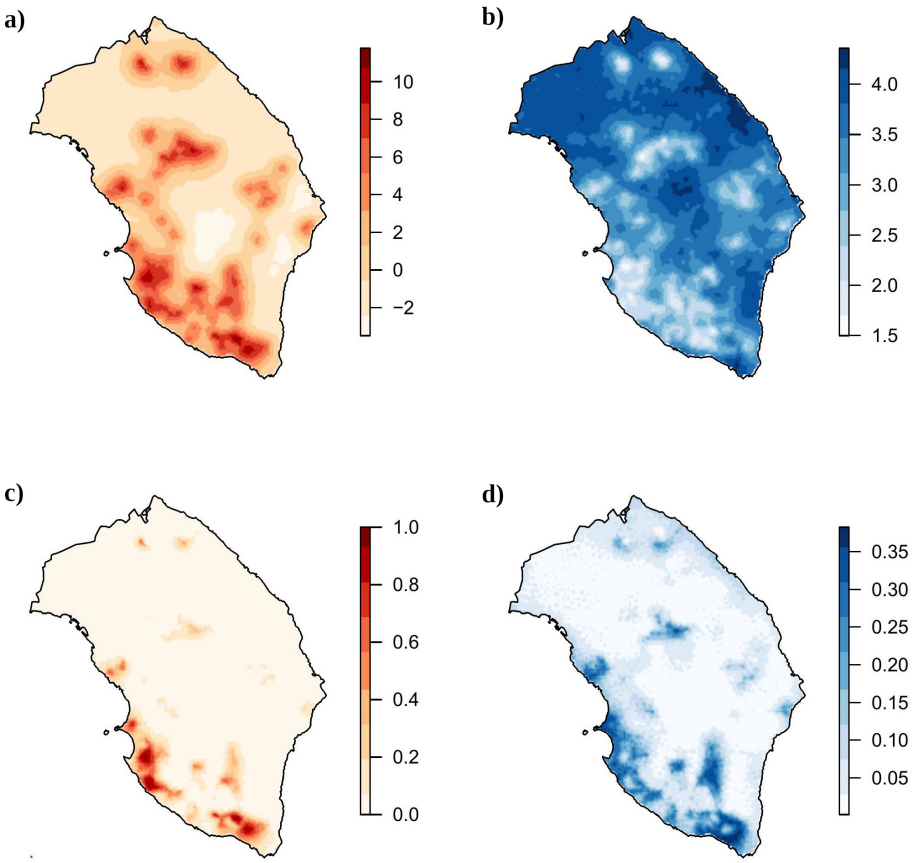

FIGURE 3.4: Model with principal components and spatial effect; (**a**) mean and (**b**) standard deviation of the posterior distribution of the spatial effect, (**c**) mean and (**d**) standard deviation of the posterior predictive distribution of the probability of *Xylella fastidiosa* presence in Lecce, Italy.

3.4 Discussion

The climatic covariates presented low variability in both regions, probably due to the limited extent of the study areas. Despite this, the four categories defined by Purcell based on the minimum winter temperature (Anas et al., 2008) were all represented in Alicante. This was noteworthy, since it

is not common to find all four categories in the same study area, and made it possible to infer whether the geographic distribution of the pathogen was somehow constrained by the low temperatures (Purcell, 1980; Lieth et al., 2011). Several studies have shown that successful *X. fastidiosa* infections (i.e., systemic host colonization) depend on certain factors like the temperature. For example, exposure of infected grapevines to low temperatures can effectively reduce or eliminate the pathogen, this phenomenon is known as "cold curing" (Purcell, 1980; Feil and Purcell, 2001). Actually, the climatic variables (temperature, rainfall) have a great influence on shaping the ecological conditions, which can (i) be more or less favorable for the insect vectors, in terms of population abundance, seasonal fluctuation, and attitude for dispersal; and (ii) influence the abiotic stresses and consequently the severity of the symptoms associated with *X. fastidiosa* infections.

Given the results obtained in Alicante, *X. fastidiosa* subsp. *multiplex* was detected in similar proportions in all Purcell's categories. This is in agreement with the known global distribution of this subspecies, which is present in warm climates, but also in areas characterized by cold winters such as Canada (Goodwin and Zhang, 1997). This is also in line with the results of Godefroid et al. (2019), suggesting that the subsp. *multiplex* may be more tolerant to cold temperatures. On the contrary, recent studies found a positive association with minimum winter temperatures in areas of France where *X. fastidiosa* subsp. *multiplex* is prevalent (Martinetti and Soubeyrand, 2019; Abboud et al., 2019). All this information together would suggested a greater environmental plasticity of subsp. *multiplex* compared to subsp. *fastidiosa*, and thus Purcell's categories for *fastidiosa* would not be applicable to *multiplex*. Nevertheless, disease severity was not recorded in the dataset from Alicante, but only the presence/absence of *X. fastidiosa*. Our empirical observations in Alicante indicated that no major differences in almond leaf scorch severity were observed among Purcell's four categories. In any case, formal quantification of disease severity in Alicante with a proper sample size and standard area diagrams would be needed to confirm this.

In Lecce, only two of Purcell's categories were represented (i.e., severe and moderate). Thus, no conclusions could be drawn in relation to the potential effect of low winter temperatures on the geographic distribution

of *X. fastidiosa* subsp. *pauca* in this region. This subspecies is prevalent in southern Italy, including the study area in Lecce (Saponari et al., 2018), and it is widespread in Central and South America (Nunney et al., 2014a; Haelterman et al., 2015; Coletta-Filho et al., 2016). It was also detected in Provence-Alpes-Côte d'Azur (France) (DG SANTE, 2020) and the Balearic Islands (Spain) (EFSA, 2019). The climates in all these regions are characterized by mild winters, suggesting that *X. fastidiosa* subsp. *pauca* has a lower tolerance to cold temperatures. Nevertheless, data from experiments under controlled conditions are needed to support this hypothesis.

Multicollinearity refers to the non-independence between the covariates, which can lead to inaccurate estimation of the parameters of the model and a bias in the statistical inference, thus inducing an incorrect identification of the relevant covariates in the model (Graham, 2003; Dormann et al., 2013). Climatic covariates are typically correlated, as was our case. In this study, two different methods were used to minimize this problem. The first one was based on the pairwise correlation of the climatic variables, selecting only those with $|\rho| < 0.7$. On the other hand, PCA was performed including all climatic variables. This is one of the most popular methods to reduce the number of covariates and avoid multicollinearity (Jolliffe, 2002; Dormann et al., 2013). Unlike the previous method based on pairwise correlations, PCA makes it possible to retain all the information provided by the set of variables through their linear combinations in PCs. Nevertheless, PCA has the disadvantage that the interpretation of the PCs linking the response variable and the covariates is not always straightforward. The zone near the coast in the study area in Alicante was characterized by high values of PC1 and low values of PC2 and PC3, meaning high values of the mean temperature of the coldest quarter and the precipitation seasonality, but low precipitation levels in the wettest month and in the mean diurnal range. In contrast, the lowest values of PC1 and the highest values of PC2 were located in a small central zone at a higher altitude, were samples not available. On the other hand, in Lecce, the lowest values of PC1, meaning low mean diurnal range and high temperatures in the coldest month, were found on the Ionian coast, where most of the positive samples were concentrated.

In Alicante, the models showed that climate effects, included as climatic

covariates or PCs, were not relevant. Therefore, no straightforward relationships could be established between the distribution of *X. fastidiosa* and temperature or precipitation in the study area. As indicated above, this may be due to the limited extent of the study area and the resulting relatively low variability of climatic variables. Nevertheless, the greater environmental plasticity of *X. fastidiosa* subsp. *multiplex* might have also played a role (EFSA, 2019). The best model based on the WAIC, LCPO and parsimony criteria included only the spatial effect. This implies that the areas close to positive findings of *X. fastidiosa* are more likely to be infested than those farther away. Our results also indicate that climatic factors are not likely to prevent the colonization of neighboring areas by *X. fastidiosa*. Therefore, control measures based on the reduction of inoculum and vector populations should be enforced to limit further disease spread.

Nevertheless, this spatial effect depends on the predefined neighborhood structure. In the case of Alicante, grid cells with a distance $\leq$ 2.5 km between their centroids were considered neighbors. This distance was established as being the shortest at which all grid cells with a resolution of 1x1 km had at least one neighbor, considering that non-sampled grid cells were not considered in the analysis. This might represent a limitation of our spatial model, which could be improved by increasing the spatial resolution of the dataset and defining alternative neighborhood structures based on actual disease spread distances.

The ability of several insect species to transmit *X. fastidiosa*, like *H. vitripennis* in southern California (Almeida and Purcell, 2003) or *P. spumarius* in Italy (Saponari et al., 2014; Cornara et al., 2017), is well documented. Daugherty and Almeida (2009) studied the ecology of two vectors of PD, *H. vitripennis* and *Graphocephala atropunctata* (Signoret) (Hemiptera: Cicadellidae), which are prevalent in coastal areas of California. Vector abundance together with the duration of the acquisition and inoculation periods greatly influenced the transmission efficiency of *X. fastidiosa* by insect species. Nevertheless, from a spatial epidemiology perspective, studies providing data on the actual distances of vector dispersal and disease spread are needed, especially for those present in Europe for which quantitative information is rather uncertain.

The models for the study area in Lecce showed that some climatic covariates could be related to the distribution of *X. fastidiosa* subsp. *pauca*. In particular, the mean diurnal range (*bio2*) and the mean temperature of the wettest quarter (*bio8*) were negatively related to the presence of the pathogen. That is, areas with higher variation in daily temperature and higher temperatures during the wettest months would have a lower probability of *X. fastidiosa* presence. Likewise, effects were observed with PC2 and PC3 in the model with PCA. In PC2, climatic covariates related to precipitation in the wettest month and the coldest quarter (*bio13* and bio19) made the greatest contribution, while PC3 was mainly defined by the *ADD* over 15 ℃ from April to October. Both PC2 and PC3 had negative coefficients in the model, indicating a lower probability of presence in areas with more precipitation in winter and higher temperatures during the vegetative growth period of the olive tree.

From these results it can be speculated that wet winters and hot summers in Lecce would be detrimental to vector activity and/or bacterial multiplication in the host plants. Nevertheless, given the environmental homogeneity found in the study area, the results of the model in terms of climatic covariates were inconclusive. The survey strategy was not uniform across the study area and most of the positive findings were concentrated around the first location of *X. fastidiosa* near Gallipoli (Martelli et al., 2016). Consequently, the heterogeneous distribution of the samples could be confounded with the climatic covariates.

Previous studies estimated the potential spread of *X. fastidiosa* in Europe, in order to implement control strategies and assess potential impacts. These studies used different methodologies like network analysis (Strona et al., 2017, 2020), spatially explicit process-based models (White et al., 2017), compartmental Susceptible-Infected-Removed (SIR) Bayesian models (Soubeyrand et al., 2018), or a coupled reaction-diffusion-absorption model, considering the spread via insects and transportation of plants (Abboud et al., 2019). Nevertheless, uncertainties on the actual vector dispersal and disease spread distances in Europe as well as on the human-assisted dispersal component limit the predictive capacity of the models. In our case, *X. fastidiosa* datasets from Alicante and Lecce were analyzed with different methodologies because they were actually two different types of spatial data.

In Alicante, the georeferenced samples were presented on a discrete space (lattice data), so that the spacial dependence was incorporated through an ICAR structure, which is a particular case of GMRF. In this way, the Besag, York and Mollié model was implemented (Besag et al., 1991) using the reparameterisation proposed by Simpson et al. (2017), which includes a standardised spatial effect. In the case of Lecce, locations were considered in a continuous space (geostatistical data), so the stochastic partial differential equation (SPDE) was used (Lindgren et al., 2011), where the GF is represented through Matérn covariance as GMRF, in order to use the INLA methodology.

The effect of spatial relationships in SDMs cannot be ignored, as is clearly illustrated by our study, where the spatial effect explained virtually all the variability found in the distribution of *X. fastidiosa* in both regions. Therefore, in our particular case studies, models which did not consider spatial autocorrelation could result in erroneous relationships between some covariates and the presence/absence of the pathogen. The use of Bayesian hierarchical models allowed a straightforward incorporation of the spatial effect, which would indeed be challenging from a frequentist approach. This methodology allowed sources of variability and unobserved uncertainty to be incorporated in a convenient way. Furthermore, INLA has proven to be a computationally efficient methodology to implement complex Bayesian hierarchical models including spatial autocorrelation.

In both study areas, the spatial component had a strong effect in the models regardless of the climatic variables. This substantial contribution of the spatial effect in the models might indicate that the current extent of *X. fastidiosa* in the study regions had arisen from a single outbreak in each zone or several nearby outbreaks that coalesced. Nevertheless, actual disease spread rates based on time-series data would be needed to confirm this hypothesis. In the case of Lecce, data were indeed available from 2013 to 2018, but with different surveillance strategies in each campaign, to comply with the updates of Decision (EU) 2015/789. This temporal and spatial heterogeneity in surveillance constrained the information that can be derived from the dataset, and only data from the first campaign and related to one province met the requirements to be used in model fitting. From an epidemiological modeling perspective, a recommendation for risk managers

would be to perform additional surveillance programs, complementary to those established by Decision (EU) 2015/789, to gather more informative epidemiological data and draw sound conclusions on the spatio-temporal scale of disease spread. Finally, the spatial models developed here may assist risk managers in designing more efficient surveillance strategies, where inspection and sampling efforts would be adjusted considering the probability of *X. fastidiosa* presence.

The study presented in this Chapter has been published in the following paper:

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

Data and code are available at: `https://bitbucket.org/mcendoya/xylella_alicante_lecce`.

Chapter 4

Non-stationary approach with dispersal barriers

The updating of information on the distribution of *X. fastidiosa* in Alicante through official surveys enabled further research, which led to the establishment of new objectives. Thus, the incorporation of different structures that may prevent the pathogen dispersal and the analysis of their effect on the spatial distribution were considered. This chapter is devoted to present the non-stationary modeling approached for the inclusion of these structures, referred to as barriers.

4.1 Introduction

An often overlooked problem in the analysis of spatial data is that models usually assume stationarity (i.e., the spatial effect is invariant to the map translation) and isotropy (i.e., the spatial effect is invariant to the map rotation), that is, the autocorrelation between two locations only depends on the Euclidean distance. However, relying on these two assumptions can produce misleading results, with unrealistic associations and/or bias in the prediction of the species distribution, when elements such as barriers that are an obstacle to the movement of the species are present in the study

area. To address this issue, Bakka et al. (2019) introduced an approach that makes it possible to deal with non-stationary spatial processes where, as in our study, stationary also includes isotropy for convenience. This approach has been applied in marine species distribution studies, where the coastline was implemented as a physical barrier (Bakka et al., 2019; Martínez-Minaya et al., 2019).

Barriers are an intrinsic part of the principles of plant disease control, i.e., exclusion, eradication, protection and resistance (Maloy, 1993). Exclusion strategies aim to prevent the pathogens from entering new areas. Barriers in the form of prohibitions restricting the import of plants, interceptions through border inspections and subsequent elimination of the pathogen are enforced by legal provisions worldwide. In the case of *X. fastidiosa*, the Commission Implementing Regulation (EU) 2020/1201 establishes special requirements for the import of host plants from third countries into the EU. When exclusion fails, eradication is attempted by removing the infected plants to limit further spread of the disease. According to Commission Implementing Regulation (EU) 2020/1201, demarcated areas consisting of an infected (i.e., infested) zone surrounded by a buffer zone should be established for *X. fastidiosa*. Eradication measures should then be implemented to ensure the removal of the infected plants and control of vector populations. Special requirements are also set for the movement of specified plants from the demarcated area.

Protection from already established diseases can be accomplished with barriers such as screenhouses, plastic covers and distance from inoculum sources that prevent pathogens and vectors from contacting host plants. Windbreaks can also prevent the movement of pathogen propagules. In areas where *X. fastidiosa* is endemic, screen and planting barriers have been evaluated to reduce vector spread (Daugherty and Almeida, 2009; Blua et al., 2005). Resistant cultivars, either alone or in mixtures, limit the infection and multiplication of plant pathogens, acting as a barrier for the onset of disease epidemics. In this regard, recent advances have been made to obtain grapevine and olive cultivars that are resistant to *X. fastidiosa* (Krivanek et al., 2006; Giampetruzzi et al., 2016).

All these examples described above illustrate to what extent the presence of barriers and their resulting non-stationarity can shape the spatial dimension of plant disease epidemics. Nevertheless, apart from performing separate directional spatial autocorrelation analyses to study whether a process is isotropic (Madden et al., 2007), the issue of non-stationarity has been scarcely explored in the context of plant disease epidemiology.

Our study focuses on the demarcated area for *X. fastidiosa* in Alicante, Spain (Appendix B, Figure B.1). Despite its relatively small extension, the study area of Alicante presents a great orographic diversity, from the sea level to mountain ranges rising to an altitude of above 1,500 m. This particular geographic setting must be taken into account to model the occurrence of *X. fastidiosa*, since it can affect the presence of host plants and also affect the behavior of the vectors, thus violating the stationarity and isotropy assumptions. In addition to these geographic barriers, the control measures for *X. fastidiosa* established by the EU legislation are aimed at limiting the spread of the disease, which also represents a potential dispersal barrier to be considered.

The aim of this study is to describe how overlooking non-stationarity in spatial models in the presence of dispersal barriers would provide misleading results when predicting plant pathogen distributions. In particular, the occurrence of *X. fastidiosa* in Alicante was analyzed with four spatial modeling scenarios, three of them including hypothetical dispersal barriers. With more realistic predictions of pathogen distributions, more targeted interventions can be implemented to reduce further disease spread, minimize the associated impacts and optimize disease management.

4.2 Methods

4.2.1 Database

The georeferenced data from the official surveys carried out for *X. fastidiosa* in Alicante, Spain, in 2018 were provided by the plant health authority (*Sanitat Vegetal, Generalitat Valenciana*). This database contained the

plant species sampled, the result of the laboratory analysis being positive (i.e., presence) or negative (i.e., absence) for *X. fastidiosa* based on real-time PCR (EPPO, 2019b), as well as the UTM coordinates of the location where the sample was taken.

Samples were also collected from plant species that were not known to be natural hosts for the *X. fastidiosa* subsp. *multiplex* strains present in the study area, such as *Olea europaea*, of which 2,414 samples were collected during that period, all of them resulting negative for *X. fastidiosa*. In order to avoid biases in the estimation due to this large number of negative samples from non-host species, only the samples from plant species having at least one positive for *X. fastidiosa* were considered for further analysis. The plant species selected were: *Prunus dulcis*, *P. armeniaca*, *P. domestica*, *Calicotome spinosa*, *Rhamnus alaternus*, *Phagnalon saxatile*, *Helichrysum italicum*, *Polygala myrtifolia*, *Rosmarinus officinalis* and *Laurus nobilis*. The dataset consisted of a total of 4,205 samples, 1,151 were positive and 3,054 were negative for *X. fastidiosa*, distributed in the demarcated area of Alicante with an extension of approximately 1,346 km^2 (GVA, 2019) (Appendix B, Figure B.1).

4.2.2 Geostatistical model

The observations of our georeferenced database of *X. fastidiosa* in Alicante were considered as geostatistical data, since they were made at continuous locations within a defined spatial domain. One of the characteristics of geostatistical models is that the main objective is to enable prediction in unsampled points within the study region (Cressie, 1993). A point-referenced spatial hierarchical model (Diggle et al., 1998) was used to fit the geostatistical data, while inference and prediction were performed within the Bayesian paradigm. As posterior distributions of the parameters and hyperparameters, along with the posterior predictive distributions of the predicted values in unobserved locations, do not have analytical expressions, the integrated nested Laplace approximation (INLA) methodology (Rue et al., 2009) was used to numerically approximate them. The outcomes of the model were summarized by the mean of the posterior distribution of the probability of *X.*

fastidiosa presence and its standard deviation to illustrate the uncertainty around the prediction.

Defining a hierarchical Bayesian spatial model can be seen as a three-step process. Firstly, a probability distribution must be identified for the observations available at the spatial locations. In this case, it was assumed that y_i, the occurrence of *X. fastidiosa* at location i, follows a Bernoulli distribution (1 indicating presence and 0 absence), that is, $y_i \sim \text{Bernoulli}(\pi_i)$, where π_i represents the probability of presence at location i. In a second step, this probability of presence π_i is linked (usually via the logit link when the response is Bernoulli) to a linear predictor and a latent Gaussian random field, whose covariance matrix Σ depends on two hyperparameters: the variance σ_u^2 of the spatial effect and the spatial range r, i.e., the distance from which two observations can be considered independent with a spatial correlation close to 0.1 (Blangiardo and Cameletti, 2015). Finally, the third step consists in assigning the corresponding priors and hyperpriors of the parameters and hyperparameters of the model. Despite its wide acceptance, INLA cannot be directly applied when dealing with continuously indexed Gaussian fields (GF). The underlying reason is that the cost of factorizing dense covariance matrices can be computationally demanding. Lindgren et al. (2011) proposed an alternative approach by using an approximate stochastic weak solution to a Stochastic Partial Differential Equation (SPDE) as a Gaussian Markov random field (GMRF) approximation to a continuous GF with Matérn covariance structure. A GMRF is a discretely indexed GF characterized by a sparse precision matrix Q, the factorizing computational cost of which is of order $O(n^{3/2})$, a large computational improvement compared to the factorization of a dense covariance matrix (of order O^n) that would imply the GF. In the approach proposed by Lindgren et al. (2011), the finite element method provides a solution to the SPDE, through the construction of a *mesh* (Appendix B, Figure B.2a), which consists in the triangulation of the study area (Bakka et al., 2018).

Using this approximation, the spatial term is reparameterized as $\boldsymbol{u} \sim N(0, \boldsymbol{Q}^{-1}(\kappa, \tau))$, where the parameters κ and τ control the spatial range (r) and the variance (σ_u^2). Specifically, $r = \sqrt{8}/\kappa$ and $\sigma_u^2 = \frac{1}{4\pi\kappa^2\tau^2}$ (Lindgren et al., 2011). However, for a more intuitive interpretation, the spatial effect

was parameterized in terms of the marginal standard deviation and the spatial range (Krainski et al., 2019).

Therefore, the hierarchical Bayesian spatial model with the Krainski et al. (2019) reparameterization can be expressed as:

$$\begin{aligned} y_i &\sim \text{Bernoulli}(\pi_i),\ i = 1, ..., n, \\ \text{logit}(\pi_i) &= \beta_0 + u_i, \\ \text{P}(\beta_0) &\propto 1, \\ \boldsymbol{u} &\sim N(0, \boldsymbol{Q}^{-1}(r, \sigma_u)), \\ r &\sim \text{PC-prior}(\mu_r, 0.5), \\ \sigma_u &\sim \text{PC-prior}(10, 0.01), \end{aligned} \tag{4.1}$$

where π_i is the probability of the presence of *X. fastidiosa* at location i, β_0 is the intercept, and $\boldsymbol{u}$ is the spatial effect. As can be observed, the linear predictor was reduced just to the intercept, the underlying reason being that previous works had indicated a dominating effect of the spatial component compared to available covariates in the demarcated area (Cendoya et al., 2020). This model already includes the scarce prior knowledge about parameters, expressed via a non-informative improper prior for the intercept, and about the hyperparameters. In this latter case, following Fuglstad et al. (2019), Penalized Complexity priors (PC-priors) were used to express vague prior knowledge about them. In particular, a PC-prior for the spatial range was defined as $\text{P}(r < \mu_r) = 0.5$, where μ_r was chosen as 50% of the diameter of the study region, while a PC-prior $\text{P}(\sigma_u > 10) = 0.01$ was defined for the standard deviation of the spatial effect.

4.2.3 Non-stationarity

The model introduced in the previous subsection assumes stationarity and isotropy. In order to deal with non-stationarity (i.e., non-stationary and anisotropic spatial processes), the approach presented by Bakka et al. (2019) was used. As happens in stationary models, estimating and predicting in non-stationary models can be rather complicated. In their proposal, Bakka et al. (2019) approximated them also by means of the SPDE approach using

the finite element method. However, in this case a system of two SPDEs is presented, one for the barrier area and the other for the remaining area, which we have also denominated normal area, adapting their terminology.

In particular, a non-stationary spatial effect $u(s)$ is the solution to the following system of stochastic differential equations:

$$\begin{aligned} u(\boldsymbol{s}) - \nabla \cdot \frac{r^2}{8} \nabla u(\boldsymbol{s}) &= r\sqrt{\frac{\pi}{2}}\sigma_u \mathcal{W}(\boldsymbol{s}), \text{ for } \boldsymbol{s} \in \Omega_n, \\ u(\boldsymbol{s}) - \nabla \cdot \frac{r_b^2}{8} \nabla u(\boldsymbol{s}) &= r_b\sqrt{\frac{\pi}{2}}\sigma_u \mathcal{W}(\boldsymbol{s}), \text{ for } \boldsymbol{s} \in \Omega_b, \end{aligned} \tag{4.2}$$

where $u(\boldsymbol{s})$ is the spatial effect, Ω_n is the normal area and Ω_b is the barrier area. r and r_b are the spatial ranges for the normal and barrier areas, respectively. σ_u is the marginal standard deviation, $\nabla = \left(\frac{\partial}{\partial_x}, \frac{\partial}{\partial_y}\right)$ and $\mathcal{W}(\boldsymbol{s})$ denotes white noise. Note that in the barrier area the correlation is eliminated by introducing a different Matérn field, with the same standard deviation, but with a spatial range close to zero.

4.2.4 Models

To analyze the effect of including barriers on the occurrence of *X. fastidiosa* in the study area, the following models were performed and compared:

i) **Stationary model**. Model in which both stationarity and isotropy are assumed, without any barrier. This model represents a scenario without any disease control interventions or geographical features potentially affecting the spread of the pathogen (Figure 4.1a).

ii) **Mountain barrier model**. Non-stationary model with barriers defined by the areas over 1,065 m. In absence of data on the maximum altitude where insect vectors can be found in the study area, this value was set based on the maximum altitude where a sample positive for *X. fastidiosa* was found in the study area. This model represents a scenario without any disease control interventions but with geographical features impeding the spread of the pathogen (Figure 4.1b).

iii) **Continuous barrier model**. Non-stationary model with a continuous barrier surrounding the infested area. This barrier consisted of a perimeter band 1,000 m wide, 500 m away from the outermost samples that were positive for *X. fastidiosa*. The width of the barrier was fixed to be lower than the spatial range (r) estimated for the stationary model. This model represents an hypothetical cordon sanitaire where all host plants were removed and measures implemented to completely impede the spread of *X. fastidiosa*. For consistency, the perimeter band was also implemented along the coastline (Figure 4.1c).

iv) **Discontinuous barrier model**. The same non-stationary model described above but with a discontinuous barrier surrounding the infested area. In this case, breaks of different sizes (1,000-3,200 m) were made in the perimeter band, facing sampled and non-sampled areas outside the barrier. This model represents a cordon sanitaire where all host plants were removed, but measures to impede the spread of *X. fastidiosa* have been implemented only in some parts (Figure 4.1d).

In the non-stationary models (ii, iii and iv), following equation (4.2), Ω_b represented the area occupied by the barriers, i.e., the area above 1,065 m in the mountain barrier model and the area of the cordon sanitaire in continuous and discontinuous barrier models. Ω_n included the remaining area in each model.

All models were fitted using the INLA methodology with the `R-INLA` package (`http://www.r-inla.org`) for R software (R Core Team, 2022). For each model a *mesh* was built, specifying in each case the barrier areas (Appendix B, Figure B.2). In the three non-stationary models, observations in the barriers were eliminated (all of them negative samples), following the assumption that *X. fastidiosa* cannot be present in this specific area (Figure 4.1).

Model performance was assessed based on the differences between the stationary model and those with barriers, along with the differences between the discontinuous and continuous ones, were obtained by subtracting the means of their corresponding posterior predictive distributions.

To visualize the spatial correlation between two observations as a function of distance, the Matérn correlation function was calculated using the posterior mean of the spatial range obtained in each model, based on the definition for this hyperparameter $r = \sqrt{8}/\kappa$.

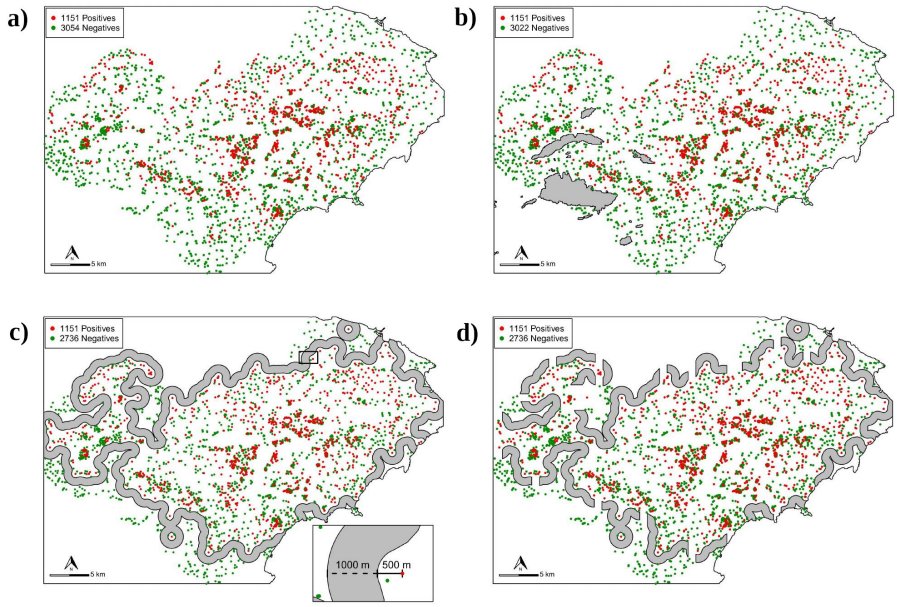

FIGURE 4.1: Positive (●) and negative (●) samples for *Xylella fastidiosa* and barriers incorporated in each model (shaded area) in the demarcated area in Alicante, Spain. (**a**) Stationary model, without barriers; (**b**) mountain barrier model; (**c**) continuous barrier model; and (**d**) discontinuous barrier model.

4.3 Results

In the stationary model, the posterior mean of the intercept was -1.68 in the linear predictor scale. Taking into account that when the spatial effect is zero, the mean posterior probability of the presence of *X. fastidiosa* is equivalent to the exponential transformation of the intercept, in this case, the probability of presence given only by the intercept was 0.19 (Figure

4.2a). The posterior mean of the spatial range was 4,030 m, with a 95% credible interval (CI) (2,907, 5,564) (Table 4.1). Therefore, we assume that two observations separated by more than this distance were not spatially correlated, that is, they are independent.

The posterior mean of the intercept in the mountain barrier, continuous barrier and discontinuous barrier models was -1.61, -1.79 and -1.57, respectively (Table 4.1). Therefore, in areas where there was no influence of the spatial effect, through the exponential transformation of these values, a probability of presence of the pathogen of 0.20 was obtained with the mountain barrier model, 0.17 with the continuous barrier model and 0.21 with the discontinuous barrier model (Figure 4.2).

TABLE 4.1: Mean and 95% credible interval (CI) for the parameter (β_0) and hyperparameters (r and σ_u) of the models for the distribution of *Xylella fastidiosa* in the demarcated area in Alicante, Spain.

Models	* Parameters and hyperparameters	Mean	95% CI
Stationary	β_0	-1.68	(-2.21, -1.23)
	r	4030.17	(2907.41, 5563.88)
	σ_u	1.52	(1.28, 1.80)
Mountain barrier	β_0	-1.61	(-2.09, -1.19)
	r	3860.88	(2918.61, 5212.18)
	σ_u	1.43	(1.20, 1.71)
Continuous barrier	β_0	-1.79	(-2.63, -1.15)
	r	6141.08	(4296.32, 9042.99)
	σ_u	1.50	(1.20, 1.88)
Discontinuous barrier	β_0	-1.57	(-2.23, -1.04)
	r	5298.90	(3813.16, 7557.78)
	σ_u	1.44	(1.17, 1.78)

*β_0 is the intercept, r is the spatial range and σ_u is the standard deviation of the spatial effect.

In the mountain barrier model, a posterior mean of the spatial range of 3,861 m was obtained with a 95% CI (2,919, 5,212). In the continuous barrier model the spatial range was greater than in the previous case and with more variability, obtaining a posterior mean of 6,141 m, with a 95% CI

(4,296, 9,043). The estimation of the discontinuous barrier model resulted in a posterior mean of the spatial range of 5,299 m with a 95% CI (3,813, 7,558) (Table 4.1).

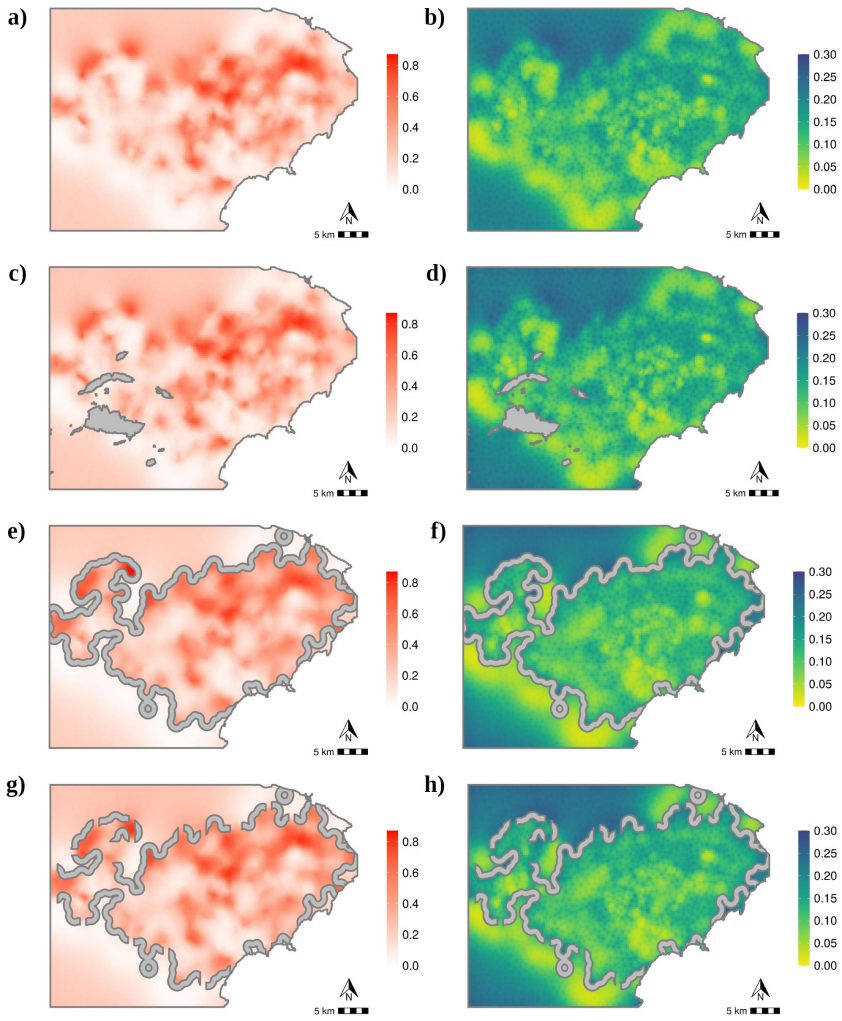

FIGURE 4.2: Mean (left) and standard deviation (right) of the posterior predictive distribution of the probability of *Xylella fastidiosa* presence in the demarcated area in Alicante, Spain. (**a**, **b**) Stationary model; (**c**, **d**) mountain barrier model; (**e**, **f**) continuous barrier model; and (**g**, **h**) discontinuous barrier model.

The Matérn correlation function was similar in the stationary and mountain barrier models, where the spatial correlation decreases quickly in the first 4,000 m. In the continuous barrier and discontinuous barrier models, the spatial correlation as a function of distance had a more gradual decrease due to the greater spatial range obtained in the estimation (Figure 4.3).

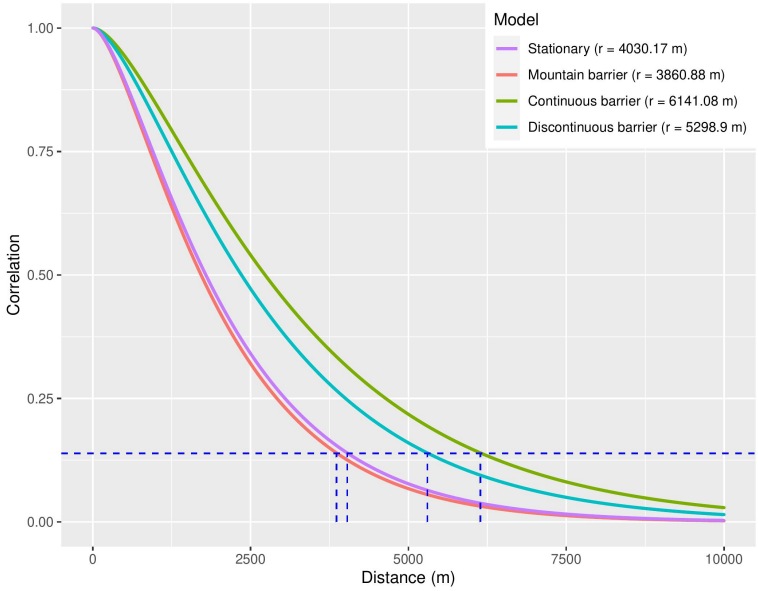

FIGURE 4.3: Representation of the Matérn correlation function for the posterior mean of the spatial range obtained in each model (vertical blue dashed lines), i.e., the distance at which the correlation is close to 0.1 (horizontal blue dashed line).

Given the model described in equation (4.1), the mean of the posterior predictive distribution, expressed in terms of probability, was defined by the intercept and the spatial effect. In general, in the four modeling scenarios the probability of the presence of *X. fastidiosa* was higher in the areas where the positive samples were concentrated, being close to zero in the areas where negative samples predominated. In the non-sampled areas at distances from the observations outside the spatial range, and thus without any influence of the spatial effect, the probability of the presence of *X. fastidiosa* only depended on the intercept. Regarding the standard deviation of the posterior predictive distribution, higher values were obtained in the

non-sampled areas, while the sampled areas where *X. fastidiosa* was not detected showed very low variability (Figure 4.2).

The range of values of the mean and standard deviation of the posterior predictive distribution was similar in all four models (Figure 4.2). However, the difference between the mean of the stationary model and the mountain barrier model was negative in the area around the barrier (Figure 4.4a). This implies that the probability of *X. fastidiosa* presence in those areas was higher in the mountain barrier model than in the stationary model.

In order to help in the interpretation of the comparison of the results of the stationary model, the continuous barrier model and the discontinuous barrier model, from now on we denominate the areas on both sides of the perimeter barrier built around the positives as external and internal zones. The maximum probability of the presence of *X. fastidiosa* was 0.46 in the area corresponding to the external zone in the stationary model (Figure 4.2a), while it was 0.29 and 0.36 in the continuous and discontinuous barrier models, respectively (Figure 4.2e, g).

The mean of the posterior predictive distribution of the probability of *X. fastidiosa* presence was higher with the stationary model than in the continuous barrier model, particularly in the northern area adjacent to the barrier. Actually, considering the difference between the mean of the posterior predictive distribution of the stationary model and the continuous barrier model, only positive values were obtained in the external area of the barrier (Figure 4.4b). This same behavior was also observed, but to a lesser extent, when the stationary and discontinuous barrier models were compared (Figure 4.4c). However, the difference between the discontinuous and continuous barrier models showed that in the areas where breaks were implemented, the probability of *X. fastidiosa* presence was similar or even increased, depending on the location. In particular, the probability of presence in the external area of the barrier increased through the breaks located in the north, while no differences were observed in those in the south-west (Figure 4.4d).

With respect to the mountain barrier model, the continuous and discontinuous barrier models showed a higher probability of *X. fastidiosa* presence in the areas adjacent to the barrier (Figure 4.2e, g). This increase in the

probability of the presence of the pathogen in the internal area adjacent to the perimeter barriers was also observed in the difference between the mean of the posterior predictive distribution of the stationary model and the continuous and discontinuous barrier models (Figure 4.4b, c).

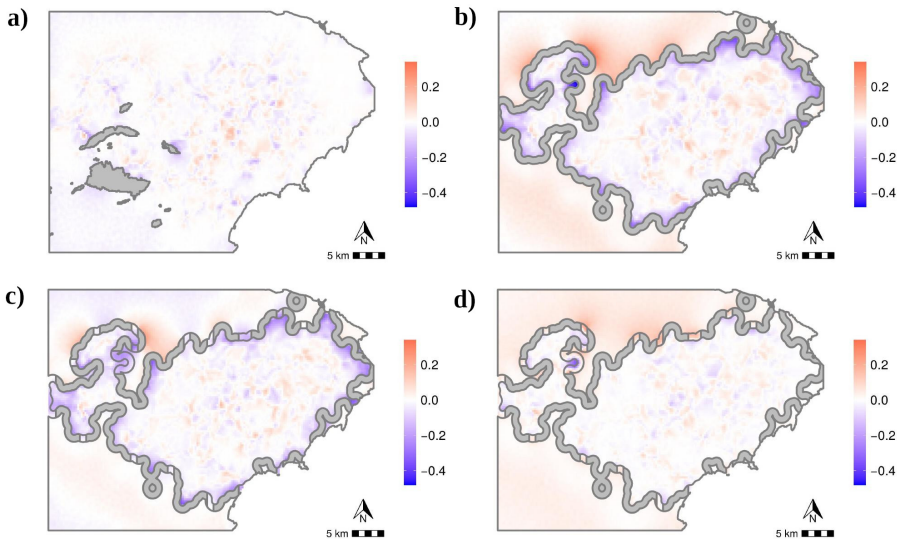

FIGURE 4.4: Differences in the mean of the posterior predictive distribution of the probability of *Xylella fastidiosa* presence in the demarcated area in Alicante, Spain. (**a**) Difference between stationary model and mountain barrier model; (**b**) difference between stationary model and continuous barrier model; (**c**) difference between stationary model and discontinuous barrier model; and (**d**) difference between discontinuous barrier model and continuous barrier model.

4.4 Discussion

The occurrence of *X. fastidiosa* in the demarcated area in Alicante was modeled using hierarchical Bayesian spatial models with the incorporation of barriers, following the methodology described by Bakka et al. (2019). Here, the main objective was to evaluate the influence of different types of barriers in the distribution of the pathogen and how ignoring non-stationarity would

result in misleading predictions. From the perspective of the SDMs, not considering the presence of elements in the landscape that prevent or hinder the spread of the organisms would be far from reality in most situations. Assuming stationarity and isotropy in this context would give inaccurate results (Bakka et al., 2019). Non-stationary models that incorporate barriers may also allow the effect of disease control interventions to be simulated.

In this case, climatic variables were not included in the models for the occurrence of *X. fastidiosa* in the demarcated area in Alicante. Previous works indicated that these variables were not relevant in this specific scenario, whereas a strong dominating effect of the spatial component was observed (Cendoya et al., 2020). Our analysis confirmed the strong spatial aggregation of *X. fastidiosa* in the demarcated area in Alicante, so the probability of *X. fastidiosa* presence was increased in the areas with higher prevalence of the pathogen compared to those where it was not detected (Figure 4.2). These results are in line with other studies highlighting the importance of incorporating the spatial structure in SDMs for plant pathogens (Meentemeyer et al., 2008).

In contrast to previous studies on marine species (Bakka et al., 2019; Martínez-Minaya et al., 2019), in the case of *X. fastidiosa* the overall values of the posterior predictive distribution of the stationary model were relatively similar to those obtained with the models that incorporated barriers (Figure 4.2). On the one hand, this was a somewhat unexpected result, considering that the barriers were assumed to be completely impervious to the spread of the pathogen. On the other hand, the results obtained here somehow illustrate the actual difficulties involved in effectively containing the spread of the pathogen by implementing dispersal barriers (Kottelenberg et al., 2021).

Nevertheless, relevant differences in the posterior predictive distribution of the probability of *X. fastidiosa* presence resulting from the incorporation of the barriers in the models can be appreciated in finer spatial detail. When the area above an altitude of 1,065 m was considered as a barrier for the spread of *X. fastidiosa*, the main difference with respect to the stationary model was found in the zone adjacent to the barrier. In this area, the probability of the presence of *X. fastidiosa* was higher in the mountain

barrier model than in the stationary model due to the smoothing effect that occurred when mountains were not considered as barriers (Figure 4.4a).

The dimensions and characteristics of the cordon sanitaire, i.e., continuous or discontinuous perimeter barriers, were based on the spatial range of approximately 4 km obtained in the stationary model (Table 4.1). To observe differences when incorporating the perimeter barrier, it should be situated less than 4 km away from the positive samples. Due to the assumed impermeability, the width of the perimeter barriers had no influence on the probability of *X. fastidiosa* presence in the area outside the barrier. This implies that the width of the hypothetical barriers used in our study cannot be interpreted in terms of the extent of the area subjected to disease control measures, such as the removal of infected host plants and vector control, as established by the Commission Implementing Regulation (EU) 2020/1201.

In the continuous barrier model, the probability of the presence of *X. fastidiosa* in the area adjacent to the outer border of the barrier was only determined by the negative samples and the intercept (Figure 4.2e), resulting in a lower probability of the presence of the pathogen compared to the stationary model (Figure 4.4b). Differences between the discontinuous and continuous barrier models showed that breaks in the perimeter barrier in areas with low sampling intensity, due to the greater uncertainty, resulted in a higher probability of *X. fastidiosa* presence (Appendix B, Figure B.3b). The increase in the probability of the presence of the pathogen through the breaks in the barrier was even greater than the difference with the stationary model (Figure 4.4c). However, no major influence of the cordon sanitaire was observed in areas with a high sampling intensity adjacent to the outer border of the barrier. In those areas, the breaks in the barrier did not increase the probability of *X. fastidiosa* presence (Appendix B, Figure B.3c).

These results may assist plant health authorities in prioritizing the areas for the implementation of surveillance and disease control barriers. The highest priority would therefore be given to non-sampled areas close to high occurrence locations, where the implementation of a barrier would lower the probability of the presence of *X. fastidiosa*. Areas where the surveys concluded that the pathogen is absent (i.e., below the design prevalence)

would be, therefore, of lower priority for the implementation of surveillance and disease control barriers, as the breaks would not increase the probability of presence of *X. fastidiosa*. These results are in line with current approaches aiming for a more targeted and risk-based management of emerging plant pathogens (Parnell et al., 2014; Hyatt-Twynam et al., 2017).

In the context of our study, the spatial aggregation obtained with the models resulted from the concurrent means of spread of *X. fastidiosa* acting during the whole time span of the epidemic. For the demarcated area in Alicante, Cornara et al. (2019) indicated that *X. fastidiosa* was detected in *P. spumarius* and *N. campestris*, with a prevalence of 27% and 1.2% of the individuals tested for the bacterium, respectively. However, the references quoted in this review do not report data on the prevalence of *X. fastidiosa* in vector populations in this region. Official samplings conducted from 2017 to 2019 by the plant health authority in the demarcated area resulted in prevalences of *X. fastidiosa* of 0.67% for *N. campestris* (n = 2,995) and 7.19% for *P. spumarius* (n = 3,157) (GVA, 2020). Similar values have been reported in the Balearic Islands, with 1.12% for *N. campestris* (n = 797) and 8.25% for *P. spumarius* (n = 5,806) (MAPA, 2021a). However, the prevalence values in Alicante are much lower than those described for *P. spumarius* in Corsica (>40%) and Apulia (>50%) (Cruaud et al., 2018; Cornara et al., 2017; Saponari et al., 2014). Although from our results we cannot conclude on the relevance of the different means of disease spread, these data suggest that vectors might not be playing a dominant role in the demarcated area in Alicante. Furthermore, it should be considered that the probability of infection of a plant by vectors depends not only on the prevalence, but also on the abundance of infectious vectors, their acquisition rate, transmission efficiency, the time period of the inoculation process and the infectivity of the vectors (Purcell, 1981). For instance, EFSA (2019) used expert knowledge elicitation (EKE) to estimate a median acquisition rate of 12.08% and a transmission efficiency of 13.58% for spittlebug vectors in olives.

The dispersal capacity of *X. fastidiosa* vectors in Europe is rather uncertain, and no studies are available for the particular epidemiological setting in Alicante. According to a Mass-Mark-Recapture assay by Lago et al. (2021) conducted in Madrid, Spain, individuals of *N. campestris* were found

at a distance of more than 2,000 m from the release point, with a relatively similar number of catches at 123 and 281 m. Studies conducted with *P. spumarius* in Apulia and Piedmont, Italy, resulted in a median dispersal from the release point of 26 m day^{-1} in an olive grove and 35 m day^{-1} in a meadow. It was estimated that 50% of the *P. spumarius* population in olives in Apulia remained within 200 m and 98% within 400 m for 2 months, with a dispersal limited to some hundreds of meters throughout the whole year (Bodino et al., 2021). EFSA (2019) conducted EKEs on the uncertainty distribution of the vector local spread and the mean distance of disease spread. The 5th, 50th and 95th percentiles of the uncertainty distribution for the vector local spread were 0.148 km, 0.767 km and 2.204 km, respectively. Percentiles for the mean distance of disease spread were 1.10 km, 5.18 km and 12.35 km, this median value being included in the 95% CI of the posterior distribution of the spatial range of our stationary model (Table 4.1). This upper bound corresponds to the estimated rate of movement of the *X. fastidiosa* front in Apulia (Kottelenberg et al., 2021). Nevertheless, these EKEs were conducted under specific assumptions and their extrapolation to the scenario in Alicante is not straightforward. Among other assumptions, values were elicited for olive orchards with herbaceous cover, without the influence of competing hosts or extreme winds on vector behavior. The movement of propagating plant material was not taken into account either. Moreover, the time component was not considered in our study.

In fact, plant propagating material is considered the main pathway for the entry of *X. fastidiosa* into new regions EFSA (2019). After the introduction of the pathogen with imported infected plant material, further spread in the area can also be driven by the movement of propagating plant material. Studies reconstructing the progression of almond leaf scorch disease in Majorca indicated that *X. fastidiosa* was introduced into this island with almond buds or stems from California, and then spread through the archipelago by grafting (Moralejo et al., 2020). Grafting experiments performed in this study resulted in a transmission of about 15% with almond buds, but other studies reported values up to 60% and 80% with almond buds and stems, respectively (Mircetich et al., 1976). In the case of Alicante, genetic studies indicated that *X. fastidiosa* might also have been introduced from California (Landa et al., 2020). In the demarcated area in Alicante,

almond groves were typically established with rootstock seeds that were later grafted on site with buds or stems of the scion (Cambra and Cambra, 1991). These grafting materials were generally obtained from almond trees in the area or from outside when a new cultivar was first introduced. In fact, previous studies suggested that the current extent of the pathogen had arisen from a single introduction (Cendoya et al., 2020; Landa et al., 2020). Nevertheless, with the information available, it is not possible to accurately trace back the movement of propagating plant material in the area and thus determine its actual role in the spread of *X. fastidiosa*. Therefore, the spatial dependence illustrated by our models should be interpreted considering any potential means of spread, including propagating plant material and insect vectors.

The spatial ranges obtained with the models varied from approximately 4 to 6 km (Table 4.1), but to relate this parameter to the actual epidemiological setting in the demarcated area in Alicante, only those from the stationary and mountain barrier models should be considered. The continuous and discontinuous barrier models incorporated hypothetical disease control interventions in the form of barriers, which are not present in the study area as such. Furthermore, imposing a cordon sanitaire implied a strong spatial aggregation in the area surrounded by this perimeter barrier, resulting in a greater spatial range compared to the other models studied. The models assuming no control interventions presented similar spatial dependence for the occurrence of *X. fastidiosa*. The posterior mean of the spatial range in the stationary model was 4,030 m with a 95% CI (2,907, 5,564), whereas for the mountain barrier it was 3,861 m with a 95% CI (2,919, 5,212) (Table 4.1). Interpreting these values in terms of spread rates is, however, difficult as the contribution of the different means of pathogen spread cannot be disentangled. Moreover, with the information available, it is not possible to determine when the pathogen was first introduced in the area and so the temporal component is missing. Studies combining dendrochronology and phylogenetic analysis indicated that the introduction of *X. fastidiosa* in Majorca occurred around 1993 (Moralejo et al., 2020). Epidemiological models dated the introduction of the pathogen in Corsica to around 2001 when hidden infection reservoirs are not considered, and around 1985 when these non-observable hosts are included in the models (Soubeyrand et al.,

2018). The rate of movement of the invasion front of *X. fastidiosa* in Apulia indicated that the disease spread started in approximately 2008 (Kottelenberg et al., 2021). Based on the low genetic diversity and the absence of recombinant events (Landa et al., 2020), it can be speculated that *X. fastidiosa* was introduced in the demarcated area in Alicante not earlier than in Majorca or Corsica.

Although spread rates cannot be inferred from our analysis, the spatial component of the models provides useful information for the management of *X. fastidiosa* in the study area. In the Matérn correlation function of the stationary and mountain barrier models, distances up to 1,792 and 1,717 m, respectively, accounted for 50% of the spatial correlation, and was less than 5% for distances longer than 5,698 and 5,459 m, respectively (Figure 4.3). Regardless of the date of introduction and the weight of the different means of spread of the pathogen in the demarcated area, the mean value of the spatial range for the stationary model indicates that host plants that were closer than 4,030 m to an infected plant would be at risk of giving being for *X. fastidiosa*. Therefore, these distances should be observed to define the buffer zone where the surveillance activities will be conducted around the infested area. Originally, the Commission Implementing Decision (EU) 2015/789 established that the buffer zone surrounding the infested zone should have a width of at least 10 km. The minimum width of the buffer zone was later reduced to 5 km by the Commission Implementing Decision (EU) 2017/2352 and currently to 2.5 km by the Commission Implementing Regulation (EU) 2020/1201. Based on our models, these minimum buffer zone widths do not cover the entire area at risk for *X. fastidiosa* occurrence in the demarcated area in Alicante. Consequently, in 2019 the plant health authority implemented an additional band of 10 km surrounding the demarcated area, where official surveillance activities are also being conducted (GVA, 2020).

It should be noted that the methodological improvement considering the non-stationarity of the spatial process did not increase the computational cost or the difficulty of its implementation (Bakka et al., 2019; Martínez-Minaya et al., 2019). To our knowledge, this study is the first to apply non-stationary models with barriers in the context of plant health. However, imposing the condition that barriers are completely impermeable implies

that the pathogen cannot be present or cross this area, which is a very strong assumption rarely met in practice. In the specific case of *X. fastidiosa*, these barriers represent areas without host plants and in which it is not possible for infected vectors or propagating plant material to pass through. Our discontinuous barrier model partially relaxed this assumption, allowing the pathogen to spread in some areas but still assuming that parts of the cordon sanitaire were completely impervious, which is seldom the case for *X. fastidiosa* and plant pathogens in general. Building on the present work, new modeling methods need to be developed to accommodate the incorporation of barriers with different levels of permeability, and thus more realistic plant health scenarios may be considered.

The study presented in this Chapter has been published in the following paper:

- **Cendoya, M., Hubel, A., Conesa, D., and Vicent, A. (2022). Modeling the spatial distribution of *Xylella fastidiosa*: A nonstationary approach with dispersal barriers. *Phytopathology*, 112(5):1036-1045.**

Data and code are available at: `https://doi.org/10.5281/zenodo.4656029`

Chapter 5

An individual-based spatial epidemiological model for the spread of plant diseases

In the previous chapters we have focused on the spatial distribution of the pathogen based on data from official surveys conducted so far. With a more updated view based on information provided by other research works, in this chapter we address the spatio-temporal perspective by means of a spatially explicit individual-based model for the disease spread under different settings with the aim to trace the potential progression of the outbreak from the origin to the current situation.

5.1 Introduction

The development of new statistical methods and computational advances have made disease modeling a widely used tool to assess risks and predict the spread of diseases, which is useful in the design of effective strategies for their control (Keeling and Rohani, 2008; Chen et al., 2014). Mathematical models have been extensively used to study the epidemiology of diseases in

humans and animals (Keeling and Rohani, 2008), and in plants (Gilligan, 2002; Madden et al., 2007). Compartmental epidemiological models assume that the population can be divided into different compartments according to the state of the disease. The simplest model is composed of three compartments dividing the population into three states, i.e., susceptible, infected and recovered individuals. Compartmental models are often built around differential equations, which describe the behavior of the disease at the population level (Kermack and McKendrick, 1927). More complex models have been developed including different compartments, sub-classes of each compartment, and even stochastic approximations with discrete time-steps considering spatial heterogeneity in different ways, as reviewed by Kleczkowski et al. (2019). Parameter estimation in these models is usually performed, in the simplest way, using classical techniques. One of these techniques, known as least squares, consists of obtaining those parameters that minimize the distance between the model and the data (Matis and Hartley, 1971; Chowell et al., 2009; Capaldi et al., 2012). However, there are currently other methodologies that are becoming increasingly important. On the one hand, a Bayesian framework through Markov chain Monte Carlo (MCMC) methods (Gibson, 1997; Gibson et al., 2004). On the other hand, metaheuristic algorithms such as the Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) (Akman and Schaefer, 2015; Akman et al., 2018).

A drawback of compartmental models based on differential equations is that it is not possible to identify the individuals fluctuating between the compartments. The rate of transmission is then proportional to the number of susceptible individuals, the proportion of infectious individuals and the rate of contacts among individuals (Kermack and McKendrick, 1927; Kleczkowski et al., 2019). An alternative approach consists in using individual-based models, which make it possible to monitor the state of each individual and the interactions among them. Nevertheless, the number of iterations increases substantially with large population sizes and long time horizons, leading to high computational costs that can limit the implementation of these models (Keeling and Rohani, 2008).

Modeling spread requires a good knowledge of the spatiotemporal behavior of the disease. When including spatial structures in the models, it

is usually assumed that interactions among individuals, and thus potential disease transmission, decrease with distance. However, long-distance spread due to human interventions or other factors cannot be ignored, as it plays an important role in shaping disease progress. It is also important to consider the spatial heterogeneity of populations, since host aggregation and distance from the source of transmission can strongly influence model outcomes (Keeling and Rohani, 2008).

The spatial component of plant disease spread has been approached from different methods, such as partial differential equations, distance class methods, spatial autocorrelation, or metapopulations (Madden et al., 2007; Meentemeyer et al., 2011). The scale and the type of spatial relationship in the models is often determined by the availability of data and the computational cost required for their analysis. Using a fine spatial scale can have a high computational cost, but spatial effects cannot be detected when aggregating the data at a larger scale. For instance, with lattice (aggregated) data modeling, the neighborhood structure is an approximation of the real dynamics, thus increasing the speed of the simulations (Keeling and Rohani, 2008). However, working at a finer scale, i.e., the individual level of a population, with individuals represented as discrete points in time and space, allows the spread of the disease to be described more accurately.

The spatiotemporal dynamics of an individual-based model can be described through the probability of a susceptible individual becoming infected. This probability is quantified as a function of the surrounding infectious population and the effect of distance between individuals on disease transmission, which is usually incorporated by different forms of kernels (e.g., Keeling et al. 2001; Deardon et al. 2010; Meentemeyer et al. 2011; Hyatt-Twynam et al. 2017). Here, we integrated spatial dependence in the model by means of a correlation function. Although several functions have been proposed to handle the spatial effect, the Matérn correlation function (Matérn, 1986) is widely used in several fields due to its flexibility, since it encompasses several functions depending on the value of its smoothing parameter (Stein, 1999; Guttorp and Gneiting, 2006). Despite its advantages, to date, the Matérn correlation function has not been integrated into spatial individual-based models for plant disease spread.

The main objective of this study was to develop a spread model for plant diseases, i) based on individuals, thus identifying their state at any given time, ii) considering spatial dependence, and iii) that was computationally efficient. For this purpose, from the available data, an optimal value for the disease transmission rate was obtained using the classical structure of a compartmental model, to which the spatial dependence was also included through the Matérn correlation function. Then, a spatially explicit algorithm for the disease spread at individial level was designed.

Almond leaf scorch disease (ALSD), caused by *X. fastidiosa*, was used as a case study. In particular, disease spread in almond trees in an area of Alicante was simulated using this individual-based model. The effects of the parameters on disease spread were evaluated, as well as the different types and locations of initial disease introduction. A user-friendly tool to visualize the results was also developed. Spatial individual-level prediction of disease spread provides a better understanding of the epidemiology, thus allowing the application of more targeted control measures and the optimization of resources.

5.2 Epidemiological model

This section describes the proposed framework for modeling plant disease spread (i.e., spatiotemporal progression), based on the disease status of each individual and the influence of their spatial dependence. The proposed methodology is structured into two distinct phases. Firstly, we employ a compartmental model, considering all individuals at each moment in time. This model is governed by a parameter-dependent system of differential equations, determining the transitions of individuals among various states, which also includes spatial dependence. Should real data be available, the parameters involved in both phases can be estimated using either classical deterministic techniques or stochastic methods, thereby acknowledging and accounting for the inherent uncertainty. Since this model only provides us with information at the population level, this concept is extended to a model based on individuals. We conclude the section by presenting an algorithm to simulate using our proposed model.

5.2.1 Compartmental model

Based on the structure of the compartmental model, individuals are classified according to their disease status as susceptible, i.e., non-infected individuals, or infected. Depending on the disease and the organism affected, the number of compartments and the transition between them can be highly variable. Models with more compartments may be closer to reality but also imply greater complexity in data fitting, require greater availability of information, and are more case-specific. Here, following a basic model with the usual states for most plant diseases, three possible disease states were defined for individuals: susceptible (S), those that are susceptible but are not infected; asymptomatic infected (I_a), those infected and infectious but without visible symptoms; and symptomatic infected (I_s), those infected and infectious with visible symptoms. Following the scheme in Figure 5.1, infection of S individuals comes from both I_a and I_s. The disease transmission rate of infectious individuals, i.e., I_a and I_s, is defined by β. Since asymptomatic individuals have a lower pathogen concentration (EFSA, 2019), the transmission rate of I_a is reduced by the parameter λ. After an asymptomatic period, the transition from I_a to I_s, i.e., the time elapsed for the symptoms expression, is given by σ. The transmission rate and population size are assumed constant over time.

However, the spatial dependence of individuals and their spatial distribution can influence disease spread. This can be taken into account by incorporating any possible spatial dependency between susceptible and infectious individuals, such as a correlation function implying a distance-dependent decrease in disease transmission. In particular, we used the Matérn correlation function (Matérn, 1986), a very flexible correlation family that generalizes many of the correlation functions widely used in spatial statistics. For two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, separated by a Euclidean distance $d_{ij} > 0$, the Matérn correlation is:

$$C(d_{ij}) = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa d_{ij})^{\nu} K_{\nu}(\kappa d_{ij}), \tag{5.1}$$

where K_{ν} is the Bessel modified function of the second kind and order ν, which is the smoothness parameter of the function; and κ is a scale

parameter defined by the range parameter r, considered as the distance at which two locations can be considered spatially uncorrelated. The Matérn class comprises different special cases with only two parameters, providing a wide range of forms of correlation functions, including the exponential correlation function when $\nu = 1/2$ (Handcock and Wallis, 1994; Guttorp and Gneiting, 2006). The literature contains different parameterizations of the Matérn correlation function which define the scale parameter in a variety of ways (see Handcock and Wallis 1994; Stein 1999; Diggle et al. 2003). We used the one described by Lindgren et al. (2011), which facilitates the parameters interpretation, where from the empirically derived definition $r = \sqrt{8\nu}/\kappa$. Through this relation, r represents the distance at which the correlation is close to 0.1 for all ν.

The flow of individuals from one compartment to another is unidirectional and with no recovery option. Thus, as the number of infected individuals increases, the number of susceptible individuals decreases, with no possible return to the susceptible state. Under these assumptions, we can define a deterministic compartmental model that describes the disease progression by means of the solution of the following Ordinary Differential Equations (ODEs):

$$\begin{aligned} \frac{dS}{dt} &= -(\lambda b I_a + b I_s)S, \\ \frac{dI_a}{dt} &= (\lambda b I_a + b I_s)S - \sigma I_a, \\ \frac{dI_s}{dt} &= \sigma I_a. \end{aligned} \tag{5.2}$$

The parameter b includes the transmission rate β and the spatial dependence as:

$$b = \beta \sum_{i \in \Omega_{S(t)}} \frac{1}{|\Omega_{S(t)}|} \sum_{j \in \{\Omega_{I_a(t)}, \Omega_{I_s(t)}\}} C_{ij}, \tag{5.3}$$

such that C_{ij} is the spatial correlation between individuals i and j, where $\Omega_{S(t)}$ is the set of susceptible individuals at time t and $\{\Omega_{I_a(t)}, \Omega_{I_s(t)}\}$ the set

of infectious individuals, asymptomatic and symptomatic, at time t. Thus, b encompasses, together with the transmission rate β, the mean for the set of susceptible individuals i of the cumulative spatial correlation for each individual i to the infectious individuals j.

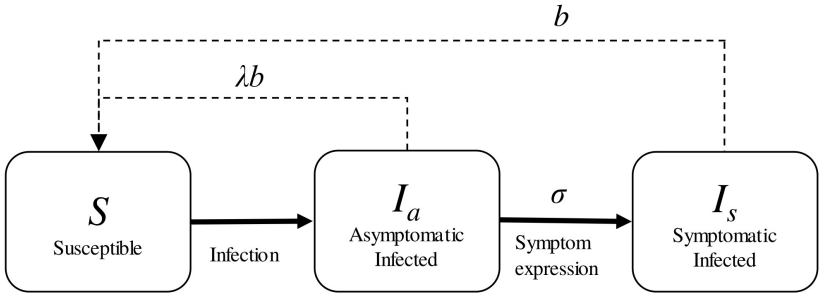

FIGURE 5.1: Compartments and dynamics of the epidemiological model. The solid arrows indicate movement of individuals between states. The population starts as susceptible (S), after infection goes through an asymptomatic infected (I_a) period and continues with the expression of symptoms, becoming symptomatic infected (I_s). The dashed arrows indicate the disease transmission (b) from infectious (I_a and I_s) to susceptible individuals, where b is given by the transmission rate and spatial correlation. λ is a transmission reduction parameter for I_a individuals.

When real data are available, specifically including the number of susceptible and infected individuals, both symptomatic and asymptomatic, at different time points, it becomes feasible to estimate the parameters of the system, λ, β, σ, and the parameters of the Matérn correlation function. From a deterministic perspective, one approach to estimate these parameters is by minimizing the distances between the solution, which relies on the parameters, and the observed data at each time point (least square method). In this work, given the limited availability of data and that the main objective is to simulate the disease spread from the origin, we focus on analyzing the impact of these parameters on the spread at individual level, where the system of equations presented is used as a basis for obtaining a reference value for the transmission rate.

5.2.2 Spatial individual-based model

The system of ODEs in equation (5.2) provides a simplified representation of disease progression at the population level. In contrast, individual-based models allow a more detailed analysis of disease spread, taking into account individual characteristics, interactions and spatial distribution. In individual-based models, the infection of a susceptible individual depends on the force of infection (φ_i), which incorporates the prevalence of infected individuals, i.e., the infected proportion of the total population, the transmission rate, and the spatial interaction between individuals (Keeling and Rohani, 2008). In line with this and with the proposed compartmental model (equation 5.2), our proposal for the force of infection parameter φ_i, for a susceptible individual i at the time t, includes a dependence on the surrounding infected individuals through the Matérn correlation $C(d_{ij})$ for each pair of individuals i and j at a Euclidean distance d_{ij}:

$$\varphi_i(t) = \lambda\beta \sum_{j \in \{\Omega_{I_a(t)}\}} C(d_{ij}) + \beta \sum_{k \in \{\Omega_{I_s(t)}\}} C(d_{ik}), \tag{5.4}$$

where j and k represent each of the I_a and I_s individuals, respectively, of the infected population at time t (i.e., $\Omega_{I_a(t)}$ and $\Omega_{I_s(t)}$); $\beta > 0$ is the transmission rate of infection; and $0 \leq \lambda \leq 1$ represents a decrease in the transmission rate of I_a with respect to I_s.

Given this force of infection, the probability of a susceptible individual becoming infected at time t is $P_i(t) = 1 - \exp(-\varphi_i(t))$ (Keeling and Rohani, 2008; Deardon et al., 2010).

5.2.3 Simulation algorithm

An algorithm to simulate disease spread based on the previous model was implemented using the `Python` programming language (Python Software Foundation, 2021). The `Numba` library (Lam et al., 2015) was also used due to its computational efficiency with high-dimensional data and its ability to run processes in parallel threads.

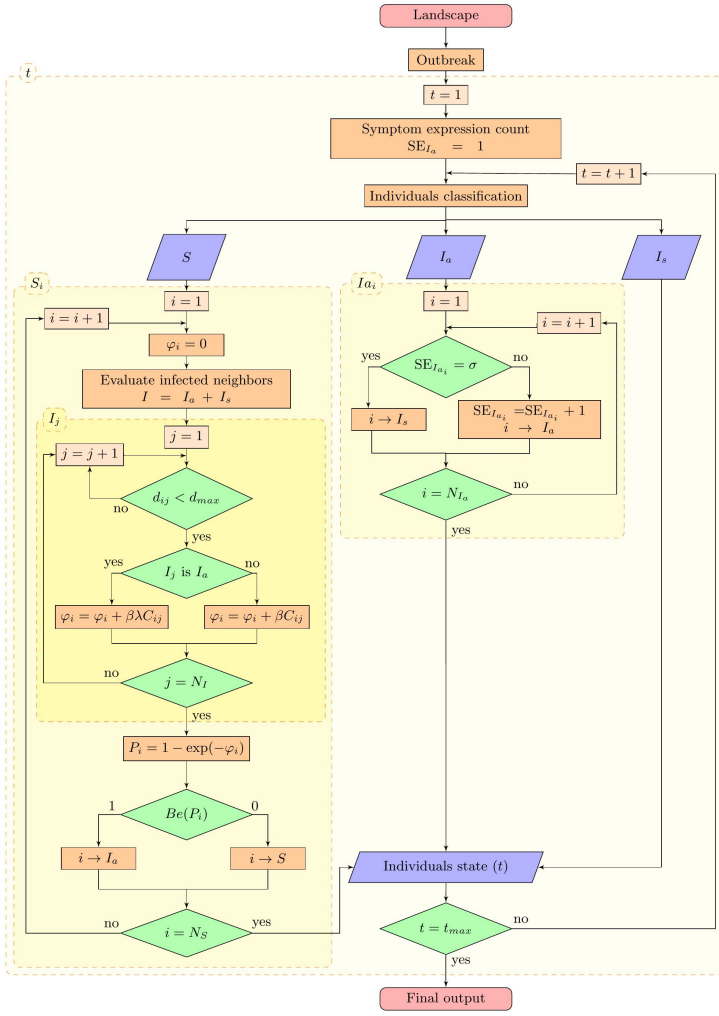

FIGURE 5.2: Diagram of the simulation algorithm. S and I indicate the susceptible and infected states, respectively, where I differentiates asymptomatic (I_a) and symptomatic (I_s). N_S, N_I and N_{I_a} are the total number of individuals in each corresponding state. d is the Euclidean distance and d_{max} is the maximum distance to evaluate neighbors. φ represents the force of infection, β is the transmission rate of infection, reduced by the λ parameter for I_a. C is the spatial correlation, P is the probability of infection, and $Be(P)$ denotes a random value of Bernoulli distribution with probability P. t represents the time and t_{max} the maximum time set for the simulation. The symptom expression SE counter starts when a susceptible individual becomes infected (I_a), so that when the given time for symptom expression (σ) is reached, I_a becomes I_s. The dashed line squares contain the loops performed by the algorithm.

As shown in the diagram of the algorithm in Figure 5.2, the landscape, i.e., the georeferenced locations of the individuals, must first be defined. The onset of the outbreak is initiated with an initial disease introduction I_a at $t = 0$. At each time t, S individuals are evaluated. For each S_i, φ_i is calculated based on the surrounding infected individuals at a distance less than $d_{max} = 1.5r$, taking into account whether they are I_a or I_s following equation (5.4). This maximum distance to compute the spatial correlation allows the algorithm to be optimized, considering that at a distance greater than r the correlation between two locations is almost zero while capturing the tail of the correlation function. The infection of individual S_i is assessed by the discretization of the probability of infection P_i, through a random variable X_i from a Bernoulli distribution $Be(P_i)$, i.e., if $X_i = 1$ individual i becomes infected I_a, and if $X_i = 0$ individual i does not become infected.

Likewise, for each t the I_a individuals are evaluated. For this purpose, a counter is initialized at $t = 1$ for the symptom expression of the individuals in this state. When the given value for the parameter σ is reached, individual i becomes I_s. Finally, the result of the state of all individuals for each t is obtained.

5.3 Case study

This section describes the rationale and the case study selected for the application of the individual-based spread model. We performed several simulations with the algorithm to assess disease spread. Simulations were performed in the study area to compare the disease spread in space and time under different scenarios, and to evaluate the efficiency of the algorithm with a large number of individuals. To interactively visualize the results of all scenarios, a `Shiny` application (Chang et al., 2021) was developed using `R` software (R Core Team, 2022). It is available at `https://spatial-ibm.shinyapps.io/spread_results_app/`.

5.3.1 Study area, population and parameters

The study area covered the *X. fastidiosa* infested area in Alicante. To simulate the spread of ALSD at the individual level, the Agricultural Plot Geographic Information System (SIGPAC) database and the grid used in the official surveillance program were merged to generate the georeferenced distribution of almond trees. From SIGPAC, we selected the plots identified as 'nut trees' and 'nut trees with other associated crops' (MAPA, 2021b). The study area was defined to the boundary of the grid cells with samples from the official surveillance program. As the locations of the trees were not available in the databases, the georeferenced distribution of almond trees was generated with a 7x7 m tree row spacing, which is frequent in the traditional system of almond tree planting (Segura et al., 2018) in the selected SIGPAC plots. Our study area thus consisted of 282,041 almond trees (Figure 5.3).

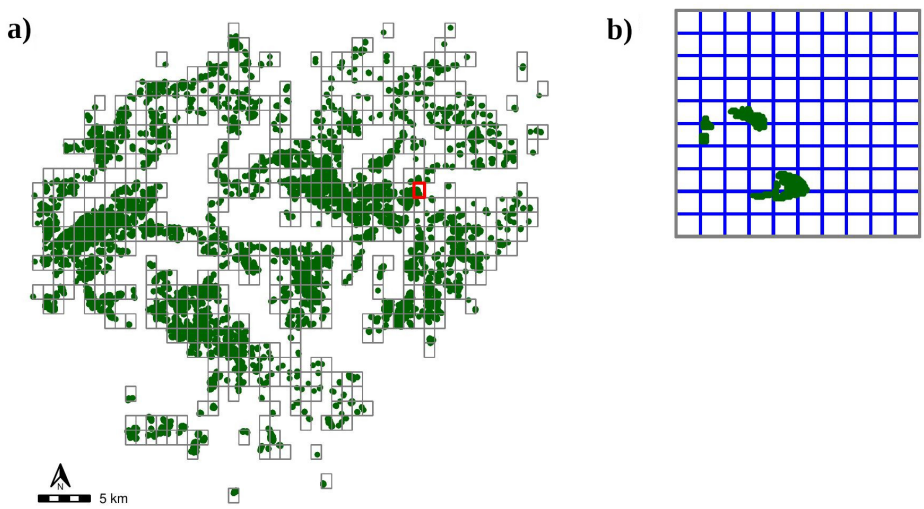

FIGURE 5.3: (**a**) Georeferenced distribution of almond trees (green) generated in the affected area of Alicante. The grid of cells aggregated to 1 km^2 is represented, where the cell marked in red has been zoomed in (**b**) with the corresponding grid of of 1 ha cells used in the simulations, and the almond trees corresponding to these cells (green).

The effect of the model parameters (β, r and ν) on disease progression was compared by simulation. 100 I_a infected individuals were initially introduced and a 360-month time period was simulated. The disease spread was also evaluated according to the type of initial introduction. Specifically, three types of initial introduction were evaluated, referred to hereafter as random, at 5 foci and at one focus. In the case of random introduction, the initial 100 I_a individuals were randomly assigned, generating a more or less geographically dispersed distribution of infected individuals over the study area. For the initial introduction at one focus, an individual was randomly selected in the study area, defining the center of a circumference with a radius of 1-km where the remaining the individuals were then randomly selected. The same process was used for the introduction at 5 foci, initially selecting 5 individuals randomly from the population, and assigning the remaining in a radius of 1-km around them so that each focus consisted of 20 individuals I_a. In this way, each initial introduction type started with the same number of I_a with the different configurations relative to geographic location. In all cases, random selection was performed by non-replacement sampling where all individuals had the same probability of being selected.

Due to the relatively slow progression of the disease (Moralejo et al., 2020), simulations were performed at monthly time-steps to improve computational efficiency. Although molecular analyses indicate that the first outbreak of the disease in Alicante was decades ago (Landa et al., 2020; Moralejo et al., 2020), it was not detected until 2017. Therefore, data for estimating parameters describing the disease from its origin are scarce. Official surveillance data for *X. fastidiosa* from 2018 to 2021 were used to compare the prevalence observed with the results of the fitted compartmental model in equation (5.2) and obtain a reference value for β. The compartmental model was fitted for 30 years with an initial outbreak of 100 individuals I_a of the approximate almond tree population in the study area (282,041 almond trees). Since there was unavailable information on the asymptomatic period, the parameters related to a reduction in the transmission rate of I_a (λ) and the rate of transition from asymptomatic to symptomatic ($\sigma = 1/\text{asymptomatic period}$) were set according to information gathered from the literature as $\lambda = 0.015$ (White et al., 2020) and $\sigma = 1/(8 \text{ months})$ (EFSA, 2019). The spatial dependence introduced in the

compartmental model was obtained from the spatial correlation through the Matérn function (equation 5.1), between each susceptible individual, i.e., negative sampling result, i.e., and each infected individual, i.e., with positive sampling result. The correlation average C of the set of individuals S was used from the accumulated correlations for each individual S following equation (5.3). The model was run with 100 values of β between 0.001 and 0.1 with C obtained given four values for the range parameter r of the Matérn correlation function: 250 m, 500 m, 750 m and 1000 m (Table 5.1). The value of β that minimized the sum of the squared error of the infected individuals (I) predicted by the model and the observations was considered optimal. The sum of the squared error was given as $\sum_{i=1}^{n} \left(I(t_i, \beta) - I(t_i)\right)^2$, where n is the total time t evaluated. For this purpose, the proportion of infected individuals predicted by the model for the last 4 years, i.e., $n = 4$, and the proportion of infected individuals accumulated from the samples in relation to the total number of samples for each year were used. As a result, the optimal value obtained for β was between 0.006 and 0.014 month^{-1}, with $r = 1000$ m and $r = 250$ m, respectively (Table 5.1). However, it must be taken into account that this optimization was only based on the data available for the last few years.

TABLE 5.1: Average of the spatial correlations C obtained for the set of susceptible individuals with each of the values of the range parameter r of the Matérn correlation function, and the corresponding optimal transmission rate β values obtained by by the sum of the squared error of the surveillance data and values predicted values by the compartmental model.

r (m)	C	β (month^{-1})
250	2.12	0.014
500	3.15	0.01
750	4.19	0.007
1000	5.39	0.006

5.3.2 Results of disease spread simulations

In the following we present the results obtained when simulating with different parameter values and different initial introductions in order to evaluate: i) the effect of the different parameters on the disease spread, ii) the variability depending on the initial introduction, and iii) the intrinsic variability of the model. Because of the large number of simulations and computational cost due to the large number of individuals, infection probabilities were calculated using a grid of 1 ha cells (Figure 5.3). At each time-step, one individual S per cell was randomly selected and the probability of infection of this individual was assigned to the other individuals S in the same cell. To test whether this approach produced biases in the results, we compared the results obtained with the simulation of disease spread by calculating the probability of infection of each individual S at each time t.

5.3.2.1 Effects of parameters

Disease spread was simulated with the combination of different values for the parameters β, r and ν, with initial introduction of the infected individuals random, at 5 foci or at one focus. For each of these types of introduction, the location of the initial infected individuals was fixed at the same location. The disease spread simulations were performed with β values of 0.005 month^{-1}, 0.015 month^{-1} and 0.03 month^{-1}, thus covering the range of optimal values of the transmission rate obtained from the surveillance data of *X. fastidiosa* in the study area with the compartmental model. Different values of the parameters of the Matérn r and ν correlation function also were tested. The values of r used, as for the β optimization, were 250 m, 500 m, 750 m and 1000 m, representing a reasonable range of values for the spread distance by insect vectors. The smoothness parameter ν was set to 0.5, 1 and 1.5. The transmission rate reduction for I_a was set as $\lambda = 0.015$ and the time to symptom expression at 8 months for all simulations (EFSA, 2019).

The results were similar for the three values of ν, regardless of the other parameters (β and r) or the type of initial introduction (Figure 5.4). The slight variations when the initial introduction was aggregated at 5 foci or

at one focus can be attributed to the intrinsic variability of the simulations. Therefore, due to the lack of variation concerning this parameter, $\nu = 1$ was fixed in subsequent simulations.

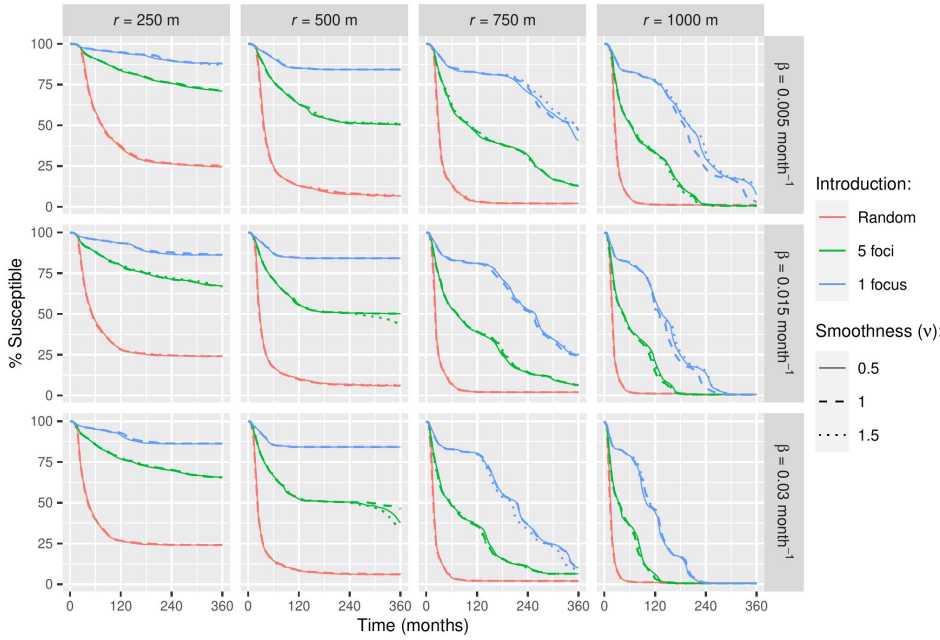

FIGURE 5.4: Percentage of susceptible individuals over time for the simulations with different values of the transmission rate (β), the parameters of the Matérn correlation function, range (r) and smoothness (ν) and the type of initial introduction (random, 5 foci and one focus).

With the same r and the same type of introduction, few differences were found when varying β. With the lowest β, the percentage of S decreased moderately slower. Despite this, after 360 months, the results were almost the same. The only notable difference after this time period were obtained with the introduction at one focus and $r = 750$ m, where the percentage of S was 36.24% higher with the lowest β compared to that obtained with the highest β, while in all other combinations this difference in the percentage of S did not exceed 7%.

Simulations with a large r value resulted in a higher infection rate. Nevertheless, the spread of the disease was different depending on the type of

initial introduction. When the initial introduction was aggregated at one focus, with low values of r the decrease in the percentage of S individuals stabilizes in a relatively short period of time. However, with $r = 1000$ m it continues to decrease at different rates over time. This behavior was also observed when the initial introduction was aggregated at 5 foci. With the two lowest values of r, however, the number of susceptible individuals continues to decrease slowly without stabilizing, and with $r = 1000$ m the fluctuations in the infection rate were less pronounced. These variations were not observed with the random initial introduction, where the number of infected individuals almost stabilized at 96 months at about 2 % and 0.6 % of S with $r = 750$ m and $r = 1000$ m, respectively, while with the lowest values of r the disease continued to spread slowly.

Large differences were found depending on the type of initial disease introduction and on the value of r. While in the case of random introduction the percentage of S fell rapidly in the first few months, the decline was more gradual when introduction was aggregated, being slower the higher the aggregation, i.e., when the introduction was at one focus. Disease spread was limited in all cases by the value of r, so that at the highest value of r, regardless of β or introduction, after 360 months only 0.6 % of S remained. At $r = 750$ m, disease spread was rapid with random introduction and stabilized after a few months at 2 % of S, whereas with both aggregated introductions the decline continued without appearing to stabilize during this period. At the two lowest values of r (250 m and 500 m), the percentage of S practically stabilized in all cases. The higher r and the less aggregated the introduction, the faster this stabilization point was reached and the lower the percentage of remaining S, i.e. the disease affected a greater number of individuals.

5.3.2.2 Variability due to initial introduction

To study the variability due to the type and geographical location of the initial introduction, 10 simulations were performed for each type of introduction, random and aggregated at 5 foci and at one focus, with each combination of β and r parameter values. In each simulation, the location of

the initially infected individuals was different, but in accordance with the type of introduction, choosing the introduction of the I_a as described above.

The dynamics of disease spread were similar to those observed previously in all cases (Figure 5.5). However, while the variability between different introduction configurations was very low for random introduction, large differences were observed for aggregated introduction, depending on the location of the initial foci of infected individuals.

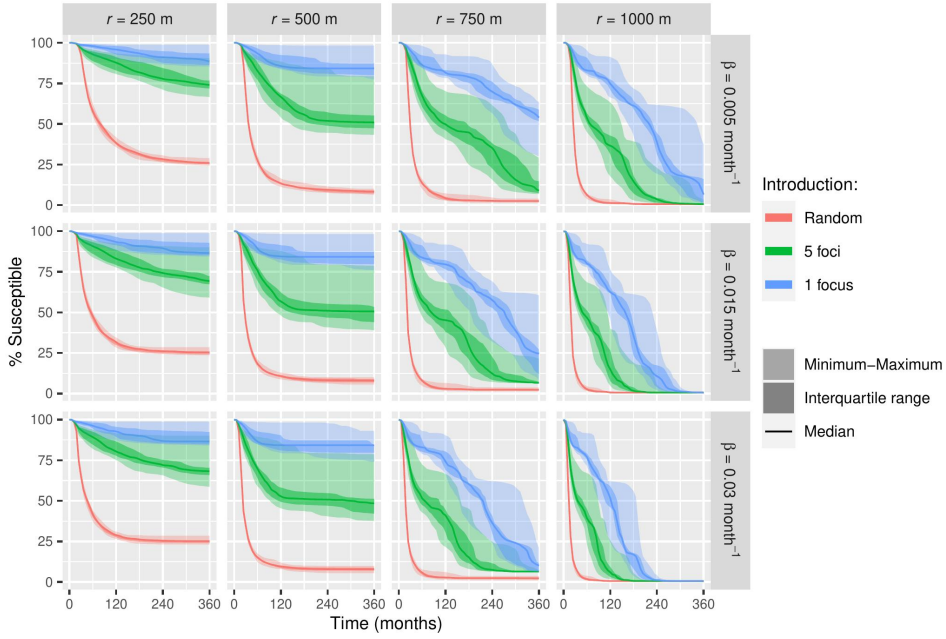

FIGURE 5.5: Minimum, maximum, interquartile range and median of the percentage of susceptible individuals over time of the 10 different geographical locations for each type of initial introduction (random, 5 foci and one focus), with different values of the transmission rate (β) and range (r).

The difference between the maximum and minimum percentage of S at 360 months ranged from 0.22 to 5.4 with random introduction, this difference decreasing as r increased, while different values of β produced similar results. With the two types of aggregate introduction and the lower values of r, the variability gradually increased in the first months and then almost

stabilized, with the median being closer to the minimum. In the simulations with one initial focus located in an area at a greater distance than r of the rest of the individuals, the spread of the disease was practically null.

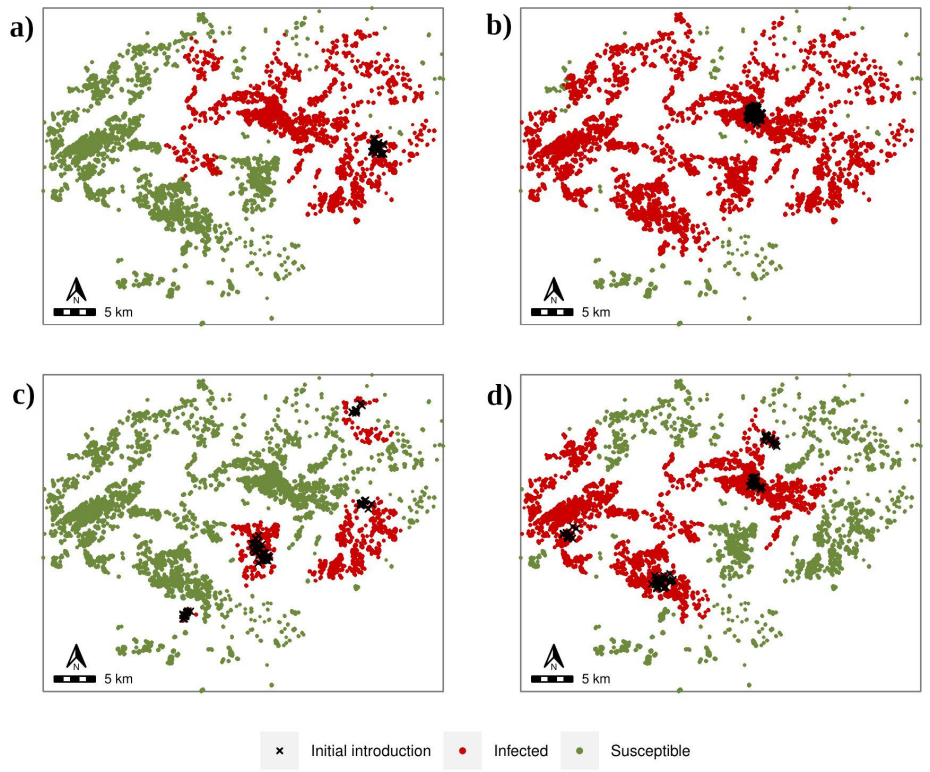

FIGURE 5.6: Simulations with which the greatest difference was obtained depending on the geographic location of initial introduction. (**a**) and (**b**) minimum and maximum number of infected individuals after 360 months, respectively, with initial introduction at one focus, transmission rate $\beta = 0.015$ month^{-1} and range $r = 750$ m. (**c**) and (**d**) minimum and maximum infected individuals after 360 months, respectively, with initial introduction at 5 foci, transmission rate $\beta = 0.03$ month^{-1} and range $r = 500$ m.

The greatest variability was observed when the introduction was at one focus, with the largest difference with $r = 750$ m and $\beta = 0.015$ month^{-1}, where , where the percentage of S was 60 % when the initial focus occurred towards the eastern end and 8 % when the introduction was in the central

area, leading to a greater disease spread (Figure 5.6a, b). However, with the same r and β, this large difference was observed in earlier months, but after 360 months the difference decreased to 14.4 %. The initial introduction at 5 foci had the highest variability in the final percentage of S with $r = 500$. Specifically, with$\beta = 0.03$ month^{-1}, the final percentage of S ranged from 79 to 38 % (Figure 5.6c, d). In these cases, the random introduction of the 5 foci resulted in two of them being very close together, even so, as shown in Figure 5.6, the location of the foci and the distance between the individuals themselves was a determining factor in the progress of the disease. For both types of aggregate introduction and the higher value of r and β 0.015 or 0.03 month^{-1}, although differences were observed at the different times, the difference in the percentage of S at the end was zero.

5.3.2.3 Variability due to intrinsic stochasticity

The intrinsic variability of the stochastic model was observed by running 10 times each combination of β, r, and the three types of initial introduction. For each type of introduction, the same configuration of the location of the initially infected individuals was fixed. Disease dynamics showed the same behavior as described above, with very little variability observed in the different simulations for each combination (Figure 5.7). The only notable differences were those obtained in the last few months with the introduction at 5 foci, $r = 500$ m, and $\beta = 0.03$ month^{-1}, where the difference between the maximum and minimum percentage of S after 360 months was 19.3 %. Despite this, the difference between the maximum and minimum of the final S percentage was between 0 and 5 % with the introduction at 5 foci for the rest of the combinations. With random introduction, the difference between the maximum and minimum percentage of S at 360 months ranged from 0 to 1.2 %, converging with the highest r values. With the introduction at one focus the difference ranged from 0 to 9.6 %, with the largest difference corresponding to the simulations performed with $r = 750$ m and $\beta = 0.005$ month^{-1}.

In order to test whether calculating the probability of infection per 1 ha cell produced a bias in the results, the simulation of disease spread was performed by calculating the probability of infection for each individual

S with $\beta = 0.03$ month^{-1}, $r = 1000$ m and each of the three types of introduction. The results obtained with these simulations did not differ from the approximation, with the percentage of S individuals within the interval obtained with the previous simulations (Figure 5.7).

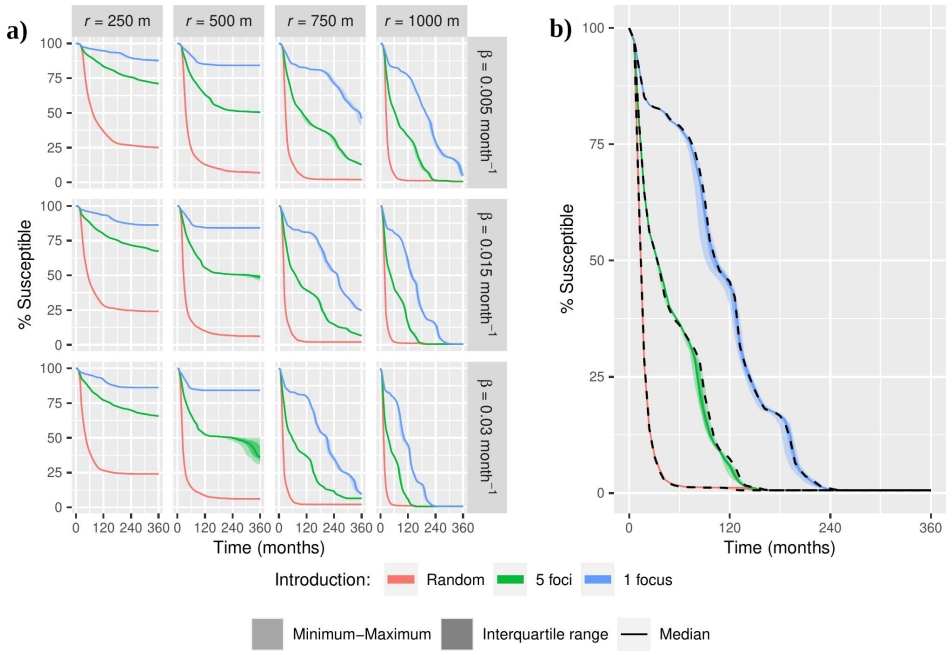

FIGURE 5.7: (**a**) Minimum, maximum, interquartile range and median of the percentage of susceptible individuals over time of the 10 simulations with each combination of the different values of the transmission rate (β), the range (r) of the Matérn correlation function, and the type of initial introduction (random, 5 foci or one focus). (**b**) The dashed line corresponds to the simulations in which the probability of infection was calculated for each susceptible individual with $\beta = 0.03$ month^{-1} and $r = 1000$ m, for each type of initial introduction (random, 5 foci and one focus), overlapping with the intervals of the 10 simulations obtained with the same values of the parameters β and r, where the probability of infection was calculated for each 1 ha cell.

Depending on the values of the parameters r and β and the type of introduction, for simulations where the probability of infection was calculated as a function of the 1 ha cell grid, the computation time varied between 53

s and 609 s, with the slower the disease progression, the longer the computation time. However, for the simulations where the probability of infection was calculated for each individual, the computational time varied between 12 min and 36 min.

5.4 Discussion

Epidemiological models are useful tools for the study of diseases, since they assist in the design and implementation of control strategies to avoid or minimize disease spread. Many factors can influence disease progress, including the environment, human interventions, and means of spread, among others. Both the complexity of taking all these factors into account and the lack of suitable data make these models an approximation of reality. Therefore, model complexity and the factors to be taken into account are often determined by the specific objectives to be achieved. Compartmental models with differential equations are widely used in the study of disease epidemics, allowing an overall description of the progression of the disease at the population level. However, these models do not allow identification of relevant factors in the potential spread of the disease, such as disease status at the individual level and spatial heterogeneity.

The main objective of this work was to develop a tool for the study of disease spread at the individual plant level with heterogeneous spatial distribution. This model can be used as a basis for different applications including epidemiological surveillance, risk assessment, disease control, and impact studies. One of the main issues when dealing with models at the individual level is the high computational cost due to the large number of iterations. In this study, the algorithm was applied to a simplified and generalizable model based on three compartments (S, I_a and I_s) and their respective interactions, also integrating the spatial dependence. This model allowed us to simulate the spread of the disease under different conditions with feasible computational times. The spatial correlation was included by means of the Matérn correlation function, due to its flexibility and the interpretability of its parameters, which is widely used in diverse scientific

disciplines (Guttorp and Gneiting, 2006). Despite this, so far it has not been integrated in individual-based models in plant disease epidemiology.

When data are available that collect the number of individuals in each of the disease states at different times, it is feasible to estimate the parameters associated with the system. As detailed above, it can be approached from a deterministic perspective, however, opting for a stochastic or random scenario is more advisable due to the presence of measurement errors in the data and the influence of external factors that can impact the system. Two main classes of differential equations with uncertainty exist, stochastic differential equations (SDEs) and random differential equations (RDEs). In the case of SDEs, differential equations are influenced by an irregular stochastic process. A classic example of an SDE is a differential equation perturbed by a term dependent on a white noise variable, calculated as the derivative of the Wiener process or Brownian motion (Oksendal, 2007; Gard, 1988). To solve an SDE, the so-called Itô calculus can be employed, based on the application of the Itô lemma (Gard, 1988). On the other hand, an RDE represents a natural generalization of the deterministic counterpart (Soong, 1973; Strand, 1970). In RDEs, the input parameters are considered random variables rather than fixed constants, respectively. Random differential equations constitute a natural generalization of their deterministic counterpart. Random effects are directly manifested in parameters, which are considered random variables or stochastic processes, i.e., parameters are assumed to have regular sample behaviour described by standard probabilistic distributions. By incorporating a stochastic approach and considering the inherent uncertainty, both SDEs and RDEs offer valuable insights and flexibility when dealing with real-world models.

In our case, given the limited availability of data, we employed a compartmental model, which includes the spatial dependence between individuals using the Matérn function, to obtain a reference value for the transmission rate parameter. Subsequently, based on this and information provided by the literature, an individual-based model is used for disease spread, where we compare how the parameters affect the solution of this spatiotemporal model. If data were available, and as previously indicated, we could have introduced uncertainty in the system of differential equations, considering the parameters as random variables. On the other hand, even if data from

the origin of the introduction are not available, the simulation model used can provide a retrospective view.

The effects of the model parameters in disease spread were compared individually by simulating all possible combinations of the selected values. Due to the lack of suitable temporal data for inference on the β parameter, it was approximated to a reasonable value based on official surveillance data for *X. fastidiosa* in the study area. Due to the limitations of obtaining this parameter, other values around this parameter were also compared by the simulations. It was found that the r parameter of the Matérn correlation function as well as the type and location of the initial introduction were highly influential in the spread of the disease. For the highest values of r, almost the entire population was infected, regardless of β or the type of initial introduction. At lower r values, at certain times the spread of the disease was stopped, even though not all individuals were infected. The time of this interruption depended on the type of initial disease introduction and the value of r. The number of infected individuals increased with r, but the type of initial introduction also had a strong influence on the spread of the disease. When the initial introduction was random, with $r = 250$ m the incidence was even higher than when it was aggregated at a focus with $r = 500$ m. These values of the range parameter can be associated with the distance of natural spread of the disease, i.e. by insect vectors, so it would be realistic to consider that if the insects have greater flight distance capacity they can infect more distant trees. Therefore, information on these could be used for the estimation of this parameter of the Matérn correlation function.

The variability of disease spread as a function of the type of initial introduction showed the importance of the location of this first introduction, particularly if it was aggregated. When the infection was initiated at an isolated location, the surrounding area without trees acted as a barrier to spread when r was less than the extent of this area where host plants are not present. On the contrary, in areas with higher population density, i.e., without large empty spaces, the infection rate was almost constant and depended on r. While a random introduction such as the one employed is unlikely to occur, aggregate introductions through the introduction of propagative material may more closely approximate reality.

Due to its simplicity, this individual-based model can be applied to the study of the spread of plant diseases caused by other pathogens by adapting the parameter values. Furthermore, with this basic proposal, the model could be extended with different disease states or dynamics adapted to each case study, such as the recovery or death of individuals, entry of new individuals into the population, elements of disease control, among others. However, it should be noted that more complexity in the model may imply a higher computational cost, as well as a larger number of parameters to estimate if data are available. Moreover, the complexity of the model would imply that it would be more case-specific and therefore less flexible.

In human disease epidemiology the contact matrix is widely used to include the spatial relationship between individuals (Mahmood et al., 2021; Amaral et al., 2023). However, in plant disease epidemiology, different spatial interactions have been associated with plant disease spread, depending on the host distribution, environmental conditions and the biology of the pathogen (Madden et al., 2007). The Matérn correlation function in this model allows the easy integration of diverse spatial relationships between individuals, thus making it capable of handling a variety of plant disease epidemics. This spread model can be used to establish areas at risk of disease, improve epidemiological surveillance, and optimize control strategies. Identifying the disease status of each individual allows the model to operate at a fine scale. With this, more realistic disease spread simulations are obtained and so the efficiency of control strategies can be better assessed.

The study presented in this Chapter has been submitted as:

- **Cendoya, M., Navarro-Quiles, A., López-Quílez, A., Vicent, A., and Conesa, D. (2022). An individual-based spatial epidemiological model for the spread of plant diseases. PREPRINT available at *Research Square*.** `https://doi.org/10.21203/rs.3.rs-2123353/v1`

and is under review in an indexed journal.

Data and code are available at `doi.org/10.5281/zenodo.7128855`. Results can be viewed at `https://spatial-ibm.shinyapps.io/spread_results_app/`.

Chapter 6

Outbreak management plans for emerging plant diseases

In order to continue with a logical progression of the research, the last main objective of this Thesis focuses on outbreak management plans. In this chapter, we use the disease spread model presented in Chapter 5 to explore the effectiveness and efficiency of different outbreak management plans, including surveillance and control strategies.

6.1 Introduction

Emerging diseases and pests are an increasing threat to wild and cultivated plants worldwide. The rates of introduction into new areas has risen over recent years mainly as a consequence of climate change and the increased volumes of travel and trade (FAO, 2019, 2021). Annually, 20-40% of the world's crop production is lost to plant diseases, at a cost of more than $220 billion, and to arthropod pests, at a cost of $70 billion (Savary et al., 2019; FAO, 2019). In a global scenario where food production must increase to satisfy a growing population (FAO et al., 2019), effective management of diseases and pests is crucial to ensure food security (Almeida, 2018; Ristaino et al., 2021).

In response to this situation, plant health has acquired unprecedented prominence in the policy agenda of the European Union (EU). The EU modernized the rules on plant health and in 2019 adopted Regulation (EU) 2016/2031 on protective measures against plant pests and pathogens (EU, 2016) (hereinafter "EU Plant Health Law"). This new EU Plant Health Law addresses the epidemiological surveillance of the EU territory, adjusting the intensity of the measures according to pest and pathogen categorization and prioritization (Montanari and Traon, 2017). They are the so-called Union quarantine "pests", which are subject to strict phytosanitary measures and 20 of them are classified as priority "pests", (EU, 2019) with enhanced requirements.

Despite the implementation of phytosanitary measures for risk reduction and border controls, new introductions of pests and diseases are likely to occur. Between 1999 and 2019, a mean of 10 new introductions of diseases and pests were reported in the EU, with up to 25 new introductions per year (Rosace et al., 2023). Eradication or containment of new outbreaks is only feasible in situations where their extent and prevalence are still limited. Once the pest or disease spreads due to delayed and incomplete detection of outbreaks, their management is much more costly and often not feasible (Václavík et al., 2010; Brown et al., 2020). In this regard, the EU Plant Health Law establishes enhanced risk-based surveillance of the EU territory for early detection of new introductions. Additionally, contingency and action plans are also foreseen for outbreak management.

Epidemiological surveillance of the territory is targeted to specific locations and hosts, following a risk-based structured design to provide a statistical basis for the interpretation of the collected data (Brown et al., 2020). The European Food and Safety Agency (EFSA) provides scientific advice to EU Member States in different fields, including epidemiological surveillance. The EFSA Plant Pest Survey Toolkit (EFSA, 2023) provides the methodological framework and tools for survey planning, specifically to prepare and design surveys under a statistical risk-based approach (EFSA, 2020a,b,c,d). Based on these EFSA guidelines, new requirements for enhanced statistically-based surveillance have been introduced for quarantine pests and pathogens in the EU. Detection surveys are conducted to detect outbreaks in their early stages or to substantiate pest freedom. Delimiting

surveys are carried out once a new outbreak is detected, to delineate the infested zone. Monitoring surveys are implemented to quantify the prevalence of the pest or pathogen in infested zones (EFSA, 2020a; Brown et al., 2020; Lázaro et al., 2021).

The epidemiological information on invading diseases is often limited to the areas of current distribution. Once an outbreak is detected in a new region, risk managers should design and implement a response plan in the absence of exhaustive information on the epidemiology of the disease in this particular area. This lack of knowledge can misguide outbreak response, leading to ineffective plans when actions are too weak or inefficient ones when they are too harsh.

Epidemiological models are increasingly used to understand and predict disease spread as well as to compare different surveillance strategies and control measures for outbreak management (Gilligan, 2008; Ojiambo et al., 2017; Parnell et al., 2017; Hyatt-Twynam et al., 2017; Mastin et al., 2020; Lázaro et al., 2021). These models use different methods to analyze and simulate the spread of diseases and their impact on host plant populations, considering factors such as the environment, host characteristics, and disease spread patterns. Overall, epidemiological models are essential for developing evidence-based disease management strategies, and thereby reducing the economic and environmental impacts of plant diseases.

Although outbreak response is mandatory under the EU Plant Health Law, the performance of outbreak management plans has not been systematically quantified. Control measures should be applied as soon as an outbreak is detected, and thus a baseline scenario of disease spread without control interventions from which to calculate the effectiveness and efficiency of outbreak management plans is seldom available. Over the last few years, individual-based models, also called agent-based models, have gained importance for the study of the outbreak management of invading plant diseases (Parnell et al., 2010; Hyatt-Twynam et al., 2017; Mastin et al., 2020). This type of bottom-up models predict population dynamics by modeling multiple individual agents that interact with one another and with their environment (Zhang and DeAngelis, 2020). In the context of plant disease

epidemiology, individual-based models allow the incorporation of the interactions between individual plant hosts and the pathogen within a spatially explicit landscape. This approach captures the effects of host heterogeneity as well as the spatial structure of disease dynamics, which are essential to design and evaluate outbreak management plans (Cunniffe et al., 2015).

Studies comparing genetic sequences of *X. fastidiosa* strains from Alicante and elsewhere managed to trace back to the origin of this introduction, which was likely initiated with infected plant material from California, USA (Landa et al., 2020). Phylogenetic studies have also suggested an approximate date of introduction in Alicante as far back as the mid-1990s(Moralejo et al., 2020).

Starting in 2017, surveys have been conducted in Alicante to delimit the infested zone and substantiate disease freedom in the surrounding buffer zone. Eradication measures have been progressively applied over recent years, including vector control and the removal of host plants in the infested zone. With a demarcated area of 139,459 ha and 187,400 almond trees already destroyed (GVA, 2023), this outbreak represents one of the largest plant disease eradication campaigns ever attempted in Europe. Considering that the outset of the outbreak has been suggested by molecular epidemiology studies (Landa et al., 2020; Moralejo et al., 2020) to have likely occurred decades ago together with the absence of disease control interventions until after 2017, this case study offers a unique opportunity for a backcasting exercise, where the baseline disease dynamics can be simulated and compared with those with different outbreak management plans.

The main objective of this study was to evaluate the performance of outbreak management plans for emerging plant diseases through an epidemiological approach. Different surveillance and control strategies were considered according to zoning (i.e., buffer zone and infested zone delimitation), including those established by the *X. fastidiosa* Regulation. Almond leaf scorch disease (ALSD) caused by *X. fastidiosa* in Alicante, mainland Spain, was selected as a case study. Surveillance was approached following EFSA (2020a) and EFSA (2020d), comparing the one-step and two-step survey approaches for the first time in plant health, and also including control strategies with a different buffer zone size and eradication radius covering

the minimum values set by the *X. fastidiosa* Regulation. The effectiveness and efficiency of different outbreak management plans were evaluated in relation to a baseline scenario without interventions.

6.2 Methods

6.2.1 Baseline scenario

An epidemiological model was used to simulate the spread of ALSD, caused by *X. fastidiosa* subsp. *multiplex* in the affected area of Alicante, mainland Spain (hereinafter "study area") over a period of 360 months (i.e., 30 years), based on the likely outset of the outbreak dated by the molecular epidemiology studies (Landa et al., 2020; Moralejo et al., 2020) and the absence of disease control interventions until after 2017. The size and shape of the study area were set based on the georeferenced data from the official survey programs implemented from 2017 to 2020. A georeferenced layer with the size and shape of almond plots was extracted from the Spanish Geographical Information System of Agricultural Plots (SIGPAC) (MAPA, 2021b) and overlaid onto the study area. The corresponding georeferenced distribution of individual almond trees within the almond plots was defined considering a 7 x 7 m tree row spacing (Segura et al., 2018). A total of 282,041 almond trees were allocated on a grid of 8,443 cells of 100 x 100 m (1 ha) resolution.

The epidemic process of ALSD was simulated starting with 5 introduction foci, each containing 10 infected trees. Five trees were randomly selected in the study area as the center of each focus. The remaining 9 infected trees in each focus were randomly allocated within a 1 km radius around the center. This pattern depicted a hypothetical introduction of infected propagating plant material imported from California (Landa et al., 2020; Moralejo et al., 2020) and then grafted in a few almond orchards in the study area. Disease spread from the introduction foci was based on an individual-based model that accounted for spatial dependence between individual almond trees (Cendoya et al., 2022b). The same parameterization by Cendoya et al. (2022b) was used to simulate the baseline scenario of disease spread without

surveillance or control interventions, referred to hereinafter as the baseline scenario without interventions. One of the features of most plant diseases, and in particular of those caused by *X. fastidiosa*, is the highly variable asymptomatic period after infection (i.e., incubation period), the duration of which depends on the characteristics of the host and the pathogen and their interactions with the environment (EFSA, 2019). Therefore, here individuals were classified according to their disease status as: susceptible (S), i.e., not infected, asymptomatic infected (I_a), and symptomatic infected (I_s).

The probability of infection P_i of a susceptible tree i depended on the force of infection φ_i at each monthly time-step t defined as:

$$\begin{aligned} P_i(t) &= 1 - \exp(-\varphi_i(t)), \\ \varphi_i(t) &= \sum_{j \in \{\Omega_{I_a(t)}, \Omega_{I_s(t)}\}} \lambda_j \beta C(d_{ij}), \end{aligned} \tag{6.1}$$

which was determined by the transmission rate $\beta > 0$ and the spatial dependence between the susceptible tree i and the infectious trees $j \in \{\Omega_{I_a(t)}, \Omega_{I_s(t)}\}$, where $\Omega_{I_a(t)}$ and $\Omega_{I_s(t)}$ represent the set of trees I_a and I_s at time t, respectively. This spatial information was included in the model by means of the Matérn correlation function $C(d_{ij})$ for two individuals i and j at a Euclidean distance d_{ij}, where the spatial range parameter r approximates the maximum distance at which two individuals are considered spatially correlated (Matérn, 1986).

The transition from I_a to I_s determining the duration of the asymptomatic period (i.e., time for symptom expression after infection). The model considered a transmission rate reduction parameter λ_j related to a lower infectivity of I_a with respect to I_s, where $0 \leq \lambda_j \leq 1$. For $j \in \Omega_{Ia})$ was defined as a constant. Since symptom expression (i.e., $j \in \Omega_{I_s}$), λ_j increased each time t until $\lambda_j = 1$ (i.e., no transmission rate reduction). This gradual increase in λ_j after symptom expression was included in the model to account for both the variability of the duration of the asymptomatic period and the increasing amount of inoculum and infectivity over time. Model parameter specifications and values considered for the disease spread are summarized in Table 6.1. The probability of infection of each S_i was

approximated at the cell level to optimize the computation. Thus, at each time, P_i was calculated for a single individual in each 1 ha cell. Given the probability obtained, infection occurred at an individual level by a random variable from a Bernoulli distribution $Be(P_i)$.

TABLE 6.1: Definition of symbols and values of the disease spread model parameters (β, λ, SE and r).

Parameter	Description	Value	Units
β	Transmission rate	0.015	month^{-1}
λ_j	Transmission rate reduction	0.015 if $j \in \Omega_{I_a(t)}$ $[0.03, 1]$ if $j \in \Omega_{I_s(t)}$	-
SE	Symptom expression	8	month
r	Spatial range	800	m

6.2.2 Outbreak management plans

Outbreak management plans were defined here as the combination of surveillance strategies and control measures in relation to the *X. fastidiosa* Regulation. The surveillance strategies include: i) annual detection surveys in areas where the presence of *X. fastidiosa* is not known; and ii) surveys upon the detection of an outbreak, to demarcate the area for eradication by delineating an infested zone around the infected trees surrounded by a buffer zone where detection surveys are implemented. According to the *X. fastidiosa* Regulation, surveillance strategies should follow the EFSA guidelines for statistically sound and risk-based surveys (EFSA, 2020a,d). Eradication measures in the *X. fastidiosa* Regulation contemplate the removal of all host plants in the infested zone and control of vector populations in the whole demarcated area by means of chemical, biological or cultural treatments.

Surveillance strategies

Based on the *X. fastidiosa* Regulation, the outbreak management plans that were evaluated considered detection surveys in the non-demarcated area

and surveys in the buffer zone (Figure 6.1). According to the *X. fastidiosa* Regulation, annual surveys (i.e., 12 months survey time) were considered in all surveillance strategies.

According to EFSA (2020a), representative sample sizes for detection surveys outside the demarcated area and in the buffer zone were calculated independently based on the statistical design methodology developed by Cannon (2002) taking into account the hypergeometric distribution:

$$n \cong \frac{\left(1-(1-CL)^{1/(N\cdot DP)}\right)\cdot\left(N-\frac{1}{2}\cdot(N\cdot DP\cdot MS-1)\right)}{MS}\,, \qquad (6.2)$$

where the sample size (n) depends on the population size (N) and three interrelated parameters: confidence level (CL), design prevalence (DP), and method sensitivity (MS).

The *confidence level* measures the reliability of the survey procedure (Montgomery and Runger, 2010) by indicating the statistical confidence that the pathogen is absent in the survey area, or that its prevalence is lower than the *design prevalence*. The *design prevalence* denotes the minimum prevalence of the pathogen in an area that the survey would detect given a *confidence level*. That is, if there are no detections in a survey, the prevalence of the pathogen in the survey area should lie between zero and the *design prevalence*. The *method sensitivity* captures the probability of a tree testing positive for the pathogen given that it is actually infected. This parameter is determined by the product of the sampling effectiveness (i.e., the probability of selecting infected plant parts from an infected tree) and the diagnostic sensitivity of the laboratory test.

Following the EFSA guidelines (EFSA, 2020a,d), one-step and two-step survey approaches were considered in the detection surveys outside the demarcated area and in the buffer zone. These two approaches basically differ in how the target population is structured in epidemiological units in the survey area. As defined in the EFSA guidelines (EFSA, 2020a,d), an epidemiological unit is the structural unit for which sample size is estimated, assuming that it is a homogeneous area resulting in similar disease epidemiology. In the one-step survey approach, the non-demarcated area and the

buffer zone are considered two different epidemiological units. In the two-step survey approach, the non-demarcated area and the buffer zone are also considered separately, but divided into a grid of 1 ha cells, each of which represents a single epidemiological unit.

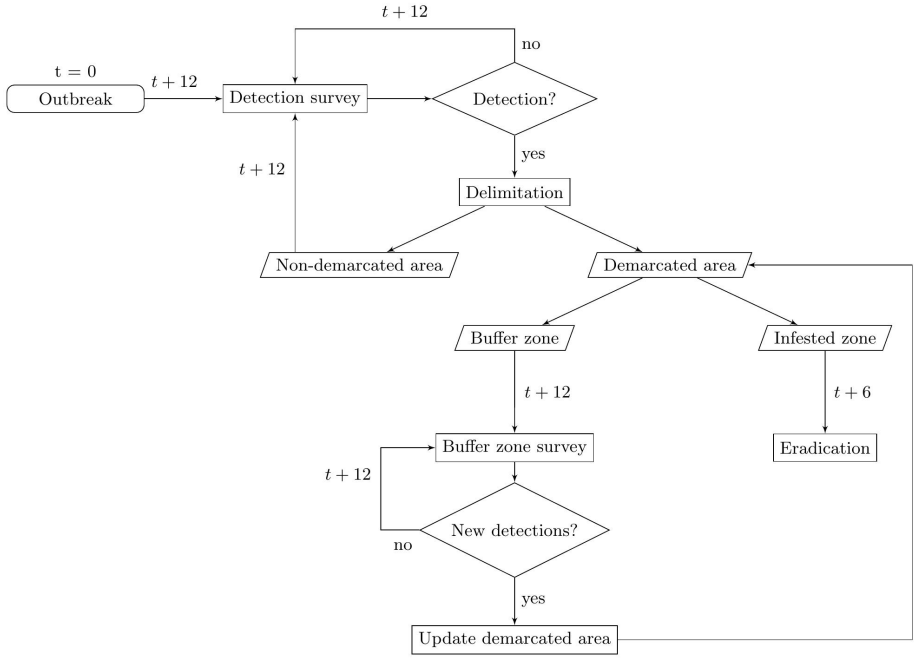

FIGURE 6.1: Flowchart outlining the surveillance and control actions addressed in the proposed generic outbreak management plan over time, with monthly time steps, for a disease that continues to spread within each time step. The outbreak management plan includes detection surveys, zoning of demarcated areas: infested and buffer zones, buffer zone survey, and eradication measures.

One-step survey approach. In the one-step survey approach, sample sizes (n) were calculated by quantifying N as the number of trees in either the non-demarcated area or the buffer zone, respectively, following equation (6.2). The CL was set according to the *X. fastidiosa* Regulation, with $CL_D = 0.8$ for detection surveys outside the demarcated area and $CL_{BZ} = 0.9$ for detection surveys in the buffer zone, referred to hereinafter as $CL_{1,\,low}$. In a second setting for the one-step survey approach, CL_D and CL_{BZ} were increased to 0.99 for both types of detection surveys, referred to

hereinafter as $CL_{1,\,high}$. In all cases, DP was set to 0.01 (*X. fastidiosa* Regulation) and MS to 0.8 (Table 6.2). The samples were randomly selected from the corresponding zone based on the resulting sample size.

Two-step survey approach. In the two-step survey approach, calculations were carried out to determine, first, the number of trees to be sampled per cell (**step 1**) and then the number of cells in the grid to be inspected, both in the non-demarcated area and the buffer zone (**step 2**) (EFSA, 2020d). In both steps, the number of samples and the number of cells were calculated using equation (6.2). In **step 1**, the sample sizes per cell (n_i) were estimated based on the mean number of trees per cell N_i, and it was updated at each survey time according to the cells selected for the detection surveys in the non-demarcated area and the buffer zone. Additionally, the confidence level (CL_i) at this stage was set separately for the non-demarcated area and the buffer zone, referred to as CL_{D_i} and CL_{BZ_i}, respectively. DP_i was set at 0.01 and MS_i at 0.8.

In **step 2**, the number of cells to be inspected (n_h) was calculated based on the total number of cells in the grid (N_h) for the the non-demarcated area and the buffer zone separately, by assigning the confidence level value set in step 1 as the method sensitivity value ($MS_h = CL_i$). Thus, a higher CL_i implies a larger sample size per cell but fewer cells to inspect. DP_h was set at 0.01, and CL_h was defined considering the minimum values established by the *X. fastidiosa* Regulation, $CL_{D_h} = 0.8$ and $CL_{BZ_h} = 0.9$ (referred to hereinafter as $CL_{2,\,high}$) and compared with lower confidence levels, $CL_{D_h} = 0.6$ and $CL_{BZ_h} = 0.7$ (referred to hereinafter as $CL_{2,\,low}$) (Table 6.2).

To maintain the balance between n_i and n_h, CL_i ($= MS_h$) was adjusted at each survey time based on the mean number of trees per cell (N_i) and the total number of cells (N_h) according to the following condition $n_i < N_i$ and $n_h < N_h$. The subsequent specifications were considered to set the value of these two parameters, denoting $min(MS_h)$ as the MS_h that guarantees $n_h < N_h$ and $max(CL_i)$ as the CL_i at which $n_i < N_i$:

$$CL_i = \begin{cases} MS_h = 0.5, & \text{if } 0.5 \in [min(MS_h),\ max(CL_i)], \\ MS_h = max(CL_i), & \text{if } 0.5 \notin [min(MS_h),\ max(CL_i)] \\ & \text{and } max(CL_i) < 0.5, \\ MS_h = min(MS_h), & \text{if } 0.5 \notin [min(MS_h),\ max(CL_i)] \\ & \text{and } min(MS_h) > 0.5, \\ max(CL_i) \text{ and } n_h = N_h, & \text{if } min(MS_h) > max(CL_i). \end{cases}$$

If $min(MS_h) < max(CL_i)$, only a few cells were available for inspection, and thus all cells were surveyed ($n_h = N_h$) considering the $max(CL_i)$.

After n_i and n_h were calculated, the cells to be surveyed were selected randomly. The selected cells were restricted to those with a total number of trees equal to or greater than the total number of samples to be surveyed and survey efforts were estimated as $n = n_i \cdot n_h$. Thus, once the cells were selected, n was reallocated proportionally to each cell based on its population size.

Control measures

Disease control measures for eradication were based on the *X. fastidiosa* Regulation, including the control of vector populations in the whole demarcated area and the removal of host plants in the infested zone, in our case 6 months after a positive finding. Two buffer zone sizes ($BZ = 2.5$ km and $BZ = 5$ km) and two eradication radii ($E = 50$ m and $E = 100$ m) were compared, with the lower values corresponding to the minimum established by the *X. fastidiosa* Regulation. For each combination of BZ and E, three different transmission rates were applied in the buffer zone (β_{BZ}) to assess their impact on disease spread. This β_{BZ} accounts for the effects of vector control in the whole demarcated area and inoculum reduction due to the removal of host plants in the infested zone. Three β_{BZ} values were compared representing different reductions in the transmission rate due to

vector control and inoculum removal: i) No reduction ($\beta_{BZ} = \beta$); ii) 50% reduction ($\beta_{BZ} = 0.5\beta$); and iii) 90% reduction ($\beta_{BZ} = 0.1\beta$).

TABLE 6.2: Parameter setting values of the surveillance strategies. CL_D and CL_{BZ} denote confidence level for detection and buffer zone surveys, respectively; MS is the method sensitivity and DP is the design prevalence.

			CL_D	CL_{BZ}	MS	DP
One-step	$CL_{1,low}$		0.8*	0.9*	0.8	0.01*
	$CL_{1,high}$		0.99	0.99	0.8	0.01*
Two-step	$CL_{2,low}$	- Step 1	Variable	Variable	0.8	0.01*
		- Step 2	0.6	0.7	CL from step 1	0.01*
	$CL_{2,high}$	- Step 1	Variable	Variable	0.8	0.01*
		- Step 2	0.8*	0.9*	CL from step 1	0.01*

*Minimum values established by the *X. fastidiosa* Regulation.

Note that in our case the host plant population in the study area corresponds only to almond trees. Following the *X. fastidiosa* Regulation, all host plants in the infested zone should be eliminated, so the extent of the infested zone was defined by the eradication radius. The settings of the different outbreak management plans evaluated in this study are summarized in Table 6.3.

The simulation algorithm was implemented in `Python` (Python Software Foundation, 2021). Due to the variability inherent to the spread and the outbreak management plans, 50 simulations of the baseline scenario were performed without interventions as well as with the incorporation of each outbreak management plan (Table 6.3), all starting from the same initial 5 introduction foci. The results of these simulations were summarized graphically and numerically using `R` software (R Core Team, 2022). To compare the outbreak management plans with each other and with the baseline scenario without interventions, the simulation with the median number of susceptible trees at $t = 360$ months was used as a reference in each case. This reference simulation is hereafter referred to as the median of the simulations.

TABLE 6.3: Summary of the outbreak management plan settings that were evaluated. Surveillance strategies assessed included the comparison of one-step vs. two-step survey approaches and various confidence level values, CL. The control measures compared considered two widths for buffer zones (BZ) and eradication radius (E) and also addressed the impact of vector control measures and inoculum reduction by varying the transmission rate among individuals within the buffer zone β_{BZ}.

		One-step		Two-step	
		$CL_{1,low}$*	$CL_{1,high}$	$CL_{2,low}$	$CL_{2,high}$*
$BZ = 2.5$ km*	$E = 50$ m*	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$
		$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$
		$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$
	$E = 100$ m	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$
		$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$
		$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$
$BZ = 5$ km	$E = 50$ m*	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$
		$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$
		$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$
	$E = 100$ m	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$	$\beta_{BZ} = \beta$
		$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$	$\beta_{BZ} = 0.5\beta$
		$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$	$\beta_{BZ} = 0.1\beta$

*Minimum values established by the *X. fastidiosa* Regulation.

Effectiveness and efficiency of outbreak management plans

Outbreak management plans involving different configurations of surveillance strategies and control measures were compared in terms of both effectiveness and efficiency. Effectiveness was defined as the reduction in disease spread, measured by the percentage reduction in the number of infected trees with each outbreak management plan (I_m) in relation to the baseline scenario without interventions (I_0). Thus, for each m outbreak management plan, effectiveness$_m = ((I_0 - I_m)/I_0) \cdot 100$, where 0% effectiveness means that

$I_m = I_0$, while 100% effectiveness implies the total removal of infected trees with the outbreak management plan ($I_m = 0$). Efficiency was measured by the percentage increase in susceptible trees that were not eradicated with each outbreak management plan (S_m) in relation to the baseline scenario without interventions (S_0). Thus, for each m outbreak management plan, $\text{efficiency}_m = ((S_m - S_0)/S_0) \cdot 100$, where 0% efficacy implies that $S_m = S_0$, while 100% efficacy means that $S_m = 2 \cdot S_0$.

In addition, inspection intensity (i.e., number of cells to be visited and inspected per year) and survey effort (i.e., number of samples collected per year) were calculated for each outbreak management plan.

6.3 Results

6.3.1 Baseline scenario

The disease spread model without interventions was used as the baseline scenario to compare the different outbreak management plans. The 50 simulations over 360 months starting from the same initial 5 introduction foci showed low variability (Appendix C, Figure C.1). As the disease spread over the study area, these initial 5 foci gradually coalesced and at 360 months they were no longer distinguishable–only the most distant trees remained susceptible (Appendix C, Figure C.2). In all simulations the disease spread quickly during the first months but gradually slowed down over time, particularly as the percentage of infected trees increased. From the initial introduction foci, which accounted for 0.02% of the infected trees, 25% of the population became infected within the first 42–48 months. In all simulations, 50% of the population became infected between months 96 and 102, and 75% between months 186 and 192. The percentage of infected trees at $t = 360$ months ranged from 92.45% to 94.16% and the median percentage of susceptible trees was 6.73%, ranging from 5.84% to 7.55%.

6.3.2 Outbreak management plans

The percentage of infected trees at $t = 360$ months showed little variability among the 50 simulations in each configuration of the outbreak management plans, while the percentages of susceptible and eradicated trees were more variable. The greatest variability was obtained with the two-step survey approach, with a maximum interquartile range of 19.61% for both susceptible and eradicated trees (Appendix C, Figure C.3). In general, the results of these simulations were consistent for all the outbreak management plans, with reduced variability in the final percentage of susceptible, infected, and eradicated trees at $t = 360$ months (Appendix C, Figure C.3). The accumulated surveillance effort and inspection intensity at $t = 360$ months in the simulations for each outbreak management plan showed almost no variability, indicating the consistency of surveillance strategies despite their inherent randomness (Appendix C, Figure C.5).

The percentage of infected and susceptible trees in $t = 360$ months of the median of the simulations obtained with the different outbreak management plans was compared to the median of the simulations of the baseline scenario without interventions. The percentage of infected and susceptible trees in $t = 360$ months of the median of the simulations obtained with the different outbreak management plans was compared with the median of the simulations of the baseline scenario without interventions (Figure 6.2). With the outbreak management plans, the maximum percentage of infected trees relative to the initial population was 36.44%. This value was obtained at $t = 180$ months using the one-step survey approach with $CL_{1,\,low}$, $BZ = 2.5$ km, $E = 50$ m and $\beta_{BZ} = \beta$. The percentage of infected trees at $t = 360$ months ranged from 0% to 17.79% depending on the surveillance strategy and control measures. With the one-step survey approach, the percentage of infected trees was between 1.17% and 17.79% with $CL_{1,\,low}$ and between 0.37% and 7.43% with $CL_{1,\,high}$. With the two-step survey approach, the percentage of infected trees ranged from 0.002% to 5.66% with $CL_{2,\,low}$ and from 0% to 1.67% with $CL_{2,\,high}$. These results implied an effectiveness of an 80.93 - 100% decrease in the percentage of infected trees relative to the baseline scenario without interventions (Appendix C, Table C.1). Complete eradication of infected trees, i.e., 100% effectiveness, was only achieved with

three configurations using the two-step survey approach with $CL_{2,\,high}$ and $BZ = 2.5$ km. In contrast, the lowest effectiveness was obtained with configurations using the one-step survey approach, with $CL_{1,\,low}$, $BZ = 5$ km and $E = 50$ m. Furthermore, the highest effectiveness in each surveillance strategy was obtained with $BZ = 2.5$ km and $E = 100$ m, and the lowest effectiveness with $BZ = 5$ km and $E = 50$ m.

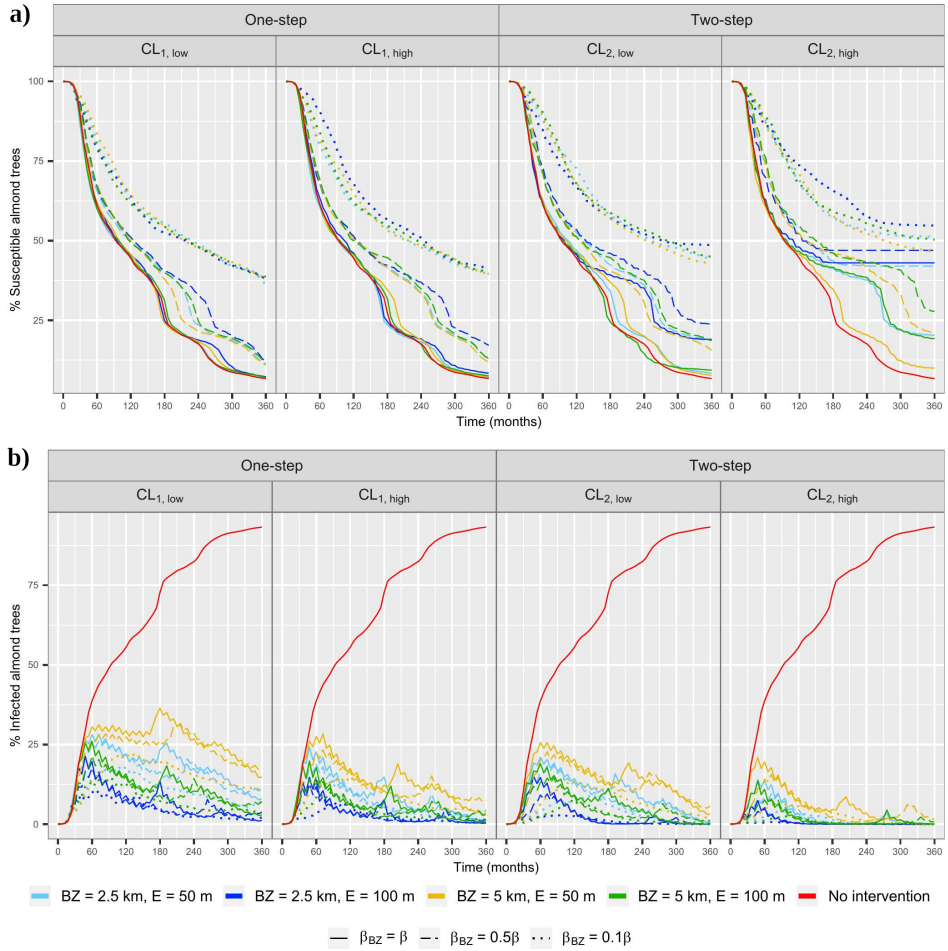

FIGURE 6.2: Effect of outbreak management plans on the spread of almond leaf scorch disease (ALSD) in the study area: percentage of (**a**) susceptible and (**b**) infected trees relative to the total population over time for the median of simulations with each outbreak management plan.

The reduction in the transmission rate in the buffer zone gave rise to the most notable differences in the percentage of susceptible trees at $t = 360$ months. With the one-step survey approach, the percentage of susceptible trees was 6.79 - 8.43% with $\beta_{BZ} = \beta$, 10.93 - 17.15% with $\beta_{BZ} = 0.5\beta$, and 36.09 - 41.13% with $\beta_{BZ} = 0.1\beta$. In all cases, the lowest percentage of susceptible trees was obtained with $CL_{1,\,low}$ and the maximum with $CL_{1,\,high}$. With the two-step survey approach, the greatest effect on the percentage of susceptible trees at $t = 360$ months was also associated with the reduction in the transmission rate in the buffer zone. However, greater differences were observed depending on CL, BZ, and E than with the one-step survey approach. With $\beta_{BZ} = \beta$ the percentage of susceptible trees was 7.77 - 19.02% with $CL_{2,\,low}$ and 10.02 - 43.04% with CL_{high}. With $\beta_{BZ} = 0.5\beta$ the percentage of susceptible trees was 15.70 - 23.89% and 21.09 - 46.95%, respectively, at $CL_{2,\,low}$ and $CL_{2,\,high}$. With $\beta_{BZ} = 0.9\beta$ the percentage of susceptible trees was 42.92 - 48.68% and 46.53 - 54.75% with both $CL_{2,\,low}$ and $CL_{2,\,high}$, respectively. Overall, the two-step survey approach was more efficient in terms of percentage increase of susceptible trees that were not eradicated in relation to the scenario without interventions (Appendix C, Table C.1). Regardless of the survey approach, reducing the transmission rate in the buffer zone, as well as the combination of $BZ = 2.5$ km and $E = 100$ m, increased the efficiency of the outbreak management plan. In contrast, the combination of $BZ = 2.5$ km and $E = 50$ m was the least efficient across all configurations.

The georeferenced distribution of tree status and surveillance results with more extreme values of efficiency for the outbreak management plans are shown in Appendix C, Figure C.7 and C.8. The outbreak management plan with the lowest efficiency was the one that combined the one-step survey approach with CL_{low}, 2.5 km buffer zone, and 50 m eradication radius (Appendix C, Figure C.7). With this configuration, the buffer zone covered practically the entire study area at $t = 180$ months, and the infested zone was still widely distributed over the study area at $t = 360$ month despite the high number of eradicated trees. The outbreak management plan that combined the two-step survey approach with CL_{high}, 2.5 km buffer zone and 100 m eradication radius was the most efficient (Appendix C, Figure C.7). With this configuration, the buffer zone was substantially reduced

compared to the previous one, resulting in 40% fewer trees eradicated and almost all the infected trees eliminated.

The one-step survey approach with $CL_{1,\,high}$ resulted in a reduction in the percentage of infected trees compared with $CL_{1,\,low}$, but with a similar percentage of eradicated trees (Figure 6.3). Likewise, in the two-step survey approach $CL_{2,\,high}$ resulted in a reduction in the percentage of infected trees, but with a lower percentage of eradicated trees, resulting in a higher percentage of susceptible trees than with $CL_{2,\,low}$. With the one-step survey approach the percentage of eradicated trees was about 91% for the most effective outbreak management plan with both $CL_{1,\,low}$ and $CL_{1,\,high}$. The most effective outbreak management plans with the two-step survey approach implied 80.98% of eradicated trees with $CL_{2,\,low}$ and 56.97% with $CL_{2,\,high}$.

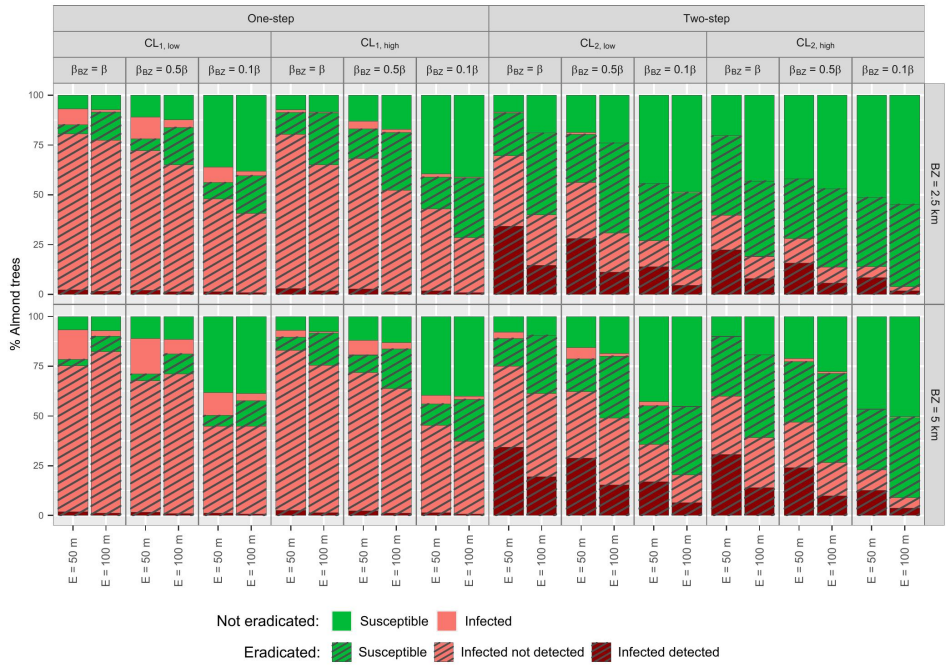

FIGURE 6.3: Percentage of susceptible, infected, and eradicated trees relative to the total population at $t = 360$ months for the median of the simulations with each outbreak management plan. Eradicated trees are differentiated into susceptible or infected at the time of eradication, and infected trees that were also detected by the surveillance strategy.

Increasing the eradication radius from 50 m to 100 m in the one-step survey approach led to an increase in the percentage of total eradicated trees at $t = 360$ months, while it had the opposite effect with the two-step survey approach (Figure 6.3). In both survey approaches, the largest eradication radius of 100 m increased the removal of susceptible trees but also resulted in a lower percentage of infected trees that were not eradicated. Nevertheless, with 100 m eradication radius a higher percentage of susceptible trees not eradicated remained after $t = 360$ months compared to a 50 m eradication radius. For the same eradication radius, a lower percentage of susceptible trees and a higher percentage of infected trees were removed with the 5 km buffer zone compared to the 2.5 km one, but this also resulted in a higher percentage of infected trees that were not eradicated. The highest percentage of susceptible trees removed was obtained with the two-step survey approach with a 2.5 km buffer zone and 100 m eradication radius, removing 45.31% of the susceptible trees and leaving 0.02% of infected trees without being eradicated. The lowest percentage of susceptible trees removed was obtained with the one-step survey approach with a 5 km buffer zone and 50 m eradication radius, removing 3.22% of susceptible trees and leaving 14.76% of infected trees without being eradicated.

The percentage of detection relative to the total number of infected trees varied greatly over time (Figure C.9). In general, the percentage of detection was higher with the two-step survey approach and increased with the confidence level in both survey approaches. The highest percentage of detection was obtained with the combination of a 2.5 km buffer zone and 100 m eradication radius. The maximum percentage of detection in a single survey campaign (i.e., survey at time t) with the one-step survey approach was 2.73% and 5.27% with $CL_{1,\,low}$ and $CL_{1,\,high}$, respectively. With the two-step survey approach, a maximum percentage of detection of 87.88% and 100% was obtained with $CL_{2,\,low}$ and $CL_{2,\,high}$, respectively.

The annual survey effort (i.e., number of samples collected per year) increased with the confidence level and was higher in the two-step than the one-step survey approach (Figure 6.4a, b). In the one-step survey approach, the annual survey effort ranged between 200 and 486 samples with $CL_{1,\,low}$ and between 573 and 1,145 samples with $CL_{1,\,high}$. In the two-step survey approach it ranged between 7,491 and 17,751 samples with $CL_{2,\,low}$ and

between 13,138 and 31,998 samples with $CL_{2,\,high}$. In the one-step survey approach, the variability in survey effort was mainly due to fewer samples in the first two years, remaining relatively constant in subsequent years. Fewer samples were also collected in the first two years with the two-step survey approach, but the survey effort was also rather variable in subsequent years. The buffer zone size, eradication radius, and transmission rate had little effect on the survey effort. In relation to the reduction in the transmission rate in the buffer zone, lower variability was obtained with the two-step survey approach but resulted in a similar overall survey effort.

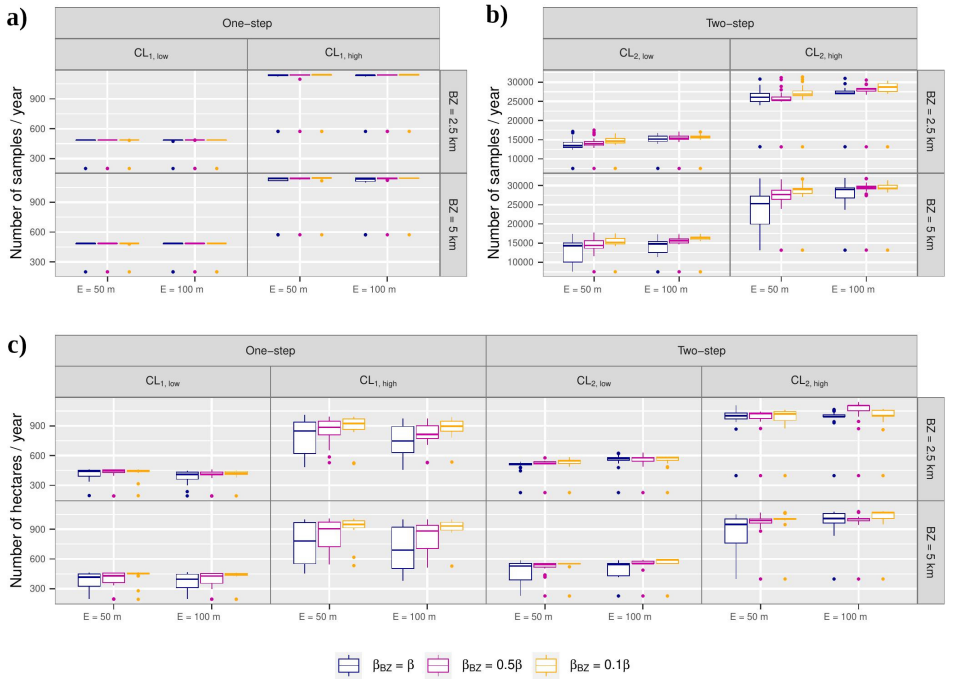

FIGURE 6.4: Number of samples per year (survey effort) (**a**, **b**) and number of inspected cells (inspection intensity) (**c**) for the median of the simulations with each outbreak management plan.

Despite the results obtained in relation to the survey effort, no substantial differences were observed in the inspection intensity with both survey approaches. With the one-step survey approach the numbers of cells inspected per year ranged from 192 to 468 and from 378 to 1,009 with $CL_{1,\,low}$ and $CL_{1,\,high}$, respectively. With the two-step survey approach the numbers

of cells inspected per year were $CL_{2,\,low}$ between 227 and 629 and between 398 and 1,139 with $CL_{2,\,high}$. Inspection intensity was lower in the first two years, before zoning. In subsequent years, higher variability was observed with CL_{high} compared to CL_{low}, particularly in the one-step survey approach (Figure 6.4c). In all cases, lower variability was obtained with the buffer zone of 2.5 km and when reducing the transmission rate.

6.4 Discussion

The disease spread model framework used in our study incorporated information from the molecular epidemiology studies available so as to be able to estimate a plausible starting time of the outbreak, which dated the possible origin of the introduction in the mid-1990s by sequencing and comparing different *X. fastidiosa* strains (Landa et al., 2020), as well as epidemiological and dendrochronological data on *X. fastidiosa* infected almond trees (Moralejo et al., 2020). Through this approach, it was possible to simulate the progression of the disease over time in the absence of diagnosis-based evidence on the date of introduction. A scenario of the dynamics of the epidemic was obtained for a period of 360 months, in which the current situation in the infested zone of Alicante closely resembles that obtained with the model between months 240 and 360. In the baseline scenario without interventions, the disease progressed quickly and 50% of the population was infected after 96 months. When the percentage of infected trees reached about 75%, disease spread slowed down considerably, becoming almost asymptotic even though not all the population was infected (Appendix C, Figure C.1). Control measures for *X. fastidiosa* have been progressively implemented in the study area since the first detection in 2017. However, due to the particular surveillance strategy first established by the EU legislation, where survey activities were strongly concentrated around the positive findings(Lázaro et al., 2021), the full extent of the infested zone was not delimited with relative accuracy until the current *X. fastidiosa* Regulation came into force. Therefore, the cumulative survey data from 2017 to 2020 used in our study provided an approximation of the extent of the infested zone that the disease would have reached without any interventions.

The disease spread model developed in our study was characterized by its simplicity and ease of application, which facilitate its generalization to other plant diseases. Although the incorporation of other elements in the model such as vector dynamics may provide more accurate results in the study of disease spread (Giménez-Romero et al., 2023), they also make it more complex and disease-specific. Despite not explicitly incorporating the role of vectors and other means of disease spread, they are considered in the model through the transmission rate and spatial correlation (White et al., 2020). The distribution of almond trees in the study area was drawn from the available official crop statistics, reproducing the heterogeneity of host distribution that characterizes this region. In contrast to the most common approach using differential equations, the individual-based model used in our study allowed us to deal with this issue, where the spatial heterogeneity of the host plants can affect disease spread. Furthermore, with the individual-based model at the tree level it was possible to quantify the number of infected and eradicated trees, inspection intensity, and survey effort for each outbreak management plan. Based on this quantitative data, further modeling work could provide an estimate of the economic and environmental costs of the different outbreak management plans, allowing risk managers to select not only the most effective or efficient in terms of disease control, but also the most economical.

As new scientific information was obtained from the different *X. fastidiosa* outbreaks in the EU, the regulation was progressively updated making the use of the EFSA guidelines mandatory for statistically sound and risk-based surveys for *X. fastidiosa* (EFSA, 2020a; EU, 2020). Nevertheless, currently our understanding of the effectiveness and efficiency of those established standards is limited, particularly with regard to long-term and fine-scale disease control. In this study we assessed the effectiveness and efficiency achieved with the minimum requirements set by the *X. fastidiosa* Regulation for surveillance and disease control, compared with several variations of these parameter settings. All the outbreak management plans evaluated included early detection, with annual surveys following the introduction of *X. fastidiosa* in the study area. Other studies have also considered the evaluation of *X. fastidiosa* disease outbreak management plans under different zoning strategies, introducing variations in the buffer zone

size or eradication radius (White et al., 2017; EFSA, 2019; Gilioli et al., 2023). These studies also used different surveillance scenarios, under different conditions of surveillance efficiency (White et al., 2017), high inspection efficiency in the first kilometer of the buffer zone and lower in the rest (EFSA, 2019), or whether detection is early or late (Gilioli et al., 2023). Here, surveillance was addressed in greater depth by comparing one-step and two-step survey approaches at different confidence levels, evaluating the effectiveness and efficiency of the different surveillance strategies, and quantifying the survey effort and inspection intensity of each of them.

The two-step survey approach is described in EFSA (2020a), Appendix C. This approach was originally developed for animal health (Martin et al., 1987; Cameron and Baldock, 1998; Wagner and Salman, 2004; FAO, 2014), but to our knowledge there are no precedents of its application to plant health. As an improvement of the methodology described in EFSA (2020a), in our study the two-step survey approach was adapted to consider the heterogeneity of the host plant population in each cell, resulting in varying densities of inspection units. For this purpose, in step 1 of the two-step survey approach the confidence level was varied as a function of the population density per cell in each survey. With this added feature, the methodology could be used to estimate the most cost-effective confidence level for the surveillance strategy (Martin et al., 1987; Cameron and Baldock, 1998).

To compare the outbreak management plans, the initial 5 introduction foci were the same for all the outbreak management plans that were tested in order to remove any possible effect of their location or size. Nevertheless, the disease spread model and surveillance strategies had associated randomness. Therefore, 50 simulations were performed for each outbreak management plan to capture the variability resulting from this randomness. Despite the intrinsic stochasticity of the spread model, simulations for the baseline scenario without interventions showed little variability (Appendix C, Figure C.1). Hence, the differences observed with the outbreak management plans were not attributable to the underlying disease spread. Overall, simulations for each outbreak management plan showed low variability, except in the two-step survey approach without reduction in the transmission rate (Appendix C, Figure C.3). In these cases, differences were observed in the percentage of susceptible trees due to eradication, while the percentage

of infected trees remained relatively constant in all simulations. However, this variability was not observed in the resulting inspection intensity and survey effort, which were relatively similar in all the simulations for each outbreak management plan (Appendix C, Figure C.5).

All the outbreak management plans implemented were effective in controlling the spread of the epidemic. Notable differences were observed until 60 months after the introduction, compared to the baseline scenario without interventions (Figure 6.2). After 360 months of surveillance and control, the percentage of infected trees in relation to the original population was lower than 18% in all the outbreak management plans. Differences in the outbreak management plans as well as the definitions of the epidemiological units here and in EFSA (2019) mean that these two studies are not quantitatively comparable, although in both cases the outbreak management plans implemented were successful in controlling disease spread.

In the outbreak management plans evaluated in EFSA (2019), a larger buffer zone was found to be more effective in disease control, although the effect was not very pronounced. However, in our study, increasing the size of the buffer zone did not result in greater effectiveness. The sample size calculated from the hypergeometric distribution converges to that given by the binomial distribution for large population sizes (EFSA, 2020a). Considering the high population densities and the fact that the same design prevalence was set in all strategies (i.e., $DP = 0.01$), similar sample sizes and thus similar survey efforts were obtained for the buffer zones of 2.5 km and 5 km (Figure 6.4). These results highlight the importance of adjusting the design prevalence based on the population size. For a same proportion of infected trees, the actual number of infected trees increases with the population size, resulting in quite different disease dynamics.

In general, a larger eradication radius was more effective in reducing the percentage of infected trees. Reductions in the transmission rate in the buffer zone to account for the effects of vector control in the buffer zone and inoculum removal in the infested zone did not have a substantial effect on the number of infected trees, but they did have a strong effect on minimizing the number of susceptible trees that were removed. That is, a greater reduction in the transmission rate reduced the number of susceptible trees that were

removed due to eradication, while still eliminating the infected ones, and thus increasing efficiency (Figure 6.2). The reduction in the transmission rate in the buffer zone was set at three levels (0, 50, 90%) to reflect the uncertainty regarding the actual effect of these control measures. Further research is thus needed to quantify the effect of different vector control and inoculum reduction approaches to select those that are more effective when it comes to improving outbreak management plans.

Despite the effectiveness of the outbreak management plans in reducing the number of infected trees, there were large differences in their efficiency depending on the surveillance strategy. These differences were reflected in the percentage of susceptible trees not removed as well as in the inspection intensity and survey effort. Increasing the confidence level resulted in greater survey efforts, associated with better effectiveness and efficiency with higher numbers of susceptible trees left without being removed by eradication. The two-step survey approach also resulted in greater survey efforts and better effectiveness. In fact, only the outbreak management plans including the two-step survey approach managed to completely control the epidemic, while leaving a higher percentage of susceptible trees without being removed due to eradication. Nevertheless, these results obtained with the two-step survey approach were achieved by implementing high survey efforts. Differences in the inspection intensity were, however, not as large between the one-step and two-step survey approaches as with both CL_{low} and CL_{high}. That is, although the two-step survey approach resulted in larger survey efforts, it involved inspection of a similar number of cells as in the one-step survey approach. Survey effort and inspection intensity are important criteria for risk managers when planning surveillance programs, as they are directly related to the logistics and working capacity in the field and in diagnostic laboratories.

The assessment of different outbreak management plans combining surveillance strategies and control measures with the same baseline scenario allowed us to quantify their effect in terms of efficacy and efficiency. Zoning and eradication measures controlled disease spread, but surveillance was the key factor enhancing the efficacy and efficiency of outbreak management plans. Surveillance strategies that involved a greater survey effort, either with a higher confidence level or through the two-step survey approach,

were more effective in terms of detection and resulted in more effective and efficient disease control.

Only outbreak management plans with the two-step survey approach were able to completely remove all infected trees. However, this approach involved a much greater survey effort compared to the one-step approach. In the two-step survey approach the confidence level and method sensitivity were set according to the number of available epidemiological units, maintaining a trade off between the two parameters. Nevertheless, these parameters can also be set based on the costs of inspection intensity and survey effort. In contrast, the one-step survey approach could only be adjusted based on the cost of survey effort, but not inspection intensity. Although the two-step survey approach implied higher cost due to the large survey effort, its higher efficiency reducing the number of susceptible trees removed by eradication should be considered along with the environmental and social impacts associated.

The study presented in this Chapter has been submitted to an indexed journal with the title: "Performance of outbreak management plans for emerging plant diseases: the case of almond leaf scorch caused by *Xylella fastidiosa* in mainland Spain"

Chapter 7

Conclusions and future work

In this thesis different models for the study of the distribution and epidemiology of plant diseases have been presented. Specifically, the presence of *X. fastidiosa* in the Mediterranean basin has been addressed as a real and current problem. In general, tools to better understand the epidemiology have been provided, as well as results that can assist in the management of the disease. The most noteworthy conclusions were:

- Spatial hierarchical Bayesian models using the INLA methodology have allowed us to effectively estimate and predict the probability of pathogen presence in the study areas based on different types of spatial data and to analyze the effect of climatic covariates taking into account the spatial component.
- The analysis of climatic and spatial effects has yielded consistent results despite the different epidemiological settings and the use of two modeling approaches adapted to the characteristics of the spatial data, highlighting the importance of taking into account the spatial component. Furthermore, it has been found that the effect of climatic variables may not be extrapolable, and may depend on the epidemiological characteristics of the outbreak.
- The analysis of the georeferenced data from the *X. fastidiosa* surveys in Alicante using a geostatistical model suggests that the minimum size

of the buffer zone established by the legislation could be insufficient, and has served as a reference for extending the buffer zone in this area. This highlights that control measures should be tailored to the epidemiological setting in each outbreak.

- The spatial hierarchical Bayesian model under the non-stationary approach has allowed to take into account physical barriers, which cannot be ignored when present in the study area due to their potential effect on pathogen spread. Furthermore, this approach has also enabled the inclusion of barriers that simulate phytosanitary control measures that can prevent the spread of the pathogen. This can be especially useful in establishing areas where control efforts should be increased.
- In order to simulate the disease spread from the origin of the outbreak, an individual-based model has been implemented. This model has allowed to handle large populations and to take into account the spatial distribution of hosts. Furthermore, the importance of both the spatial distribution of individuals and the characteristics of the initial introduction is highlighted due to the impact on disease progression.
- To assess the potential performance of outbreak management, back casting simulations and the use of genomic data to date and locate potential introduction foci have been useful. Comparison of the different outbreak management plans has highlighted that surveillance for early detection is critical. Furthermore, we provide a comparison of the effectiveness and efficiency of the surveillance strategies and control measures applied. The results indicated a higher effectiveness and efficiency of the two-step surveillance approach with respect to the one-step approach, but involving a much higher surveillance effort. Additionally, the important role of reducing the transmission rate in the buffer zone leading to an increase in the effectiveness of outbreak management plans is shown.

Overall, the models presented and applied to *X. fastidiosa* can be extended to plant diseases caused by other pathogens. However, the scope of research is still very broad and several challenges remain to be addressed in this context. Therefore, several proposals are presented in the following for future consideration:

- The incorporation of vector information, as well as taking into account other environmental variables, such as wind or plant biodiversity, which may affect pathogen distribution and disease progression, could provide more accurate results.

- The non-stationary approach used assumes that the barriers are impermeable, however, it is a strong assumption in plant health when considering dispersal control barriers. Therefore, further methodological developments needed to consider permeable barriers in plant health.

- New spatio-temporal data could be used to estimate the parameters involved in the disease spread model to provide a more realistic understanding of disease progression in the study area.

- Based on the disease spread model, the integration of surveillance and control data in the study area could be used to predict disease spread taking into account these measures. Also, further consideration should be given to the impact of design prevalence according to population size on surveillance strategies. Furthermore, data on different control measures, such as vector control or inoculum reduction, could be used to accurately predict the effect of these measures.

References

Abboud, C., Bonnefon, O., Parent, E., and Soubeyrand, S. (2019). Dating and localizing an invasion from post-introduction data and a coupled reaction–diffusion–absorption model. *Journal of mathematical biology*, 79(2):765–789.

Akman, D., Akman, O., and Schaefer, E. (2018). Parameter estimation in ordinary differential equations modeling via particle swarm optimization. *Journal of Applied Mathematics*, 2018:9160793.

Akman, O. and Schaefer, E. (2015). An evolutionary computing approach for parameter estimation investigation of a model for cholera. *Journal of Biological Dynamics*, 9:147–158.

Almeida, R. P. (2018). Emerging plant disease epidemics: Biological research is key but not enough. *PLoS biology*, 16(8):e2007020.

Almeida, R. P. P., Blua, M. J., Lopes, J. R. S., and Purcell, A. H. (2005). Vector transmission of *Xylella fastidiosa*: applying fundamental knowledge to generate disease management strategies. *Annals of the Entomological Society of America*, 98(6):775–786.

Almeida, R. P. P. and Nunney, L. (2015). How do plant diseases caused by *Xylella fastidiosa* emerge? *Plant Disease*, 99(11):1457–1467.

Almeida, R. P. P. and Purcell, A. H. (2003). Transmission of *Xylella fastidiosa* to grapevines by *Homalodisca coagulata* (Hemiptera: Cicadellidae). *Journal of Economic Entomology*, 96(2):264–271.

Almeida, R. P. P. and Purcell, A. H. (2006). Patterns of *Xylella fastidiosa* colonization on the precibarium of sharpshooter vectors relative to transmission to plants. *Annals of the Entomological Society of America*, 99(5):884–890.

Amaral, A. V. R., González, J. A., and Moraga, P. (2023). Spatio-temporal modeling of infectious diseases by integrating compartment and point process models. *Stochastic Environmental Research and Risk Assessment*, 37(4):1519–1533.

Anas, O., Harrison, U. J., and Brannen, P. M. (2008). The effect of warming winter temperatures on the severity of Pierce's disease in the Appalachian mountains and Piedmont of the southeastern United States. *Plant Health Progress*, 9(1):13.

Aniţa, S., Capasso, V., and Scacchi, S. (2021). Controlling the spatial spread of a *Xylella* epidemic. *Bulletin of Mathematical Biology*, 83:1–26.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1443.

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D., and Rue, H. (2019). Non-stationary Gaussian models with physical barriers. *Spatial Statistics*, 29:268 – 288.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, New York.

Belmonte, A., Gadaleta, G., and Castrignanò, A. (2023). Use of geostatistics for multi-scale spatial modeling of *Xylella fastidiosa* subsp. *pauca* (*Xfp*) infection with unmanned aerial vehicle image. *Remote Sensing*, 15(3):656.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–225.

Besag, J., York, J., and Mollie, A. (1991). Bayesian image-restoration, with 2 applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):21–22.

Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons, Chichester, UK.

Blua, M., Campbell, K., Morgan, D., and Redak, R. (2005). Impact of a screen barrier on dispersion behavior of *Homalodisca coagulata* (Hemiptera: Cicadellidae). *Journal of Economic Entomology*, 98(5):1664–1668.

Bodino, N., Cavalieri, V., Dongiovanni, C., Simonetto, A., Saladini, M. A., Plazio, E., Gilioli, G., Molinatto, G., Saponari, M., and Bosco, D. (2021). Dispersal of *Philaenus spumarius* (Hemiptera: Aphrophoridae), a vector of *Xylella fastidiosa*, in olive grove and meadow agroecosystems. *Environmental entomology*, 50(2):267–279.

Bolin, D. and Lindgren, F. (2011). Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *The Annals of Applied Statistics*, pages 523–550.

Bosso, L., Febbraro, M. D., Cristinzio, G., Zoina, A., and Russo, D. (2016a). Shedding light on the effects of climate change on the potential distribution of *Xylella fastidiosa* in the Mediterranean basin. *Biological Invasions*, 18:1759–1768.

Bosso, L., Russo, D., Di Febbraro, M., Cristinzio, G., and Zoina, A. (2016b). Potential distribution of *Xylella fastidiosa* in Italy: a maximum entropy model. *Phytopathologia Mediterranea*, 55(1):62–72.

Brotons, L., Thuiller, W., Araújo, M. B., and Hirzel, A. H. (2004). Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, 27(4):437–448.

Brown, N., Pérez-Sierra, A., Crow, P., and Parnell, S. (2020). The role of passive surveillance and citizen science in plant health. *CABI Agriculture and Bioscience*, 1:1–16.

Bull, C., Boer, S. D., Denny, T., Firrao, G., Saux, M. F.-L., Saddler, G., Scortichini, M., Stead, D., and Takikawa, Y. (2012). List of new names of plant pathogenic bacteria (2008-2010). *Journal of Plant Pathology*, 94(1):21–27.

Cambra, M. and Cambra, R. (1991). *Diseños de plantación y formación de árboles frutales (8ª ed.)*. Cuadernos (Estación Experimental de Aula Dei). CSIC - Estación Experimental de Aula Dei (EEAD). In Spanish.

Cameron, A. R. and Baldock, F. C. (1998). Two-stage sampling in surveys to substantiate freedom from disease. *Preventive Veterinary Medicine*, 34:19–30.

Camino, C., Calderon, R., Parnell, S., Dierkes, H., Chemin, Y., Roman-Ecija, M., Montes-Borrego, M., Landa, B. B., Navas-Cortes, J. A., Zarco-Tejada, P. J., and Beck, P. S. A. (2021). Detection of *Xylella fastidiosa* in almond orchards by synergic use of an epidemic spread model and remotely sensed plant traits. *Remote Sensing of Environment*, 260:112420.

Cannon, R. M. (2002). Demonstrating disease freedom - combining confidence levels. *Preventive Veterinary Medicine*, 52:227–249.

Capaldi, A., Behrend, S., Berman, B., Smith, J., Wright, J., and Lloyd, A. L. (2012). Parameter estimation and uncertainty quantification for an epidemic model. *Mathematical Biosciences and Engineering*, 9:553–576.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76:1–32.

Cavalieri, V., Altamura, G., Fumarola, G., di Carolo, M., Saponari, M., Cornara, D., Bosco, D., and Dongiovanni, C. (2019). Transmission of *Xylella fastidiosa* subspecies *pauca* sequence type 53 by different insect species. *Insects*, 10(10):324.

Cendoya, M., Hubel, A., Conesa, D., and Vicent, A. (2022a). Modeling the spatial distribution of *Xylella fastidiosa*: A nonstationary approach with dispersal barriers. *Phytopathology*, 112(5):1036–1045.

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

Cendoya, M., Navarro-Quiles, A., López-Quílez, A., Vicent, A., and Conesa, D. (2022b). An individual-based spatial epidemiological model for the spread of plant diseases. *Research Square*.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2021). *Shiny: web application framework for R*. R package version 1.7.1.

Chen, D., Moulin, B., and Wu, J. (2014). *Analyzing and modeling spatial and temporal dynamics of infectious diseases*. John Wiley & Sons, New Jersey.

Chowell, G., Hyman, J. M., Bettencourt, L. M. A., and Castillo-Chávez, C. (2009). *Mathematical and statistical estimation approaches in epidemiology*. Springer Dordrecht, New York.

Clark, J. S. and Gelfand, A. E. (2006). *Hierarchical modelling for the environmental sciences: statistical methods and applications*. OUP Oxford.

Coletta-Filho, H. D., Francisco, C. S., Lopes, J. R. S., De Oliveira, A. F., and Da Silva, L. F. D. O. (2016). First report of olive leaf scorch in Brazil, associated with *Xylella fastidiosa* subsp. *pauca*. *Phytopathologia Mediterranea*, 55(1):130–135.

Cornara, D., Bosco, D., and Fereres, A. (2018). *Philaenus spumarius*: when an old acquaintance becomes a new threat to European agriculture. *Journal of Pest Science*, 91(3):957–972.

Cornara, D., Morente, M., Markheiser, A., Bodino, N., Tsai, C. W., Fereres, A., Redak, R. A., Perring, T. M., and Spotti Lopes, J. R. (2019). An overview on the worldwide vectors of *Xylella fastidiosa*. *Entomologia Generalis*, 39(3-4):157–181.

Cornara, D., Saponari, M., Zeilinger, A. R., Stradis, A. D., Boscia, D., Giuliana Loconsole, Bosco, D., Martelli, G. P., Almeida, R. P. P., and Porcelli, F. (2017). Spittlebugs as vectors of *Xylella fastidiosa* in olive orchards in Italy. *Journal of Pest Science*, 90(2):521–530.

Cressie, N. A. C. (1993). *Statistics for spatial data (revised edition)*. John Wiley & Sons, Inc.

Cruaud, A., Gonzalez, A.-A., Godefroid, M., Nidelet, S., Streito, J.-C., Thuillier, J.-M., Rossi, J.-P., Santoni, S., and Rasplus, J.-Y. (2018). Using insects to detect, monitor and predict the distribution of *Xylella fastidiosa*: a case study in Corsica. *Scientific Reports*, 8(1):1–13.

Cunniffe, N. J., Koskella, B., Metcalf, C. J. E., Parnell, S., Gottwald, T. R., and Gilligan, C. A. (2015). Thirteen challenges in modelling plant diseases. *Epidemics*, 10:6–10.

Da Rocha, J., Zambolim, L., Maciel-Zambolim, E., and Ribeiro do Vale, F. (2010). Temporal and spatial dynamics of coffee leaf scorch caused by *Xylella fastidiosa*. *Australasian Plant Pathology*, 39:234–240.

Damian, D., Sampson, P. D., and Guttorp, P. (2001). Bayesian estimation of semi-parametric non-stationary spatial covariance structures. *Environmetrics: The official journal of the International Environmetrics Society*, 12(2):161–178.

Damian, D., Sampson, P. D., and Guttorp, P. (2003). Variance modeling for non-stationary spatial processes with temporal replications. *Journal of Geophysical Research: Atmospheres*, 108(D24).

Daugherty, M. P. and Almeida, R. P. P. (2009). Estimating *Xylella fastidiosa* transmission parameters: decoupling sharpshooter number and feeding period. *Entomologia Experimentalis et Applicata*, 132(1):84–92.

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., and Bodik, R. (2017). Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26:403–413.

Dean, C. B., Ugarte, M. D., and Militino, A. F. (2001). Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics*, 57:197–202.

DeAngelis, D. L. and Grimm, V. (2014). Individual-based models in ecology after four decades. *F1000prime reports*, 6.

Deardon, R., Brooks, S. P., Grenfell, B. T., Keeling, M. J., Tildesley, M. J., Savill, N. J., Shaw, D. J., and Woolhouse, M. E. (2010). Inference for individual-models of infectius diseases in large populations. *Statistica Sinica*, 20:239.

DG SANTE (2020). Latest developments of *Xylella fastidiosa* in the EU territory. `https://ec.europa.eu/food/plant/plant_health_biosecurity/legislation/emergency_measures/xylella-fastidiosa/latest-developments_en`. Accessed: 22 July 2020.

DG SANTE (2023). Latest developments of *Xylella fastidiosa* in the EU territory. https://food.ec.europa.eu/plants/plant-health-and-biosecurity/legislation/control-measures/xylella-fastidiosa/latest-developments-xylella-fastidiosa-eu-territory_en. Accessed: 10 August 2021.

Diekmann, O. and Heesterbeek, J. A. P. (2000). *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, volume 5. John Wiley & Sons.

Diggle, P., Tawn, J., and Moyeed, R. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350.

Diggle, P. J., Ribeiro, P. J., and Christensen, O. F. (2003). *An introduction to model-based geostatistics*. Springer, New York.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., Mcclean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., and Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1):27–046.

Elith, J. and Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):677–697.

European and Mediterranean Plant Protection Organization (EPPO) (2019a). Eppo global database: *Xylella fastidiosa (XYLEFA)*. https://gd.eppo.int/taxon/XYLEFA/distribution. Accessed: 20 April 2023.

European and Mediterranean Plant Protection Organization (EPPO) (2019b). PM 7/24 (4) *Xylella fastidiosa*. *EPPO Bulletin*, 49(2):175–227.

European Food Safety Authority (EFSA) (2015). Scientific opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory, with the identification and evaluation of risk reduction options. *EFSA Journal*, 13(1):3989.

European Food Safety Authority (EFSA) (2019). Update of the scientific opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory. *EFSA Journal*, 17(5).

European Food Safety Authority (EFSA) (2020). Update of the *Xylella* spp. host plant database - systematic literature search up to 30 june 2019. *EFSA Journal*, 18(4).

European Food Safety Authority (EFSA) (2023). Toolkit for plant pest surveillance in the EU. https://efsa.onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)1831-4732.toolkit-plant-pest-surveillance. Accessed: 17 April 2023.

European Food Safety Authority (EFSA), Gibin, D., Pasinato, L., and Delbianco, A. (2023). Update of the *Xylella* spp. host plant database – systematic literature search up to 31 December 2022. *EFSA Journal*, 21(6):e08061.

European Food Safety Authority (EFSA), Lázaro, E., Parnell, S., Civera, A. V., Schans, J., Schenk, M., Abrahantes, J. C., Zancanaro, G., and Vos, S. (2020a). General guidelines for statistically sound and risk-based surveys of plant pests. *EFSA Supporting Publications*, 17(9):1919E.

European Food Safety Authority (EFSA), Lázaro, E., Parnell, S., Civera, A. V., Schans, J., Schenk, M., Abrahantes, J. C., Zancanaro, G., and Vos, S. (2020b). Guidelines for statistically sound and risk-based surveys of *Agrilus planipennis*. *EFSA Supporting Publications*, 17(12):1983E.

European Food Safety Authority (EFSA), Lázaro, E., Parnell, S., Vicent Civera, A., Schans, J., Schenk, M., Schrader, G., Abrahantes, J. C., Zancanaro, G., and Vos, S. (2020c). Guidelines for statistically sound and risk-based surveys of *Phyllosticta citricarpa*. *EFSA Supporting Publications*, 17(7):1893E.

European Food Safety Authority (EFSA), Lázaro, E., Parnell, S., Vicent Civera, A., Schans, J., Schenk, M., Schrader, G., Cortiñas Abrahantes, J., Zancanaro, G., and Vos, S. (2020d). Guidelines for statistically sound and risk-based surveys of *Xylella fastidiosa*. *EFSA Supporting Publications*, 17(6):1873E.

European Union (EU) (2016). 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) No 228/2013, (EU) No 652/2014 and (EU) No 1143/2014 of the European Parliament and of the Council 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.

European Union (EU) (2019). 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.

European Union (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.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, methods and applications*. Springer Science & Business Media.

Feil, H. and Purcell, A. H. (2001). Temperature-dependent growth and survival of *Xylella fastidiosa* in vitro and in potted grapevines. *Plant Disease*, 85(12):1230–1234.

Fernández, M. G. R. (2020). Análisis de distribución potencial de *Xylella fastidiosa* subsp. *multiplex* ST-6 y *Philaenus spumarius* en el sur de la Península Ibérica mediante el modelo ecológico de nicho MaxEnt. *GeoFocus. International Review of Geographical Information Science and Technology*, (25):77–102. In Spanish.

Ferreira, G., de A. Mascaro, F., Pria, M. D., Ribeiro, P., and De Mio, L. M. (2016). Spatial analysis of plum leaf scald in São Paulo State, Brazil. *Journal of Plant Pathology*, pages 511–518.

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315.

Food and Agriculture Organization of the United Nations (FAO) (2014). Risk-based disease surveillance—A manual for veterinarians on the design and analysis of surveillance for demonstration of freedom from disease. *FAO Animal Production and Health Manual No 17*.

Food and Agriculture Organization of the United Nations (FAO) (2019). New standards to curb the global spread of plant pests and diseases. `https://www.fao.org/news/story/en/item/1187738/icode/`. Accessed: 11 April 2023.

Food and Agriculture Organization of the United Nations (FAO) (2021). Climate change fans spread of pests and threatens plants and crops, new FAO study. `https://www.fao.org/news/story/en/item/1402920/icode/`. Accessed: 24 July 2023.

Food and Agriculture Organization of the United Nations (FAO), IFAD, UNICEF, WFP, and WHO (2019). *The State of food Security nad Nutrition in the World, 2019. Safeguarding against economic slowdowns and downturns. Rome.* FAO.

Franklin, J. (2010). *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, New York.

Fuglstad, G. A., Simpson, D., Lindgren, F., and Rue, H. (2019). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452.

Gard, T. (1988). *Introduction to stochastic differential equations*. Marcel Dekker, New York.

Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.

Generalitat Valenciana (GVA) (2019). Situación de *Xylella fastidiosa* en la Comunitat Valenciana, marzo 2019. `http://agroambient.gva.es/documents/163214705/163847802/20190326+Situacion+de+X+fastidiosa+en+C+Valenciana+web.pdf/ab1a7a09-3d26-4b61-9b7a-ab30d72105cf`. Accessed: 18 February 2021.

Generalitat Valenciana (GVA) (2020). Plan de acción frente a *Xylella fastidiosa* en la Comunitat Valenciana. http://agroambient.gva.es/documents/163214705/163847802/Plan+accion+Diciembre+2020_firmado.pdf/856ce97f-f733-4ea8-8095-88149f6e78e7. Accessed: 18 February 2021.

Generalitat Valenciana (GVA) (2021). Plan de acción frente a *Xylella fastidiosa* en la Comunitat Valenciana. https://portalagrari.gva.es/documents/366567370/368987753/Plan+acci%C3%B3n+Xylella_enero+2023+B18_firmado.pdf/f94e83d1-7341-0451-73bc-c5adf1c360d0?t=1674027938932. Accessed: 10 August 2023.

Generalitat Valenciana (GVA) (2023). Situación de *Xylella fastidiosa* en la Comunitat Valenciana. Servicio de Sanidad Vegetal, DG Agricultura, Ganadería y Pesca, Generalitat Valenciana. https://portalagrari.gva.es/documents/366567370/368358522/20230110+Situacion+de+Xf+en+la+C+Valenciana.pdf/24dfb27c-18b7-13b5-8732-b5dd092f97d5?t=1673437548566. Accessed: 1 May 2023.

Giampetruzzi, A., Morelli, M., Saponari, M., Loconsole, G., Chiumenti, M., Boscia, D., Savino, V. N., Martelli, G. P., and Saldarelli, P. (2016). Transcriptome profiling of two olive cultivars in response to infection by the CoDiRO strain of *Xylella fastidiosa* subsp. *pauca*. *BMC Genomics*, 17(1):1–18.

Gibson, G., Kleczkowski, A., and Gilligan, C. (2004). Bayesian analysis of botanical epidemics using stochastic compartmental models. *Proceedings of the national academy of sciences*, 101(33):12120–12124.

Gibson, G. J. (1997). Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46:215–233.

Gilioli, G., Simonetto, A., Colturato, M., Bazarra, N., Fernández, J. R., Naso, M. G., Donato, B., Bosco, D., Dongiovanni, C., Maiorano, A., Mosbach-Schulz, O., Cortés, J. A. N., and Saponari, M. (2023). An eco-epidemiological model supporting rational disease management of *Xylella fastidiosa*. An application to the outbreak in Apulia (Italy). *Ecological Modelling*, 476:110226.

Gilligan, C. A. (2002). An epidemiological framework for disease management. *Advances in Botanical Research*, 38:1–64.

Gilligan, C. A. (2008). Sustainable agriculture and plant diseases: an epidemiological perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363:741.

Giménez-Romero, M. A., Moralejo, D. E., and Matías, D. M. (2023). A compartmental model for *Xylella fastidiosa* diseases with explicit vector seasonal dynamics. *Phytopathology*.

Godefroid, M., Cruaud, A., Streito, J. C., Rasplus, J. Y., and Rossi, J. P. (2019). *Xylella fastidiosa*: climate suitability of European continent. *Scientific Reports*, 9(1):8844.

Godefroid, M., Cruaud, A., Streito, J. C., Rasplus, J. Y., and Rossi, J. P. (2022). Forecasting future range shifts of *Xylella fastidiosa* under climate change. *Plant Pathology*, 71:1839–1848.

Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. Chapman and Hall/CRC, Boca Raton.

Goodwin, P. and Zhang, S. (1997). Distribution of *Xylella fastidiosa* in southern Ontario as determined by the polymerase chain reaction. *Canadian Journal of Plant Pathology*, 19(1):13–18.

Graham, M. H. (2003). Confronting multicollinearity in ecological multiple regression. *Ecology*, 84(11):2809–2815.

Groves, R., Chen, J., Civerolo, E., Freeman, M., and Viveros, M. (2005). Spatial analysis of almond leaf scorch disease in the San Joaquin Valley of California: factors affecting pathogen distribution and spread. *Plant disease*, 89(6):581–589.

Guisan, A. and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135:147–186.

Guttorp, P. and Gneiting, T. (2006). Studies in the history of probability and statistics XLIX on the Matérn correlation family. *Biometrika*, 93:989–995.

Haelterman, R., Tolocka, P., Roca, M., Guzmán, F., Fernández, F., and Otero, M. (2015). First presumptive diagnosis of *Xylella fastidiosa* causing olive scorch in Argentina. *Journal of Plant Pathology*, 97(2):393.

Handcock, M. S. and Wallis, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields. *Journal of the American Statistical Association*, 89:368–378.

Heinze, G., Wallisch, C., and Dunkler, D. (2018). Variable selection - A review and recommendations for the practicing statistician. *Biometrical Journal*, 60(3):431–449.

Hernández, O. G. and García, L. V. (2019). La dimensión geográfica de las invasiones biológicas en el Antropoceno: el caso de *Xylella fastidiosa*. *Boletín de la Asociación de Geógrafos Españoles*, (80):1–32. In Spanish.

Hernández, O. G. and García, L. V. (2018). Incidencia de *Xylella fastidiosa* en las Islas Baleares y distribución potencial en la península ibérica. *Investigaciones Geográficas*, pages 55–72. In Spanish.

Hijmans, R. J. and van Etten, J. (2012). raster: Geographic analysis and modeling with raster data. R package version 2.0-12. `http://CRAN.R-project.org/package=raster`.

Hyatt-Twynam, S. R., Parnell, S., Stutt, R. O. J. H., Gottwald, T. R., Gilligan, C. A., and Cunniffe, N. J. (2017). Risk-based management of invading plant disease. *New Phytologist*, 214(3):1317–1329.

Janse, J. D. and Obradovic, A. (2010). *Xylella fastidiosa*: its biology, diagnosis, control and risks. *Journal of Plant Pathology*, 92:S35–S48.

Jeger, M. et al. (2000). Theory and plant epidemiology. *Plant Pathology*, 49(6):651–658.

Jolliffe, I. T. (2002). *Principal component analysis*. Springer, 2 edition.

Keeling, M. J. and Rohani, P. (2008). *Modeling infectious diseases in humans and animals*. Princeton University Press, Princeton.

Keeling, M. J., Woolhouse, M. E., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J., and Grenfell, B. T. (2001). Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294:813–817.

Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 115(772):700–721.

Kirkeby, C., Brookes, V. J., Ward, M. P., Dürr, S., and Halasa, T. (2021). A practical introduction to mechanistic modeling of disease transmission in veterinary science. *Frontiers in Veterinary Science*, 7:546651.

Kleczkowski, A., Hoyle, A., and McMenemy, P. (2019). One model to rule them all? Modelling approaches across OneHealth for human, animal and plant epidemics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1775).

Kottelenberg, D., Hemerik, L., Saponari, M., and van der Werf, W. (2021). Shape and rate of movement of the invasion front of *Xylella fastidiosa* spp. *pauca* in Puglia. *Scientific Reports*, 11:1061.

Krainski, E., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgren, F., and Rue, H. (2019). *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Chapman and Hall/CRC, Boca Raton.

Krivanek, A., Riaz, S., and Walker, M. (2006). Identification and molecular mapping of PdR1, a primary resistance gene to Pierce's disease in *Vitis*. *Theoretical and Applied Genetics*, 112(6):1125–1131.

Lago, C., Morente, M., De las Heras-Bravo, D., Martí-Campoy, A., Rodríguez-Ballester, F., Plaza, M., Moreno, A., and Fereres, A. (2021). Dispersal of *Neophilaenus campestris*, a vector of *Xylella fastidiosa*, from olive groves to oversummering hosts. *Journal of Applied Entomology*.

Lam, S. K., Pitrou, A., and Seibert, S. (2015). Numba: A LLVM-based Python JIT Compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC - LLVM '15*, pages 1–6, New York, USA. ACM Press.

Landa, B. B., Castillo, A. I., Giampetruzzi, A., Kahn, A., Román-Écija, M., Velasco-Amo, M. P., Navas-Cortés, J. A., Marco-Noales, E., Barbé, S., Moralejo, E., Coletta-Filho, H. D., Saldarelli, P., Saponari, M., and Almeida, R. P. P. (2020). Emergence of a plant pathogen in Europe associated with multiple intercontinental introductions. *Applied and Environmental Microbiology*, 86(3):e01521–19.

Laranjeira, F. F., Silva, L. G., Fonseca, E. L., Silva, S. X., Rocha, J. B., Santos-Filho, H. P., Ledo, C. A. S., and Hau, B. (2008). Prevalence, incidence and distribution of citrus variegated chlorosis in Bahia, Brazil. *Tropical Plant Pathology*, 33:339–347.

Latimer, A. M., Wu, S., Gelfand, A. E., and Silander Jr., J. A. (2006). Building statistical models to analyze species distributions. *Ecological Applications*, 16(1):33–50.

Lázaro, E., Sesé, M., López-Quílez, A., Conesa, D., Dalmau, V., Ferrer, A., and Vicent, A. (2021). Tracking the outbreak: an optimized sequential adaptive strategy for xylella fastidiosa delimiting surveys. *Biological Invasions*, 23(10):3243–3261.

Leroux, B. G., Lei, X., and Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 179–191. Springer, Springer, New York, NY.

Li, B., Sanderlin, R., Melanson, R. A., and Yu, Q. (2011). Spatio-temporal analysis of a plant disease in a non-uniform crop: A Monte Carlo approach. *Journal of Applied Statistics*, 38(1):175–182.

Lieth, J. H., Meyer, M. M., Yeo, K. H., and Kirkpatrick, B. C. (2011). Modeling cold curing of Pierce's disease in *Vitis vinifera* 'Pinot Noir' and 'Cabernet Sauvignon' grapevines in California. *Phytopathology*, 101(12):1492–1500.

Lindgren, F., Bolin, D., and Rue, H. (2022). The spde approach for gaussian and non-gaussian fields: 10 years and still running. *Spatial Statistics*, 50:100599.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337.

Maanen, A. V. and Xu, X. M. (2003). Modelling plant disease epidemics. *European Journal of Plant Pathology*, 109:669–682.

Madden, L. V. and Hughes, G. (1995). Plant disease incidence: distributions, heterogeneity, and temporal analysis. *Phytopathology*, 33:529–564.

Madden, L. V., Hughes, G., and van den Bosch, F. (2007). *The study of plant disease epidemics*. APS Press, St. Paul, Minnesota.

Madden, L. V., Louie, R. A. J. J., and Knoke, J. K. (1982). Evaluation of tests for randomness of infected plants. *Phytopathology*, 72:195–198.

Mahmood, M., Mateu, J., and Hernández-Orallo, E. (2021). Contextual contact tracing based on stochastic compartment modeling and spatial risk assessment. *Stochastic Environmental Research and Risk Assessment*, pages 1–25.

Maloy, O. C. (1993). *Plant disease control: principles and practice.* John Wiley and Sons, Inc., New York.

Martelli, G. P. (2016). The current status of the quick decline syndrome of olive in southern Italy. *Phytoparasitica*, 44:1–10.

Martelli, G. P., Boscia, D., Porcelli, F., and Saponari, M. (2016). The olive quick decline syndrome in south-east Italy: a threatening phytosanitary emergency. *European Journal of Plant Pathology*, 144(2):235–243.

Martin, O., Fernandez-Diclo, Y., Coville, J., and Soubeyrand, S. (2021). Equilibrium and sensitivity analysis of a spatio-temporal host-vector epidemic model. *Nonlinear Analysis: Real World Applications*, 57:103194.

Martin, S. W., Meek, A. H., and Willeberg, P. (1987). Sampling methods. In *Veterinary Epidemiology: Principles and Methods*, chapter 2, pages 22–47. Iowa State University Press.

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

Martínez-Minaya, J., Cameletti, M., Conesa, D., and Pennino, M. G. (2018a). Species distribution modeling: a statistical review with focus in spatio-temporal issues. *Stochastic Environmental Research and Risk Assessment*, 32(11):3227–3244.

Martínez-Minaya, J., Conesa, D., López-quílez, A., and Vicent, A. (2018b). Spatial and climatic factors associated with citrus black spot. A Bayesian analysis of disease spread in South Africa. *European Journal of Plant Pathology*, 151(4):991–1007.

Martínez-Minaya, J., Conesa, D., Bakka, H., and Pennino, M. G. (2019). Dealing with physical barriers in bottlenose dolphin (*Tursiops truncatus*) distribution. *Ecological Modelling*, 406:44–49.

Mastin, A. J., Gottwald, T. R., van den Bosch, F., Cunniffe, N. J., and Parnell, S. (2020). Optimising risk-based surveillance for early detection of invasive plant pathogens. *PLOS Biology*, 18:e3000863.

Matis, J. and Hartley, H. (1971). Stochastic compartmental analysis: model and least squares estimation from time series data. *Biometrics*, 27:9160793.

Matérn, B. (1986). *Spatial variation*. Springer, Berlin, Germany, 2nd edition.

Meentemeyer, R. K., Anacker, B. L., Mark, W., and Rizzo, D. M. (2008). Early detection of emerging forest disease using dispersal estimation and ecological niche modeling. *Ecological Applications*, 18(2):377–390.

Meentemeyer, R. K., Cunniffe, N. J., Cook, A. R., Filipe, J. A. N., Hunter, R. D., Rizzo, D. M., Gilligan, C. A., Cunniffe, N. J., Cook, A. R., Filipe, J. A. N., Hunter, R. D., Rizzo, D. M., and Gilligan, C. A. (2011). Epidemiological modeling of invasion in heterogeneous landscapes: Spread of sudden oak death in California (1990-2030). *Ecosphere 2(2): Article 17*, 2:1–24.

Ministerio de Agricultura, Pesca y Alimentación (MAPA) (2021a). Anexo I del Reglamento de Ejecución (UE) 2019/1715 de la Comisión, vigor 14/12/2019). `http://www.caib.es/sites/xf/ca/ultima_notificacion/`. Accessed: 1 April 2021.

Ministerio de Agricultura, Pesca y Alimentación (MAPA) (2021b). Resumen de los datos contenidos en el SIGPAC para la campaña 2021.

Mircetich, S. M., Lowe, S., Moller, W., and Nyland, G. (1976). Etiology of almond leaf scorch disease and transmission of the causal agent. *Phytopathology*, 66(1):17–24.

Montanari, F. and Traon, D. (2017). Modernising eu policy against phytosanitary risks-the new eu plant health law. *European Food and Feed Law Review*, 12:131.

Montgomery, D. C. and Runger, G. C. (2010). *Applied statistics and probability for engineers*. John wiley & sons.

Moralejo, E., Gomila, M., Montesinos, M., Borràs, D., Pascual, A., Nieto, A., Adrover, F., Gost, P. A., Seguí, G., Busquets, A., Jurado-Rivera, J. A., Quetglas, B., García, J. d. D., Beidas, O., Juan, A., Velasco-Amo, M. P., Landa, B. B., and Olmo, D. (2020). Phylogenetic inference enables reconstruction of a long-overlooked outbreak of almond leaf scorch disease (*Xylella fastidiosa*) in Europe. *Communications Biology*, 3(1).

Nathan, R., Klein, E., Robledo-Arnuncio, J. J., and Revilla, E. (2012). Dispersal kernels: review. In *Dispersal Ecology and Evolution*. Oxford University Press.

Nunney, L., Ortiz, B., Russell, S. A., Ruiz Sánchez, R., and Stouthamer, R. (2014a). The complex biogeography of the plant pathogen *Xylella fastidiosa*: genetic evidence of introductions and subspecific introgression in Central America. *PLoS ONE*, 9(11):e112463.

Nunney, L., Schuenzel, E. L., Scally, M., Bromley, R. E., and Stouthamer, R. (2014b). Large-scale intersubspecific recombination in the plant-pathogenic bacterium *Xylella fastidiosa* is associated with the host shift to mulberry. *Applied and Environmental Microbiology*, 80(10):3025–3033.

Ojiambo, P., Yuen, J., Van den Bosch, F., and Madden, L. (2017). Epidemiology: past, present, and future impacts on understanding disease dynamics and improving plant disease management — a summary of focus issue articles. *Phytopathology*, 107(10):1092–1094.

Oksendal, B. (2007). *Stochastic differential equations: An introduction with applications*. Springer, Berlin.

Park, Y.-L., Perring, T. M., Krell, R. K., Hashim-Buckey, J. M., and Hill, B. L. (2011). Spatial distribution of Pierce's disease related to incidence, vineyard characteristics, and surrounding land uses. *American journal of enology and viticulture*, 62(2):229–238.

Parnell, S., Bosch, F. V. D., Gottwald, T., and Gilligan, C. A. (2017). Surveillance to inform control of emerging plant diseases: An epidemiological perspective. *Annual review of phytopathology*, 55:591–610.

Parnell, S., Gottwald, T., Gilligan, C., Cunniffe, N., and Van Den Bosch, F. (2010). The effect of landscape pattern on the optimal eradication zone of an invading epidemic. *Phytopathology*, 100(7):638–644.

Parnell, S., Gottwald, T., Riley, T., and Van Den Bosch, F. (2014). A generic risk-based surveying method for invading plant pathogens. *Ecological Applications*, 24(4):779–790.

Pettit, L. I. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1):175–184.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, 124.

Pou, M. M. (2004). *El almendro. Manual técnico.* Mundi-Prensa Libros. In Spanish.

Purcell, A. (1997). *Xylella fastidiosa*, a regional problem or global threat? *Journal of Plant Pathology*, 79:99–105.

Purcell, A. H. (1980). Environmental therapy for Pierce's disease of grapevines. *Plant Disease*, 64(4):388.

Purcell, A. H. (1981). Vector preference and inoculation efficiency as components of resistance to Pierce's disease in European grape cultivars. *Phytopathology*, 71(4):429.

Python Software Foundation (2021). The Python Language Reference - Python 3.8.12 documentation. Accessed: 12 April 2022.

R Core Team (2022). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rallo, L. and Cuevas, J. (2017). Fructificación y producción. In Barranco, D., Fernández Escobar, R., and Rallo, L., editors, *El cultivo del olivo*, chapter 5, pages 145–186. Mundi-Prensa Libros, Madrid, España, 7 edition. In Spanish.

Randall, J. J., Goldberg, N. P., Kemp, J. D., Radionenko, M., French, J. M., Olsen, M. W., and Hanson, S. F. (2009). Genetic analysis of a novel *Xylella fastidiosa* subspecies found in the southwestern United States. *Applied and Environmental Microbiology*, 75(17):5631–5638.

Riebler, A., Sørbye, S. H., Simpson, D., and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–1165.

Ristaino, J. B., Anderson, P. K., Bebber, D. P., Brauman, K. A., Cunniffe, N. J., Fedoroff, N. V., Finegold, C., Garrett, K. A., Gilligan, C. A., Jones, C. M., et al. (2021). The persistent threat of emerging plant disease pandemics to global food security. *Proceedings of the National Academy of Sciences*, 118(23):e2022239118.

Ristaino, J. B. and Gumpertz, M. L. (2003). New frontiers in the study of dispersal and spatial analysis of epidemics caused by species in the genus *Phytophthora*. *https://doi.org/10.1146/annurev.phyto.38.1.541*, 38:541–576.

Roos, M. and Held, L. (2011). Sensitivity analysis in Bayesian generalized linear mixed models for binary data. *Bayesian Analysis*, 6(2):259–278.

Rosace, M. C., Cendoya, M., Mattion, G., Vicent, A., Battisti, A., Cavaletto, G., Marini, L., and Rossi, V. (2023). A spatio-temporal database of plant pests' first introductions across the EU and potential entry pathways. *Scientific Data*, 10(731).

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and Its Application*, 4:395–421.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119.

Saponari, M., Boscia, D., Nigro, F., and Martelli, G. (2013). Identification of DNA sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (southern Italy). *Journal of Plant Pathology*, 95(3).

Saponari, M., Giampetruzzi, A., Loconsole, G., Boscia, D., and Saldarelli, P. (2018). *Xylella fastidiosa* in olive in Apulia: where we stand. *Phytopathology*, 109(2):175–186.

Saponari, M., Loconsole, G., Cornara, D., Yokomi, R. K., De Stradis, A., Boscia, D., Bosco, D., Martelli, G. P., Krugner, R., and Porcelli, F. (2014). Infectivity and transmission of *Xylella fastidiosa* by *Philaenus spumarius* (Hemiptera: Aphrophoridae) in Apulia, Italy. *Journal of Economic Entomology*, 107(4):1316–9.

Savary, S., Willocquet, L., Pethybridge, S. J., Esker, P., McRoberts, N., and Nelson, A. (2019). The global burden of pathogens and pests on major food crops. *Nature ecology & evolution*, 3(3):430–439.

Schaad, N. W., Postnikova, E., Lacy, G., Fatmi, M., and Chang, C.-J. (2004). *Xylella fastidiosa* subspecies: *X. fastidiosa* subsp *piercei*, subsp. nov., *X. fastidiosa* subsp. *multiplex* subsp. nov., and *X. fastidiosa* subsp. *pauca* subsp. nov. *Systematic and Applied Microbiology*, 27(3):290–300.

Schmidt, A. M., Guttorp, P., and O'Hagan, A. (2011). Considering covariates in the covariance structure of spatial processes. *Environmetrics*, 22(4):487–500.

Schmidt, A. M. and O'Hagan, A. (2003). Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(3):743–758.

Segura, J. M. A., Cabetas, M. J. R., et al. (2018). El almendro: diversidad de marcos de plantación. *Agricultura: Revista agropecuaria y ganadera*, (1014):100–105. In Spanish.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical science*, 32(1):1–28.

Sisterson, M. S., Ledbetter, C. A., Chen, J., Higbee, B. S., Groves, R. L., and Daane, K. M. (2012). Management of almond leaf scorch disease: long-term data on yield, tree vitality, and disease progress. *Plant Disease*, 96(7):1037–1044.

Soong, T. (1973). *Random differential equations in science and engineering*. Academic Press, New York.

Soubeyrand, S., de Jerphanion, P., Martin, O., Saussac, M., Manceau, C., Hendrikx, P., and Lannou, C. (2018). Inferring pathogen dynamics from temporal count data: the emergence of *Xylella fastidiosa* in France is probably not recent. *New Phytologist*, 219(2):824–836.

Stein, M. L. (1999). *Interpolation of spatial data*. Springer, New York.

Strand, J. (1970). Random ordinary differential equations. *Journal of Differential Equations*, 7(3):538–553.

Strona, G., Carstens, C. J., and Beck, P. S. (2017). Network analysis reveals why *Xylella fastidiosa* will persist in Europe. *Scientific Reports*, 7(1):71.

Strona, G., Castellano, C., Fattorini, S., Ponti, L., Gutierrez, A. P., and Beck, P. S. (2020). Small world in the real world: long distance dispersal governs epidemic dynamics in agricultural landscapes. *Epidemics*, 30:100384.

Su, C. C., Deng, W. L., Jan, F. J., Chang, C. J., Huang, H., Shih, H. T., and Chen, J. (2016). *Xylella taiwanensis* sp. nov., causing pear leaf scorch disease. *International Journal of Systematic and Evolutionary Microbiology*, 66(11):4766–4771.

Tubajika, K., Civerolo, E., Ciomperlik, M., Luvisi, D., and Hashim, J. (2004). Analysis of the spatial patterns of Pierce's disease incidence in the lower San Joaquin Valley in California. *Phytopathology*, 94(10):1136–1144.

Umlauf, N., Adler, D., Kneib, T., Lang, S., and Zeileis, A. (2015). Structured additive regression models: An R interface to BayesX. *Journal of Statistical Software*, 63:1–46.

Václavík, T., Kanaskie, A., Hansen, E. M., Ohmann, J. L., and Meentemeyer, R. K. (2010). Predicting potential and actual distribution of sudden oak death in Oregon: prioritizing landscape contexts for early detection and eradication of disease outbreaks. *Forest Ecology and Management*, 260(6):1026–1035.

Van der Plank, J. E. (1963). *Plant diseases: epidemics and control*. Academic Press, New York.

Wagner, B. and Salman, M. D. (2004). Strategies for two-stage sampling designs for estimating herd-level prevalence. *Preventive Veterinary Medicine*, 66:1–17.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.

Wells, J. M., Raju, B. C., Hung, H.-Y., Weisburg, W. G., Mandelco-Paul, L., and Brenner, D. J. (1987). *Xylella fastidiosa* gen. nov., sp. nov: gram-negative, xylem-limited, fastidious plant bacteria related to *Xanthomonas* spp. *International Journal of Systematic and Evolutionary Microbiology*, 37(2):136–143.

White, S. M., Bullock, J. M., Hooftman, D. A., and Chapman, D. S. (2017). Modelling the spread and control of *Xylella fastidiosa* in the early stages of invasion in Apulia, Italy. *Biological Invasions*, 19(6):1825–1837.

White, S. M., Navas-Cortés, J. A., Bullock, J. M., Boscia, D., and Chapman, D. S. (2020). Estimating the epidemiology of emerging *Xylella fastidiosa* outbreaks in olives. *Plant Pathology*, 69(8):1403–1413.

Willem, L., Verelst, F., Bilcke, J., Hens, N., and Beutels, P. (2017). Lessons from a decade of individual-based models for infectious disease transmission: A systematic review (2006-2015). *BMC Infectious Diseases*, 17:1–16.

Yoon, S. and Lee, W.-H. (2023). Spatial analysis of climatic and dispersion characteristics of xylella fastidiosa outbreak by insect vectors. *Journal of Asia-Pacific Entomology*, 26(1):102011.

Zhang, B. and DeAngelis, D. L. (2020). An overview of agent-based models in plant biology and ecology. *Annals of Botany*, 126(4):539–557.

Appendix A

Supplementary figures and tables for Chapter 3

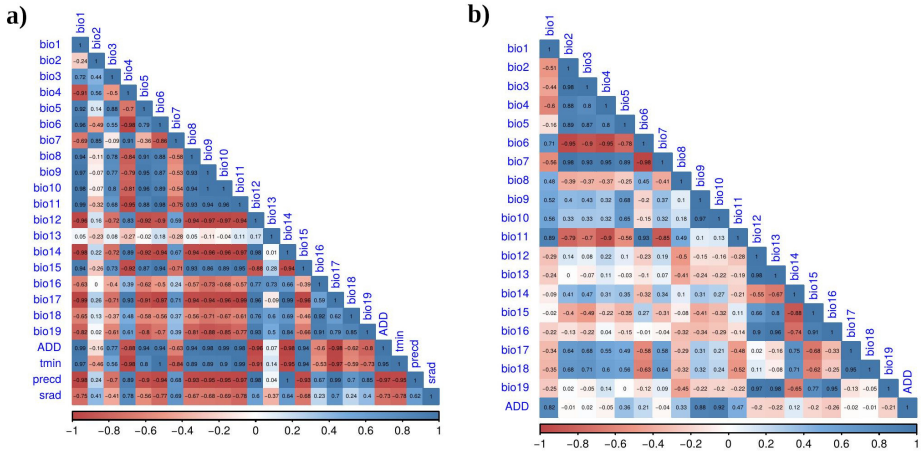

Figure A.1: Correlation matrix. (**a**) Pearson correlation coefficient for bioclimatic variables (*bio*), accumulated degree-days over 15 ℃ from February to October (*ADD*), average minimum temperature in winter (*tmin*), average precipitation during the dry season (*precd*) and solar radiation (*srad*) in Alicante, Spain; (**b**) Pearson correlation coefficient for bioclimatic variables (*bio*) and accumulated degree-days over 15 ℃ from April to October (*ADD*) in Lecce, Italy.

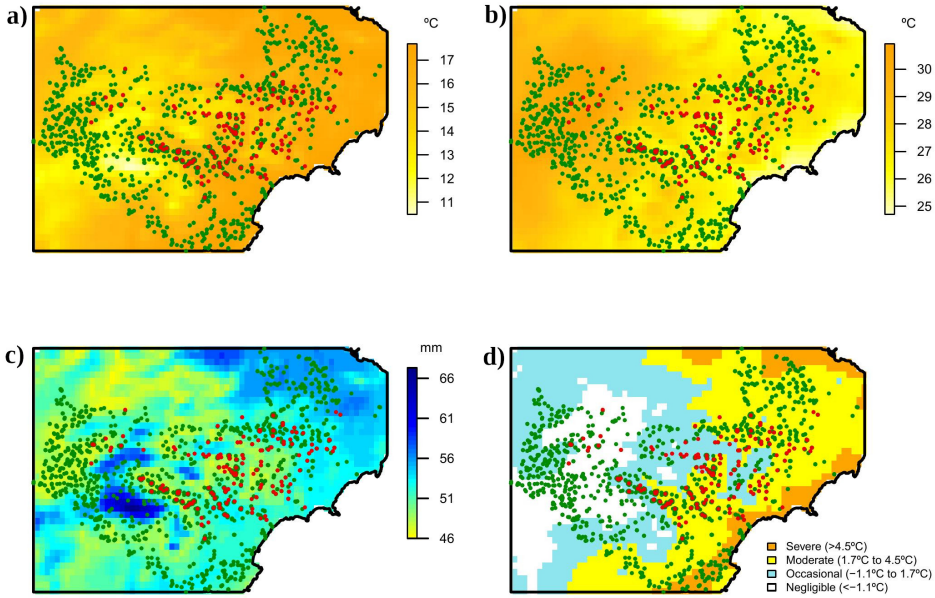

FIGURE A.2: Geographical distribution of (•) presence and (•) absence of *Xylella fastidiosa* over climatic covariates in the demarcated area in Alicante, Spain; (**a**) annual mean temperature (*bio1*), (**b**) temperature annual range (*bio7*), (**c**) precipitation of the wettest month (*bio13*), and (**d**) Purcell's categories based on minimum winter temperature (Anas et al., 2008).

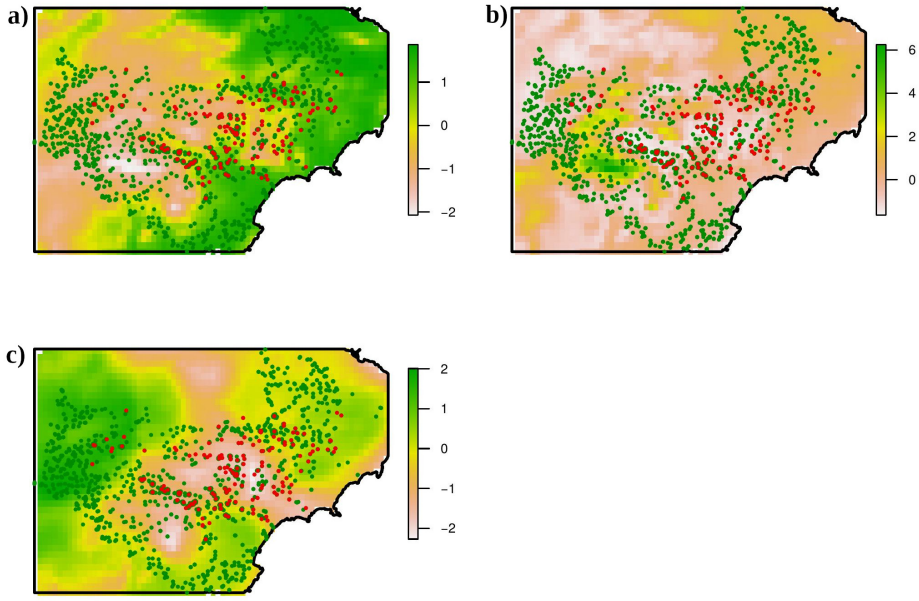

FIGURE A.3: Geographical distribution of (•) presence and (•) absence of *Xylella fastidiosa* in the demarcated area in Alicante, Spain, over the principal components, (**a**) PC1, (**b**) PC2 and (**c**) PC3.

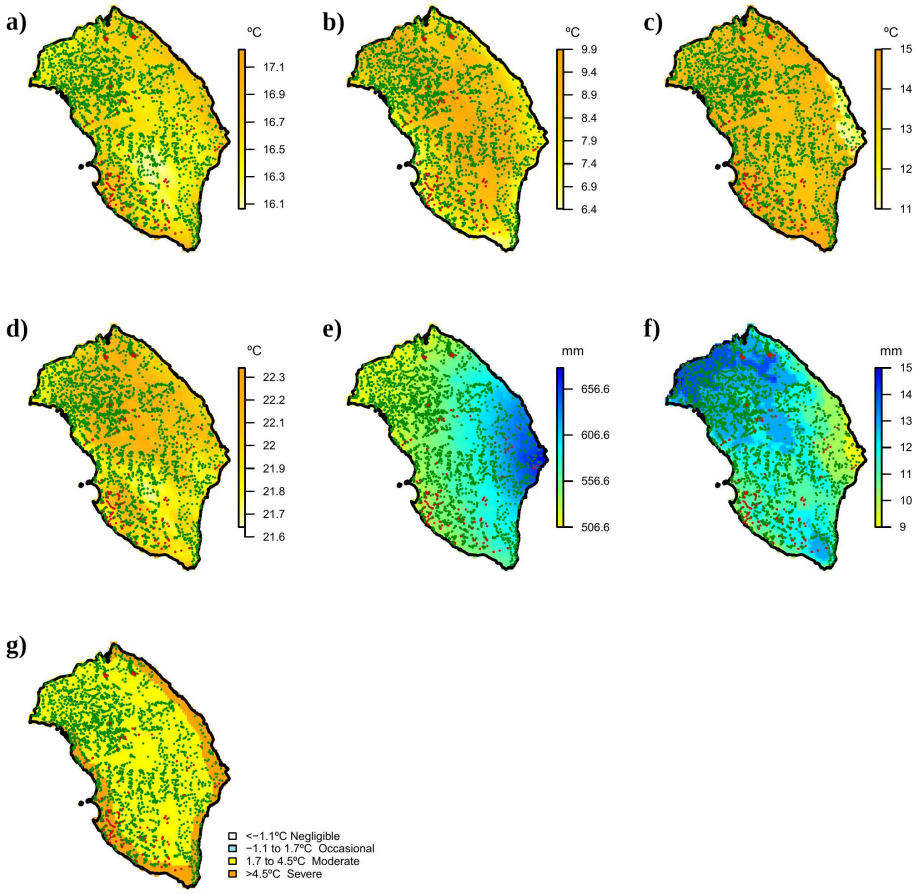

FIGURE A.4: Geographical distribution of (•) presence and (•) absence of *Xylella fastidiosa* over climatic covariates in Lecce, Italy (**a**) annual mean temperature *bio1*, (**b**) mean diurnal range *bio2*, (**c**) mean temperature of the wettest quarter *bio8*, (**d**) mean temperature of the driest quarter *bio9*, (**e**) annual precipitation *bio12*, (**f**) precipitation of the driest month *bio14*, and (**g**) Purcell's categories based on minimum winter temperature (Anas et al., 2008).

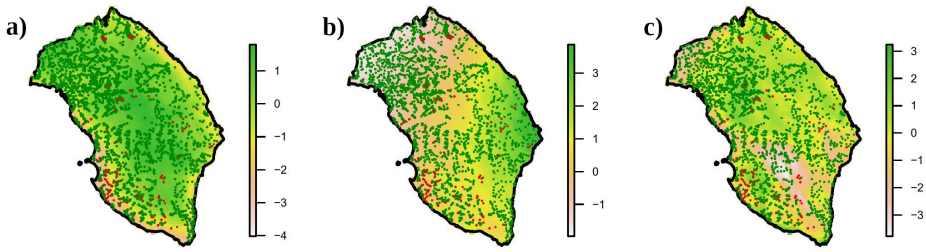

FIGURE A.5: Geographical distribution of (●) presence and (●) absence of *Xylella fastidiosa* in Lecce, Italy, over the principal components (**a**) PC1, (**b**) PC2, and (**c**) PC3.

TABLE A.1: Best models for the distribution of *Xylella fastidiosa* in the demarcated area in Alicante, Spain.

	Model[a]	**WAIC**[b]	**LCPO**[c]
Models with climatic covariates			
1.	$\beta_0 + \boldsymbol{u}$	617.627	1.638
2.	$\beta_0 + bio7 + \boldsymbol{u}$	617.922	1.650
3.	$\beta_0 + bio1 + \boldsymbol{u}$	618.027	1.639
4.	$\beta_0 + bio1 + bio13 + \boldsymbol{u}$	620.013	1.653
5.	$\beta_0 + bio13 + \boldsymbol{u}$	620.293	1.641
6.	$\beta_0 + bio1 + bio7 + \boldsymbol{u}$	620.530	1.636
7.	$\beta_0 + bio7 + bio13 + \boldsymbol{u}$	620.620	1.661
8.	$\beta_0 + PRC + \boldsymbol{u}$	621.386	1.672
9.	$\beta_0 + bio1 + bio7 + bio13 + \boldsymbol{u}$	622.304	1.641
10.	$\beta_0 + bio1 + PRC + \boldsymbol{u}$	622.520	1.675
Models with principal components (PC)			
1.	$\beta_0 + \boldsymbol{u}$	617.627	1.638
2.	$\beta_0 + PC2 + \boldsymbol{u}$	618.887	1.656
3.	$\beta_0 + PC1 + \boldsymbol{u}$	619.558	1.635
4.	$\beta_0 + PC1 + PC2 + \boldsymbol{u}$	620.005	1.656
5.	$\beta_0 + PRC + \boldsymbol{u}$	621.399	1.672
6.	$\beta_0 + PC2 + PRC + \boldsymbol{u}$	622.634	1.689
7.	$\beta_0 + PC2 + PC3 + \boldsymbol{u}$	623.140	1.599
8.	$\beta_0 + PC1 + PC2 + PC3 + \boldsymbol{u}$	623.171	1.634
9.	$\beta_0 + PC1 + PC3 + \boldsymbol{u}$	623.364	1.608
10.	$\beta_0 + PC3 + \boldsymbol{u}$	623.530	1.584

[a] Intercept (β_0), annual mean temperature (*bio1*), temperature annual range (*bio7*), precipitation of the wettest month (*bio13*), first principal component ($PC1$), second principal component ($PC2$), third principal component ($PC3$), Purcell's risk categories (PRC) and spatial effect ($\boldsymbol{u}$).

[b] Watanabe Akaike information criterion (WAIC).

[c] Logarithmic score of conditional predictive ordinate (LCPO).

TABLE A.2: Principal component analysis (PCA) loadings for climatic variables in the demarcated area in Alicante, Spain, and Lecce, Italy.

	Alicante			Lecce		
Climatic covariates	PC1	PC2	PC3	PC1	PC2	PC3
bio1	0.963	-0.260	0	-0.636	-0.152	0.731
bio2	-0.236	0	0.967	0.955	0	0.138
bio3	0.750	0	0.633	0.923	-0.117	0.160
bio4	-0.926	0	0.347	0.919	0.102	0
bio5	0.888	-0.267	0.343	0.804	0	0.506
bio6	0.947	-0.146	-0.277	-0.964	0	0
bio8	0.934	-0.217	0.107	-0.502	-0.388	0.236
bio9	0.900	-0.388	0.119	0.286	-0.170	0.930
bio10	0.930	-0.320	0.134	0.216	-0.130	0.952
bio11	0.972	-0.202	-0.107	-0.870	-0.147	0.350
bio12	-0.882	0.460	0	0.228	0.922	0
bio13	0.297	0.936	-0.110	0	0.969	0
bio14	-0.937	0.322	0	0.458	-0.782	0
bio15	0.976	0	0	-0.397	0.869	-0.148
bio16	-0.404	0.902	0	0	0.967	-0.16
bio17	-0.967	0.222	0	0.755	-0.327	0
bio18	-0.445	0.836	0	0.796	-0.240	0
bio19	-0.659	0.721	0	0.101	0.959	0
ADD	0.968	-0.228	0	-0.141	-0.121	0.973
tmin	0.949	-0.179	-0.257	-	-	-
precd	-0.925	0.346	0	-	-	-
srad	-0.789	-0.147	0.258	-	-	-
% variability	69.7	18.2	8.5	38.9	28.7	20.1

TABLE A.3: Best models for the distribution of *Xylella fastidiosa* in Lecce, Italy.

	Model[a]	**WAIC**[b]	**LCPO**[c]
Models with climatic covariates			
1.	$\beta_0 + bio1 + bio2 + bio9 + \boldsymbol{u}$	1091.485	0.131
2.	$\beta_0 + bio2 + bio8 + bio9 + bio12 + \boldsymbol{u}$	1095.184	0.132
3.	$\beta_0 + bio2 + bio8 + bio12 + \boldsymbol{u}$	1095.307	0.132
4.	$\beta_0 + bio1 + bio2 + bio8 + bio12 + \boldsymbol{u}$	1095.709	0.132
5.	$\beta_0 + bio2 + bio8 + bio9 + \boldsymbol{u}$	1095.959	0.132
6.	$\beta_0 + bio2 + bio8 + bio9 + bio12 + bio14 + \boldsymbol{u}$	1096.218	0.132
7.	$\beta_0 + bio2 + bio8 + \boldsymbol{u}$	1096.343	0.132
8.	$\beta_0 + bio2 + bio8 + bio12 + bio14 + \boldsymbol{u}$	1096.348	0.132
9.	$\beta_0 + bio2 + bio12 + \boldsymbol{u}$	1096.362	0.132
10.	$\beta_0 + bio1 + bio2 + bio8 + \boldsymbol{u}$	1096.692	0.132
Models with principal components (PC)			
1.	$\beta_0 + PC1 + PC2 + PC3 + \boldsymbol{u}$	1078.769	0.130
2.	$\beta_0 + PC1 + PC3 + \boldsymbol{u}$	1079.789	0.130
3.	$\beta_0 + PC2 + PC3 + \boldsymbol{u}$	1080.430	0.130
4.	$\beta_0 + PC1 + PC2 + \boldsymbol{u}$	1081.277	0.130
5.	$\beta_0 + PC3 + \boldsymbol{u}$	1081.607	0.133
6.	$\beta_0 + PC1 + \boldsymbol{u}$	1081.974	0.133
7.	$\beta_0 + PC2 + \boldsymbol{u}$	1082.573	0.133
8.	$\beta_0 + \boldsymbol{u}$	1083.363	0.133
9.	$\beta_0 + PC1 + PC2$	1711.042	0.133
10.	$\beta_0 + PC1 + PC2 + PC3$	1712.478	0.133

[a] Intercept (β_0), annual mean temperature (*bio1*), mean diurnal range (*bio2*), mean temperature of the wettest quarter (*bio8*), mean temperature of the driest quarter (*bio9*), annual precipitation (*bio12*), precipitation of the driest month (*bio14*), first principal component ($PC1$), second principal component ($PC2$), third principal component ($PC3$), and spatial effect ($\boldsymbol{u}$).

[b] Watanabe Akaike information criterion (WAIC).

[c] Logarithmic score of conditional predictive ordinate (LCPO).

Appendix B

Supplementary figures for Chapter 4

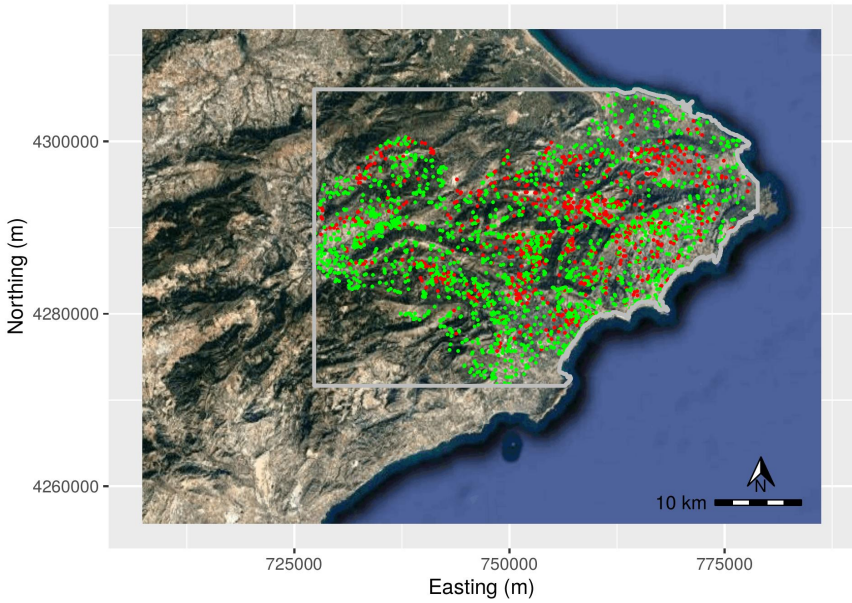

Figure B.1: Study area (delimited by the grey line) in Alicante, Spain, in UTM coordinate system, with the positive (●) and negative (●) samples for *Xylella fastidiosa* in 2018.

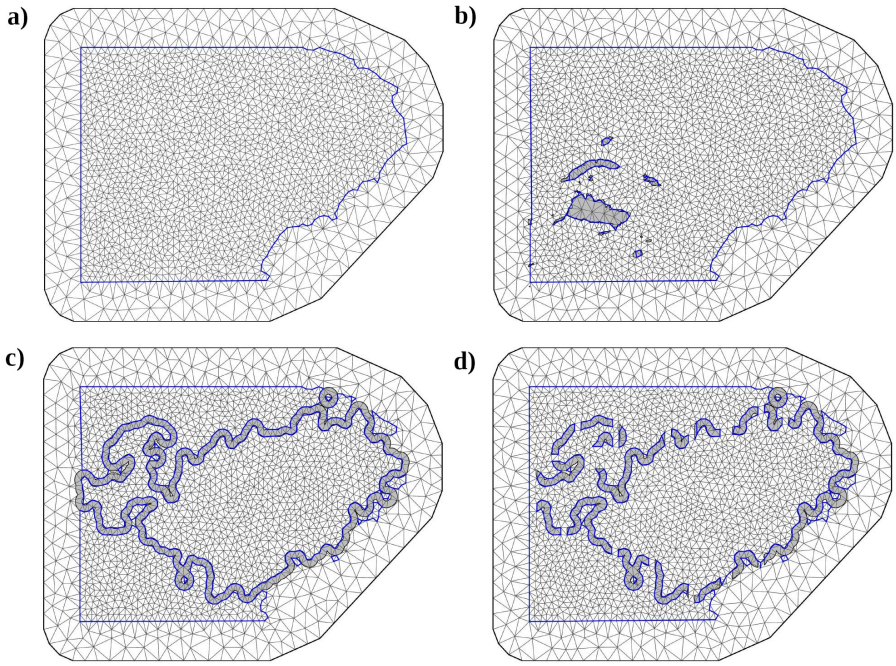

FIGURE B.2: Triangulation of the study region (*mesh*) for each model. (**a**) Stationary model; (**b**) mountain barrier model; (**c**) continuous barrier model; and (**d**) discontinuous barrier model.

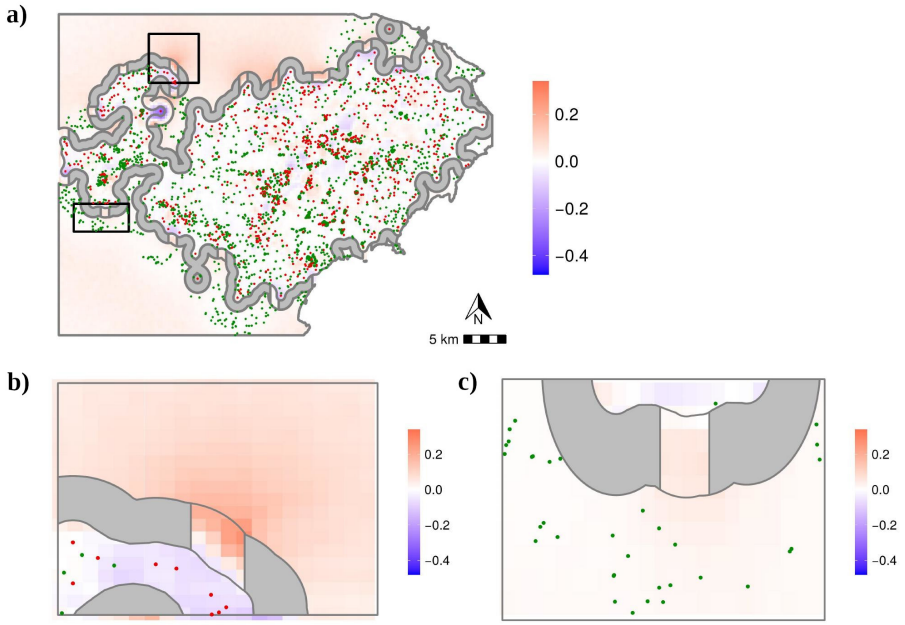

FIGURE B.3: (**a**) Difference in the mean of the posterior predictive distribution of the probability of *Xylella fastidiosa* presence between discontinuous and continuous barrier models, with the positive (●) and negative (●) samples for *X. fastidiosa*. (**b**) Detail of a break with non-sampled area adjacent to the outer border of the barrier. (**c**) Detail of a break with high sampling intensity adjacent to the outer border of the barrier.

Appendix C

Supplementary figures and tables for Chapter 6

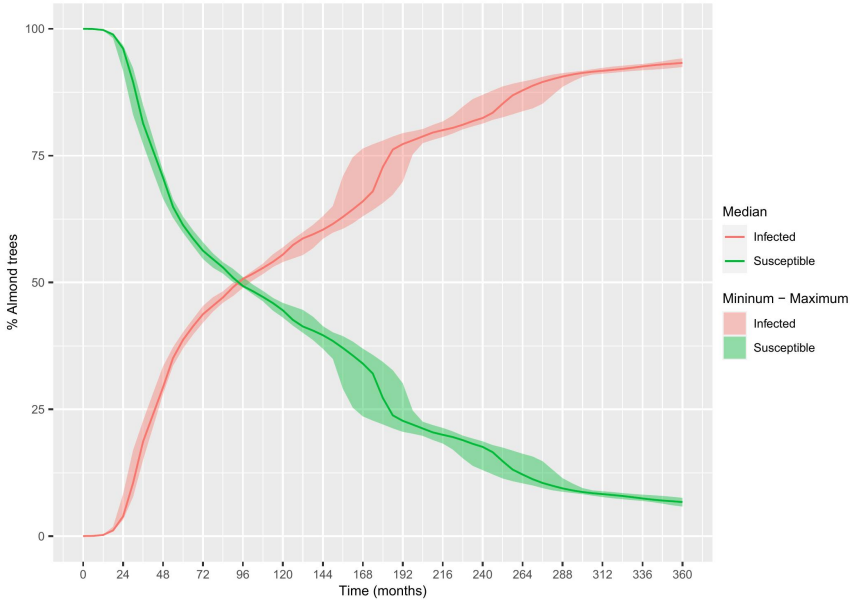

Figure C.1: Percentage of susceptible (green) and infected (red) trees relative to the total population over time for 360 months for the baseline scenario without interventions. The line represents the median of the simulations, based on the median number of trees in the 50 simulations at $t = 360$ months. The shaded area represents the interval between the minimum and maximum of the 50 simulations.

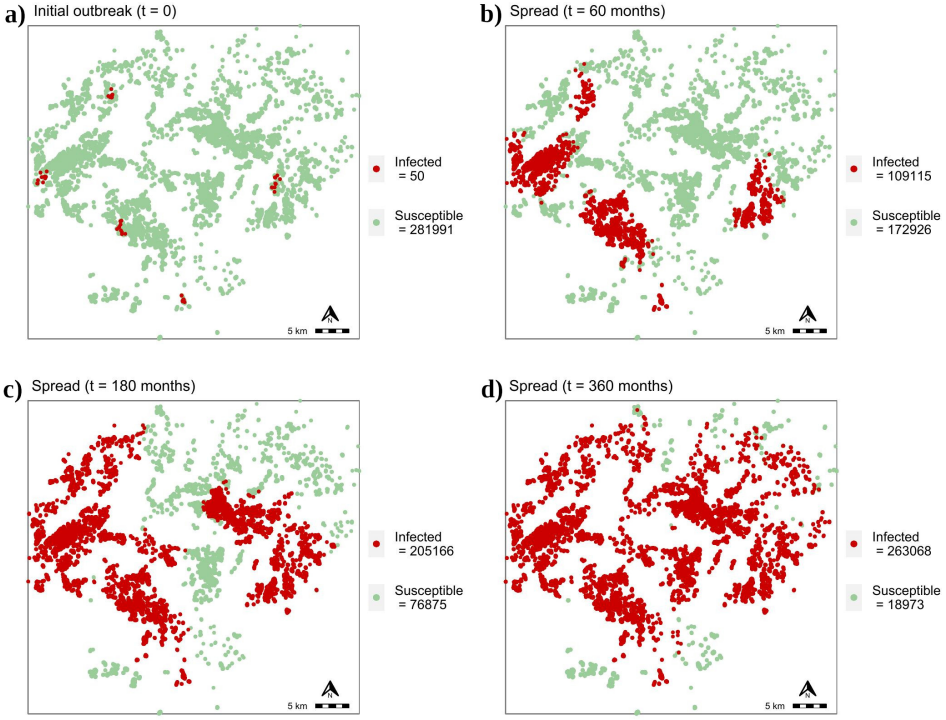

FIGURE C.2: Georeferenced distribution and number of susceptible and infected trees in the study area. (**a**) Initial outbreak at $t = 0$ distributed at 5 foci. Disease spread in the baseline scenario without interventions for the median of the simulations at (**b**) $t = 60$ months, (**c**) $t = 180$ months, and (**d**) $t = 360$ months.

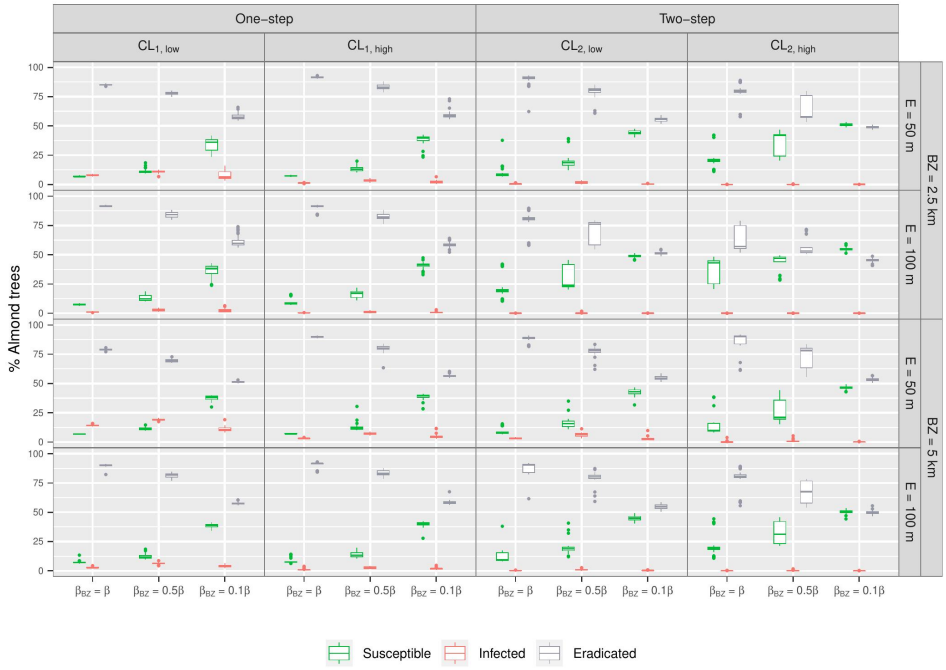

FIGURE C.3: Percentage of susceptible, infected, and eradicated trees relative to the total population at $t = 360$ months of the 50 simulations with each outbreak management plan, including surveillance strategies with the one-step and two-step survey approaches at different values of the confidence level (CL_{low} and CL_{high}), and control strategies including buffer zone (BZ), eradication radius (E), and reduction in the transmission rate in the buffer zone (β_{BZ}).

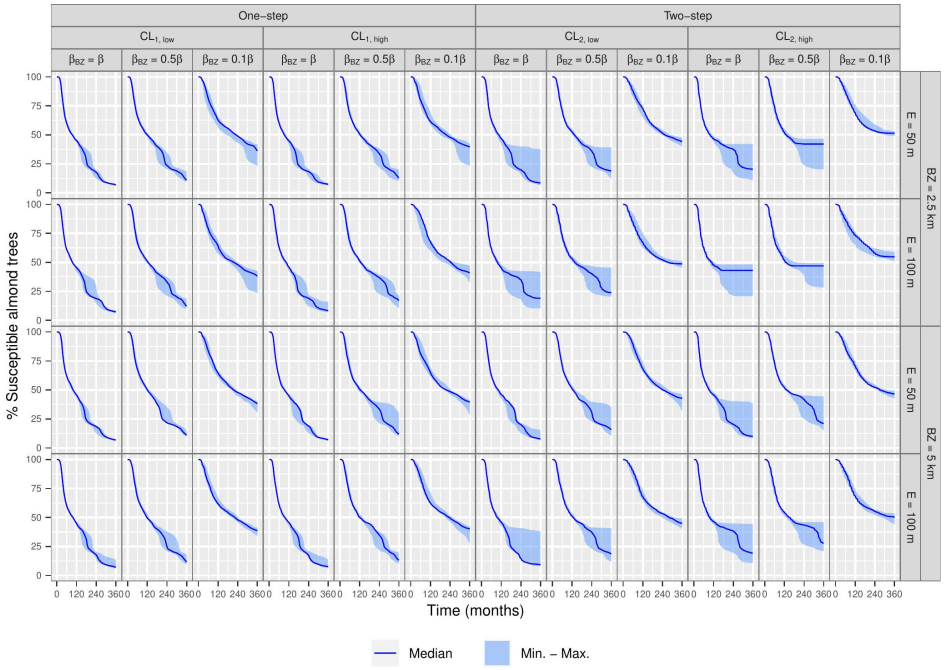

FIGURE C.4: Percentage of susceptible trees relative to the total population over time with each outbreak management plan, including surveillance strategies with the one-step and two-step survey approaches at different values of the confidence level (CL_{low} and CL_{high}), and control strategies including buffer zone (BZ), eradication radius (E), and reduction in the transmission rate in the buffer zone (β_{BZ}). The line represents the median of the simulations, based on the median number of trees in the 50 simulations at $t = 360$ months. The shaded area represents the interval between the minimum and maximum of the 50 simulations.

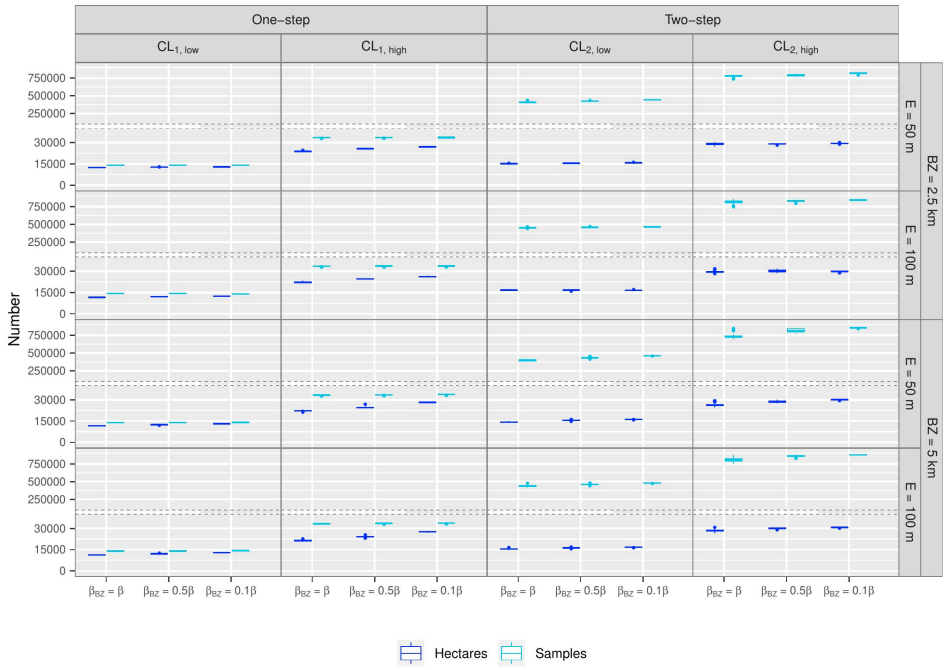

FIGURE C.5: Number of hectares and total accumulated samples after $t = 360$ months of the 50 simulations with each outbreak management plan, including surveillance strategies with the one-step and two-step survey approaches at different values of the confidence level (CL_{low} and CL_{high}), and control strategies including buffer zone (BZ), eradication radius (E), and reduction in the transmission rate in the buffer zone (β_{BZ}).

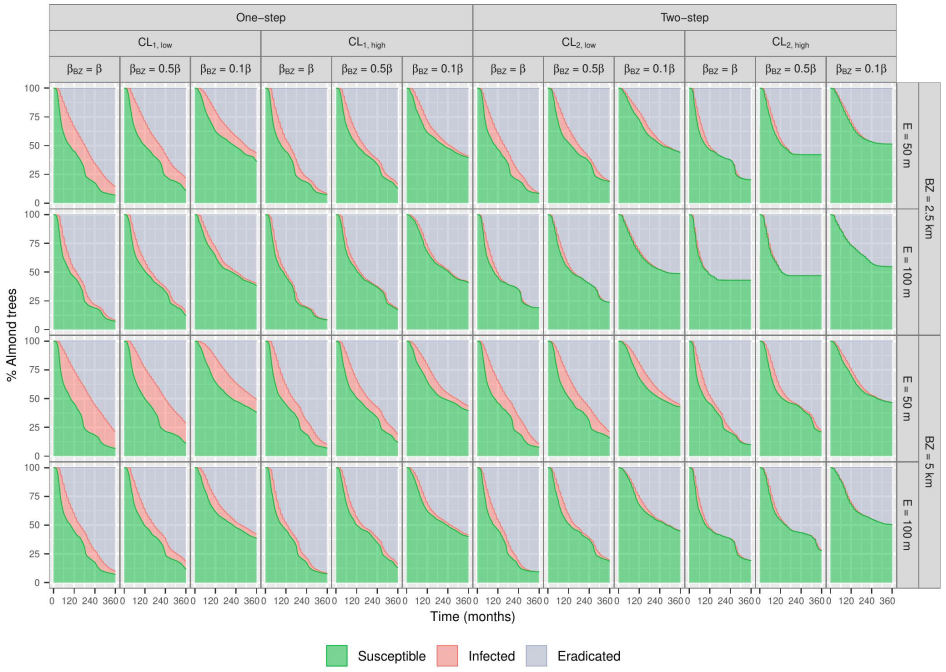

FIGURE C.6: Percentage of susceptible, infected, and eradicated trees relative to the total population over time for the median of the simulations with each outbreak management plan, including surveillance strategies with the one-step and two-step survey approaches at different values of the confidence level (CL_{low} and CL_{high}), and control strategies including buffer zone (BZ), eradication radius (E), and reduction in the transmission rate in the buffer zone (β_{BZ}).

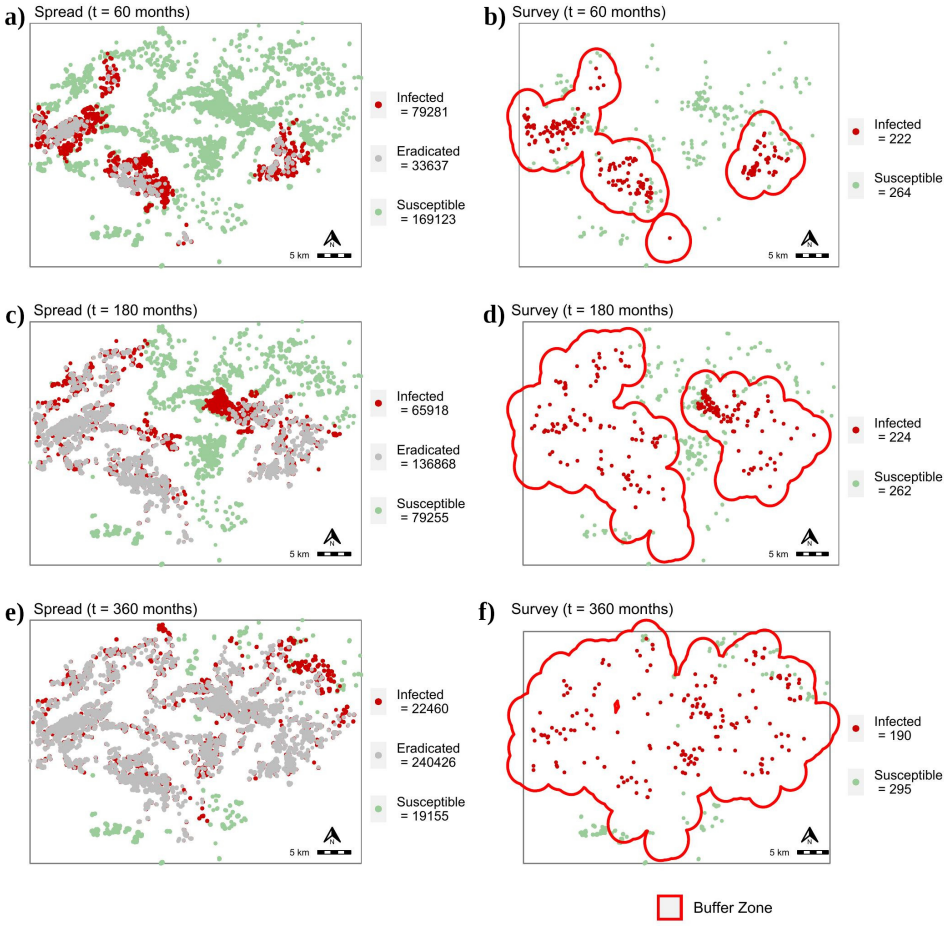

FIGURE C.7: Georeferenced distribution and number of susceptible, infected, and eradicated trees (left), and surveillance and buffer zone delimitation (right), for the less efficient outbreak management plan, which included the one-step survey approach, CL_{low}, $BZ = 2.5$ km, $E = 50$ m, $\beta_{BZ} = \beta$, at (**a, b**) $t = 60$ months, (**c, d**) $t = 180$ months, and (**e, f**) $t = 360$ months.

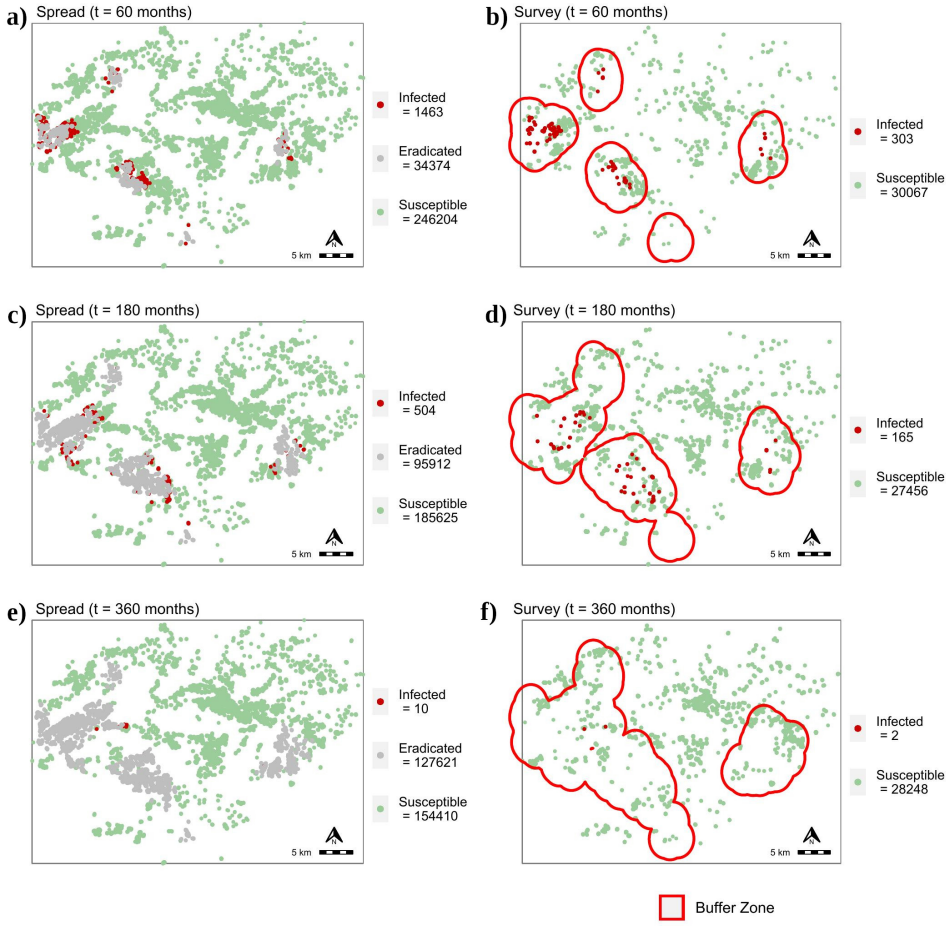

FIGURE C.8: Georeferenced distribution and number of susceptible, infected, and eradicated trees (left), and surveillance and buffer zone delimitation (right), for the most efficient outbreak management plan, which included the two-step survey approach, CL_{high}, $BZ = 2.5$ km, $E = 100$ m, $\beta_{BZ} = 0.1\beta$, at (**a, b**) $t = 60$ months, (**c, d**) $t = 180$ months, and (**e, f**) $t = 360$ months.

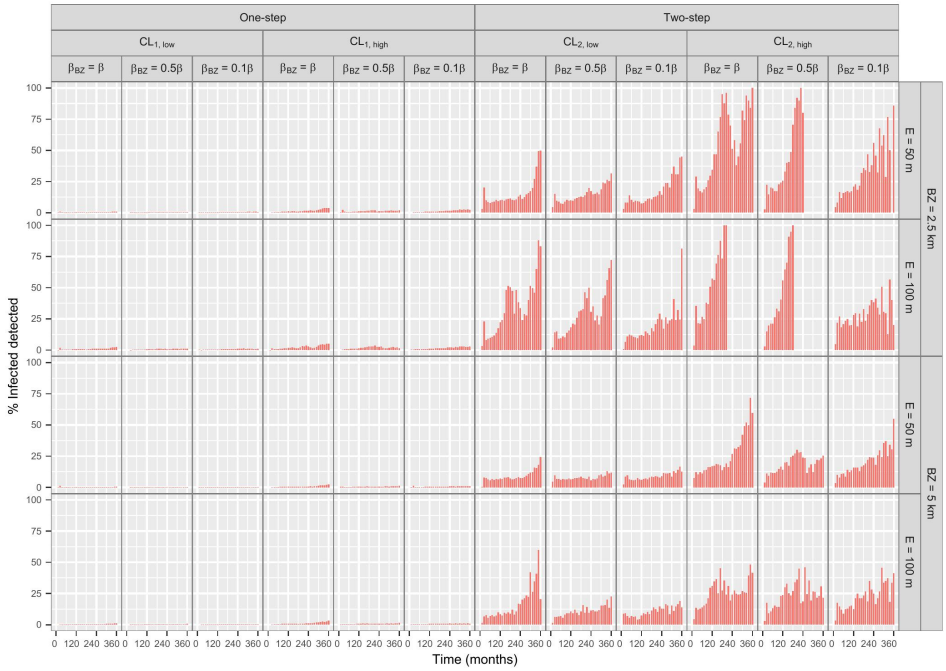

FIGURE C.9: Percentage of infected samples (detections) per year relative to the total of the infected population for the median of the simulations with each outbreak management plan, including surveillance strategies with the one-step and two-step survey approaches at different values of the confidence level (CL_{low} and CL_{high}), and control strategies including buffer zone (BZ), eradication radius (E), and reduction in the transmission rate in the buffer zone (β_{BZ}).

TABLE C.1: Ranking of outbreak management plans in terms of effectiveness, i.e., percentage reduction in the number of infected trees with each outbreak management plan in relation to the baseline scenario without interventions, and efficiency, i.e., percentage increase in the number of susceptible trees that were not eradicated with each outbreak management plan in relation to the baseline scenario without interventions. The ranking position for both effectiveness and efficiency is indicated in parentheses.

Surveillance approach	CL	BZ (km)	E (m)	β_{BZ}	Effectiveness		Efficiency	
Two-step	$CL_{2,high}$	2.5	100	0.5β	(1)	100	(5)	597.9603
Two-step	$CL_{2,high}$	2.5	100	β	(1)	100	(9)	539.7196
Two-step	$CL_{2,high}$	2.5	50	0.5β	(1)	100	(11)	524.9091
Two-step	$CL_{2,low}$	2.5	100	β	(2)	99.9977	(25)	182.6965
Two-step	$CL_{2,high}$	2.5	50	0.1β	(3)	99.9973	(2)	663.1055
Two-step	$CL_{2,high}$	2.5	100	0.1β	(4)	99.9962	(1)	713.8407
Two-step	$CL_{2,high}$	2.5	50	β	(5)	99.9943	(23)	202.3349
Two-step	$CL_{2,low}$	2.5	100	0.1β	(6)	99.9776	(4)	623.5861
Two-step	$CL_{2,low}$	2.5	100	0.5β	(7)	99.9753	(21)	255.0625
Two-step	$CL_{2,low}$	5	100	β	(8)	99.9723	(38)	39.435
Two-step	$CL_{2,high}$	5	100	0.1β	(9)	99.9658	(3)	649.3016
Two-step	$CL_{2,high}$	5	50	β	(10)	99.9574	(37)	48.9011
Two-step	$CL_{2,high}$	5	100	β	(11)	99.8886	(24)	186.4544
Two-step	$CL_{2,high}$	5	50	0.1β	(12)	99.8719	(6)	591.6724
Two-step	$CL_{2,low}$	2.5	50	0.1β	(13)	99.8308	(8)	557.3763
Two-step	$CL_{2,low}$	5	100	0.1β	(14)	99.6491	(7)	569.1983
One-step	$CL_{1,high}$	2.5	100	β	(15)	99.6066	(40)	25.3044
Two-step	$CL_{2,low}$	2.5	50	β	(16)	99.4735	(39)	25.5099
One-step	$CL_{1,high}$	2.5	100	0.1β	(17)	99.4507	(12)	511.353
Two-step	$CL_{2,low}$	2.5	50	0.5β	(18)	99.196	(26)	179.397
One-step	$CL_{1,high}$	5	100	β	(19)	99.1538	(42)	12.755
Two-step	$CL_{2,high}$	5	100	0.5β	(20)	99.1303	(20)	313.9145
One-step	$CL_{1,low}$	2.5	100	β	(21)	98.7482	(44)	8.6017
Two-step	$CL_{2,low}$	5	100	0.5β	(22)	98.6673	(27)	177.6419
One-step	$CL_{1,high}$	2.5	50	β	(23)	98.5566	(43)	8.8652
One-step	$CL_{1,high}$	2.5	50	0.1β	(24)	98.4057	(14)	488.7208
One-step	$CL_{1,high}$	2.5	100	0.5β	(25)	98.3768	(28)	154.9465
One-step	$CL_{1,high}$	5	100	0.1β	(26)	98.3442	(13)	496.1682
Two-step	$CL_{2,high}$	5	50	0.5β	(27)	98.2354	(22)	213.4507
Two-step	$CL_{2,low}$	5	50	0.1β	(28)	97.8941	(10)	538.0014
One-step	$CL_{1,low}$	2.5	100	0.1β	(29)	97.7337	(17)	467.9545
One-step	$CL_{1,low}$	5	100	β	(30)	96.8913	(45)	5.803
One-step	$CL_{1,high}$	5	100	0.5β	(31)	96.672	(30)	96.3685
Two-step	$CL_{2,low}$	5	50	β	(32)	96.572	(41)	15.4272
One-step	$CL_{1,high}$	5	50	β	(33)	96.4382	(46)	4.3377
One-step	$CL_{1,low}$	5	100	0.1β	(34)	95.9664	(16)	475.228

Continued on next page

Survey approach	CL	BZ (km)	E (m)	β_{BZ}	Effectiveness		Efficiency	
One-step	$CL_{1,low}$	2.5	100	0.5β	(35)	95.9588	(32)	83.076
One-step	$CL_{1,high}$	2.5	50	0.5β	(36)	95.7391	(31)	92.8794
One-step	$CL_{1,high}$	5	50	0.1β	(37)	95.3423	(15)	487.6561
Two-step	$CL_{2,low}$	5	50	0.5β	(38)	93.9362	(29)	133.321
One-step	$CL_{1,low}$	5	100	0.5β	(39)	92.3837	(34)	72.9616
One-step	$CL_{1,high}$	5	50	0.5β	(40)	92.0359	(33)	78.0952
One-step	$CL_{1,low}$	2.5	50	0.1β	(41)	91.7151	(19)	436.4413
One-step	$CL_{1,low}$	2.5	50	β	(42)	91.4623	(48)	0.9593
One-step	$CL_{1,low}$	2.5	50	0.5β	(43)	88.1476	(36)	62.4256
One-step	$CL_{1,low}$	5	50	0.1β	(44)	87.6614	(18)	467.8174
One-step	$CL_{1,low}$	5	50	β	(45)	84.1794	(47)	1.3493
One-step	$CL_{1,low}$	5	50	0.5β	(46)	80.9228	(35)	65.9095