

Article

A Stochastic Simulation Framework to Predict the Spatial Spread of *Xylella fastidiosa*

Nikolaos Marios Polymenakos , Iosif Polenakis , Christos Sarantidis , Ioannis Karydis 
and Markos Avlonitis 

Department of Informatics, Ionian University, 7 Tsirigoti Square, 49132 Corfu, Greece;
polymenakos@ionio.gr (N.M.P.); karydis@ionio.gr (I.K.);

* Correspondence: ipolenakis@ionio.gr

Abstract

The spread of *Xylella fastidiosa*, a xylem-limited bacterial pathogen, has caused widespread mortality among olive trees in Apulian region, Italy in more than a decade, and represents a significant threat to Mediterranean agroecosystems. To encourage evidence-based containment strategies, we developed a stochastic, spatiotemporal simulation model that represents pathogen transmission at the individual-tree level. This work integrates high-resolution georeferenced olive-tree data and implicitly incorporates vector population dynamics through a tree-specific vulnerability index, which considers local host density and landscape connectivity. Vector dispersal is approximated using a radial transmission kernel, which preserves host–vector spatial interactions while avoiding the explicit modeling of insect trajectories. The system’s spatial structure is additionally formulated as a proximity graph, facilitating network-based analysis of spread pathways. A series of Monte Carlo simulation experiments is employed for calibration against the observed epidemic footprint, while validation utilizes independent infection records and global sensitivity analysis of key parameters. The findings indicate that the model effectively replicates realistic propagation patterns, and its calibrated parameters are consistent with out-of-sample data. This makes it an appropriate exploratory tool for scenario testing, assessing the potential impact of intervention strategies, and offering risk-based decision support for handling *Xylella fastidiosa* outbreaks. Subsequently, graph centrality metrics are used to identify epidemiologically critical trees that function as transmission bridges, thus representing priority targets for surveillance or removal efforts. Thus, multiple tests have been conducted using betweenness and closeness centrality, while comparing both methods leads to effective node-tree removal decisions.

Keywords: *Xylella fastidiosa*; spatial epidemiology; stochastic modeling; vector-borne plant disease; disease management

MSC: 92D30; 65C05; 05C90; 60H30



Academic Editors: Thomas M.
Michelitsch and Alejandro
Pérez Rascos

Received: 30 January 2026

Revised: 20 February 2026

Accepted: 22 February 2026

Published: 2 March 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Xylella fastidiosa (Xf) is a Gram-negative, xylem-restricted bacterium discovered by Wells et al. [1], belonging to the *Xanthomonadaceae* family [2]. It is recognized as one of the most dangerous plant bacteria worldwide, categorized as a priority quarantine pest in the European Union according to Bragard et al. in 2019 [3]. The pathogen colonizes the xylem vessels of host plants, forming biofilms that obstruct water and nutrient transport,

eventually leading to leaf scorching, canopy dieback, and plant death. According to European Food Safety Authority (EFSA) [4], at least 696 different plant species are recorded as hosts; however, not all of them are susceptible to every subspecies of the pathogen. Xf (subspecies *pauca*) was first identified in Europe in October 2013, in the Gallipoli region of the Salento peninsula (Apulia, Southern Italy) by Saponari et al. [5]. It was found to be the causal organism of a devastating disease afflicting olive trees (*Olea europaea*) named the Olive Quick Decline Syndrome (OQDS), which is marked by the rapid withering of the trees.

Unlike airborne pathogens, Xf is transmitted exclusively by insect vectors. In Europe, the primary vector, especially in olive orchards, is the meadow spittlebug, *Philaeus spumarius* (Hemiptera: Aphrophoridae) [6]. The epidemiology of the disease is strictly linked to the vector's life cycle and feeding behavior. Nymphs, which appear in the spring on herbaceous weeds, are not responsible for the transmission of bacteria between the olive trees; spreading takes place only through adult insects, during the late spring and summer months [7]. As a result, the probability of contracting is clearly seasonal and increases during the peak period of the vector's activity, a critical factor for any practical epidemiological model. In addition, the bacteria are actively acting against the xylem flow within the plant, rather than passively, which further complicates the window of opportunity for vector acquisition and the timing of the appearance of symptoms [8].

The spread of *Xylella fastidiosa*, a xylem-limited bacterial pathogen, has caused widespread olive tree mortality in the Apulian region of Italy for over a decade, posing a serious threat to Mediterranean agroecosystems. Mathematical and computational modeling plays a crucial role in addressing such challenges by linking theory to practice and serving as a virtual laboratory for exploring complex, non-linear interactions among biological, environmental, and social factors [9]. In this work, we developed a stochastic spatiotemporal simulation model that represents pathogen transmission at the individual-tree level. The model integrates high-resolution georeferenced olive-tree data and implicitly accounts for vector population dynamics through a tree-specific vulnerability index based on local host density and landscape connectivity. Vector dispersal is approximated using a radial transmission kernel, preserving host–vector spatial interactions without explicitly modeling insect trajectories, while the spatial structure of the system is represented as a proximity graph to enable network-based analysis of spread pathways. Model calibration is achieved through Monte Carlo simulation experiments matched against the observed epidemic footprint, and validation is performed using independent infection records and global sensitivity analysis.

Motivation & Contributions

Considering the central importance of olive cultivation for Mediterranean countries, the conservation of these crops is not only economically vital, but also essential to preserve the cultural heritage that has been thriving in the region for millennia, as mentioned in these studies [10–12]. Therefore, there is an urgent need to set up a comprehensive infrastructure dedicated to the monitoring, research and management of Xf. This pro-active approach is essential to protect the large olive groves, which contribute significantly to the local and national economy.

Such infrastructure would set a precedent for agricultural preparedness, ensuring an effective response to potential future outbreaks in Greece and elsewhere. Immediate and decisive action is needed to safeguard the olive-growing tradition for future generations. In this context, reliable computer modeling is the cornerstone of this defense infrastructure, enabling authorities to test containment strategies in silico before an outbreak occurs.

However, despite the availability of different modeling approaches, flexible simulation tools are needed to integrate high-resolution spatial data with stochastic transmission dynamics. This is particularly important in the Mediterranean regions currently classified as free of Xf, where the lack of historical data on infections makes it impossible to use purely statistical models. In such cases, biologically informed geometrical graphs provide a proactive tool for what-if analysis, enabling the quantification of the risk based on the current spatial arrangement of the host population.

The main motivation for this work is to contribute to this agricultural preparedness framework by developing a spatially explicit geometrical simulation tool. By presenting olive orchards as a network with nodes corresponding to the actual location of the trees, we aim to provide a realistic risk assessment tool. In addition to its methodological contribution, the effort is motivated by a local perspective: due to the geographical proximity of Corfu to Italy and its location along plausible routes of pathogen introduction, the local olive groves may be exposed to increased exposure, which reinforces the need for anticipative and practical decision support tools.

The contributions of this study are:

1. Construction of a realistic spatial network of olive trees using real coordinate data, to simulate the Xf spreading in the olive orchard.
2. Implementation of a stochastic Susceptible-Infected (SI) epidemiological model using Monte Carlo simulations to capture the uncertainty of disease spread.
3. Parameterization of the model with biologically relevant data, specifically the seasonal transmission efficiency and dispersal kernel of the vector *Philaenus spumarius*.
4. Evaluation of the progression of the epidemic over time and identification of key nodes-trees as transmission bridges.
5. Demonstration and assessment of targeted mitigation strategies via betweenness and closeness centrality metrics.

The structure of the paper is as follows: Section 2 reviews the necessity of pathogen analysis and existing modeling frameworks, while Section 3 details the proposed methodology. Section 4 outlines the simulation and mitigation results. The paper finally concludes in Section 5 where results, discussion and future work are presented.

2. Background & Related Work

Olive cultivation (*Olea europaea*) is a cornerstone of the identity and economy of the Mediterranean region. Paleobotanical and genetic studies show that olive trees are not only a crop, but also a cultural symbol, which has shaped rural Mediterranean landscapes for thousands of years [11,12]. The emergence of the phytopathogenic bacteria Xf, which is the causative agent of Olive Quick Decline Syndrome (OQDS) is therefore an unprecedented threat to this social and ecological system.

Since its initial detection, the epidemic has expanded northward at an alarming rate of 10 to 20 km per year [13,14]. The consequences for the Apulian region have been catastrophic, leading to what has been described as a “ten-year evaluation” of complex and often fraught eradication efforts [15]. Beyond the irreparable cultural damage from the loss of millions of centennial olive trees, the economic impact is severe [16]. Forecasts suggest that without effective containment, cumulative losses could reach EUR 1.9–5.2 billion over 50 years, threatening major EU producers like Italy, Greece, and Spain [16]. The threat to Greece is considered particularly high due to its geography, as the Hellenic Ministry of Rural Development and Food confirms [17]. This economic danger is compounded by the loss of ecosystem services and, as the EFSA notes, a long asymptomatic period in infected trees that makes visual inspection ineffective and reinforces the need for advanced modeling tools [3].

2.1. Spatial and Network Modeling in Plant Epidemiology

Given the practical constraints and high costs of large-scale field experiments, mathematical modeling has become an essential tool for comprehending the epidemiology of Xf [14]. Deterministic population dynamics and spatially explicit stochastic simulations are just two of the modeling frameworks that have been put forth over the past ten years to predict disease progression and assess containment tactics.

Much of the existing work relies on deterministic formulations, typically reaction-diffusion equations or compartmental schemes (e.g., Susceptible-Infected-Recovered). For example, Anița et al. [18] introduced a spatial reaction-diffusion model aimed at optimizing control actions such as weed removal, showing that, at least theoretically, the epidemic front could be stopped. Additionally, Gilioli et al. [19] proposed an eco-epidemiological differential-equation framework (XEM) to represent interactions among the host, the vector, and the bacterium. Although these continuous models are informative for broad trends and mean-field behavior, they can struggle to represent the fine-scale, discrete heterogeneity that characterizes real agricultural landscapes. In addition, cultivar-level differences may matter in practice, since susceptible and resistant olive cultivars can display distinct physiological responses to Xf infection [20].

Recent research is increasingly using statistical inference and methods that take spatial characteristics into account, acknowledging the limitations of continuous models. Soubeyrand et al. [21] highlighted the difficulties in reconstructing spread using monitoring records by employing temporal surveillance count data to determine the invasion history in Corsica (France). Subsequently, Chapman et al. [22] evaluated dispersal characteristics and compared different transmission methods using Approximate Bayesian Computation (ABC). Through spatial Bayesian hierarchical modeling, Cendoya et al. [23] significantly demonstrated that the geographical dependency indicated in the data can be a more powerful predictor of Xf spread than climate factors alone.

In this context, Network Theory offers a robust alternative for modeling plant epidemics, treating trees or orchards as nodes and potential transmission pathways as edges. Strona et al. [24] provided a pivotal analysis using network science, demonstrating that the specific spatial configuration of olive groves in Southern Italy creates a highly connected orchard network with large connected components and short effective distances that facilitates rapid disease propagation and persistence. Building on this network-centric view, Martinetti et al. [25] combined network analysis with machine learning to identify high-risk “lookout” locations for early surveillance. Unlike mean-field models, network-based approaches such as the lattice models explored by Fierro et al. [26] or the spatially explicit simulations by White et al. [14] allow for the incorporation of stochastic events, long-distance dispersal kernels, and vector seasonality. A critical challenge in network-based approaches is specifying the dispersal kernel. Risk assessments and modeling studies distinguish between predominantly local spread and less frequent longer-range dispersal events, which can strongly influence regional dynamics [3]. In a geometric-graph setting, this can be represented by combining dense, short-range connections (within and between nearby orchards) with a sparse set of longer-range links, offering a pragmatic way to capture the apparent leap-frogging of the epidemic front.

The integration of high-resolution spatial data into a geometric graph framework, combined with Monte Carlo simulation, is therefore a promising frontier for developing predictive tools that reflect the stochastic nature of pathogen invasions. The critical issue remains the accurate representation of dispersal at scales: while local transmission processes are relatively better characterised, rare long-distance events may disproportionately shape regional dynamics and are difficult to capture by averaging. These jumps may be caused by wind-borne passive transport or through human-assisted movement (e.g., vectors or

infected plant material). By adopting a multi-scale dispersal kernel, network models can reproduce the apparent leapfrog behaviour of the epidemic front reported in Apulia [3,14].

2.2. The Role of the Vector and the Anthropogenic Landscape

Epidemiological modeling of Xf requires a thorough understanding of the behavior of its insect vector, *Philaenus spumarius*. Bodino et al. [7] analyzed the dispersal dynamics of the vector between olive groves and natural vegetation, demonstrating that the insect's mobility exhibits seasonal fluctuations that are decisive for disease transmission.

Moreover, the structure of the terrain seems to be crucial. The anthropogenic environment, specifically the road network, acts as a key accelerator of disease diffusion, according to recent research by Bajocco et al. [27]. This is probably because infected insects are passively transported via automobiles ("hitchhiking"). These results support the idea that high-resolution spatial data on agricultural distribution and human infrastructure must be included in realistic spread models.

2.3. Deterministic Approaches and Control Strategies

Early research focused on using differential equation models (SIR-type) to comprehend fundamental dynamics. In order to identify the best control tactics, Brunetti et al. [28] and Anita et al. [18] created deterministic models with an emphasis on weed removal and vector population reduction. Simultaneously, White et al. [14] used simulations to evaluate the effectiveness of buffer zones, emphasizing how critical it is to remove affected trees quickly in order to stop the invasion front. More recently, Ali et al. [29] suggested using fractional-order models to better explain the system's "memory" effect on the evolution of disease.

2.4. Stochastic Models and Bayesian Inference

Given the inherent uncertainty of biological systems, research has shifted towards stochastic and statistical methods. Strona et al. [24], applying Network Analysis, demonstrated that the high connectivity of olive groves in Southern Italy makes the complete eradication of the pathogen extremely difficult, suggesting that the disease tends to become endemic.

A significant breakthrough was the introduction of Bayesian inference. Soubeyrand et al. [21] developed mechanistic-statistical models that allowed for the reconstruction of the invasion history, revealing that "cryptic" (asymptomatic) spread often precedes detection by several years. This approach was extended by Cendoya et al. [23], who compared spatial models across different geographical zones (Spain and Italy).

Regarding the most recent developments, Boters-Pitarch et al. [30] presented stochastic models that account for climatic parameters, such as wind direction, to predict dispersal. Finally, Chapman et al. [22] applied simulations based on Approximate Bayesian Computation (ABC) to accurately quantify the impact of various containment scenarios, explicitly accounting for parameter uncertainty.

2.5. Research Gap and Contribution Positioning

Despite significant progress, a gap remains in integrating high-resolution spatial data, seasonal vector dynamics, and multi-scale dispersal dynamics within a single, operational simulation framework. Many models are either continuous systems that lack fine-scale landscape detail and potential cultivar-level differences in susceptibility [18–20], or statistical models that require extensive historical data, limiting their use in disease-free regions [21,23]. Furthermore, few frameworks explicitly couple the crucial role of anthropogenic structures like road networks [27], the biological seasonality of the

vector [7], and the multi-scale nature of dispersal, which combines local spread with long-range “jumps” [3,14].

Our work addresses this gap by developing a spatially explicit, network-based simulation framework (available in [31]), that synthesizes these critical elements. We represent individual trees as nodes in a unweighted undirected graph, applying principles from network science to identify vulnerable components and transmission pathways [24,25]. Into this structure, we incorporate a multi-level dispersal kernel and a time-varying seasonality factor to create a flexible tool for assessing risk. This allows for the evaluation of targeted mitigation strategies (such as the removal of critical “bridge” nodes) in a way that is sensitive to the real-world spatial configuration of the host population, extending prior simulation studies [14,22].

3. Proposed Methodology

In this study, we developed a stochastic framework to simulate the spatiotemporal dynamics of *Xf*. Our approach bridges high-resolution spatial data with a probabilistic engine to capture both local and long-distance transmission events.

The proposed simulation pipeline is structured into five main layers, as depicted in the methodological flowchart, Figure 1. The process integrates georeferenced tree data into a graph-based structure, applies monthly seasonality factors and transmission kernels for spread simulation, with a sensitivity-driven calibration and validation, and concludes with two spreading mitigation methods.

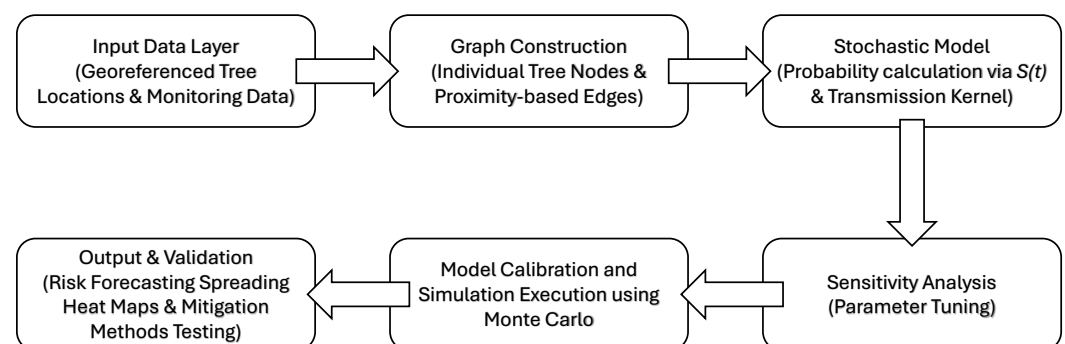


Figure 1. Methodological workflow of the proposed stochastic simulation framework.

3.1. Study Area and Data Description

The study focuses on the Apulia region in southern Italy, an area severely affected by the spread of *Xylella fastidiosa* subsp. *pauca*. The region is characterized by extensive olive monocultures, creating a continuous landscape that facilitates vector-mediated transmission. The climatic conditions of mild winters and hot, dry summers strongly influence the population dynamics of the primary vector, *Philaenus spumarius*.

The simulation employs a dataset of olive trees geolocated and monitored during the 2017 epidemic wave, that of the official monitoring database of the Apulia Region Phytosanitary Service [32]. The unfiltered data had been pre-processed to obtain a study region with $N = 262$ trees (`filtered_trees_2017.csv`). The ground truth is represented by the dataset on which the models are going to be calibrated and validated, and the entries are determined by a unique Node ID, geographical coordinates, and disease status. In order to simplify the process of estimating the infection time, the sampling plan was focused on the high-density areas and locations where the changes in the state (i.e., negative to positive) were reported. The latter dataset is comprised of the negative and positive observations that had been verified in late 2017.

3.1.1. Rationale Olive Orchard Selection

The selection of the study area was strategically focused on high-density olive orchards where documented transitions from asymptomatic to symptomatic states (negative-to-positive, N-to-P) were recorded during the 2017 monitoring cycle. This specific subset was prioritized over larger, sparser regional datasets because N-to-P transitions are critical for calibrating the stochastic transmission rate (β) and capturing the actual epidemic progression, rather than a static snapshot of infection. Furthermore, the high host density in this region was essential for the construction of a meaningful proximity graph, allowing for the analysis of both local diffusion and long-range jump dynamics. In our approach, the study area was not chosen randomly. Our simulation framework requires high-resolution data that captures the *transition* of the pathogen's state. Specifically, we required a subset of the official Apulian monitoring database that met two critical criteria:

1. **Dynamic Information (Negative-to-Positive Transitions):** To calibrate a stochastic spread model, it is essential to have records of trees that were initially asymptomatic/negative and became confirmed positive during the monitoring period. Most regions in the national database provide only static “snapshot” data; the selected area is one of the few that provides the necessary temporal resolution to model the actual progression of the disease.
2. **Structural Integrity (Host Density):** A key feature of our work is the Proximity Graph. For the graph to be epidemiologically meaningful, the host (olive tree) density must be high enough to allow for realistic transmission kernels. Sparse areas would result in disconnected graphs that fail to capture the “bridge” dynamics we aimed to study.

Given that this specific subset provided the highest quality of both spatial (density) and temporal (N-to-P transitions) data, the use of temporal splitting (30-day vs. Full Year) was the most rigorous way to test the model's predictive horizon without sacrificing data quality. While this approach means the parameters are “tuned” to the local Apulian context, the robustness of the conclusions lies in the model's internal consistency and its ability to identify “Super-Spreaders”. Even if the absolute transmission rate (β) might vary in a different geographic region, the *relative importance* of nodes based on their network centrality remains a valid and powerful insight for decision-support.

3.1.2. Olive Orchard Demarcation

The methodology for selecting the experimental area is outlined in this section. The core process was to define a rectangular area of interest using a custom-made algorithm, to ensure precision and repeatability, with the final subset of data being saved for further analysis. The filtering process begins by loading a dataset of geographic coordinates from a source file. The core logic for selecting the desired subset of data is described in the following flowchart:

The filtering boundary limits are derived from the set of user-defined input points $P = \{a_1, b_1, a_2, b_2\}$, as illustrated in Figure 2. The coordinates for the bounding box corners z_1 (top-left) and z_2 (bottom-right) are calculated by extracting the extrema of the input coordinates:

$$\begin{aligned} \text{lat}_{\max} &= \max_{p \in P}(\text{lat}_p) & \text{lat}_{\min} &= \min_{p \in P}(\text{lat}_p) \\ \text{lon}_{\max} &= \max_{p \in P}(\text{lon}_p) & \text{lon}_{\min} &= \min_{p \in P}(\text{lon}_p) \end{aligned}$$

Consequently, the definitive corner points for the selection area are established as:

$$z_1 = (\text{lon}_{\min}, \text{lat}_{\max}) \quad \text{and} \quad z_2 = (\text{lon}_{\max}, \text{lat}_{\min})$$

This method ensures that a precise and well-defined geographic area is isolated for further analysis. Figure 2 provides a detailed view of this selection with annotations.

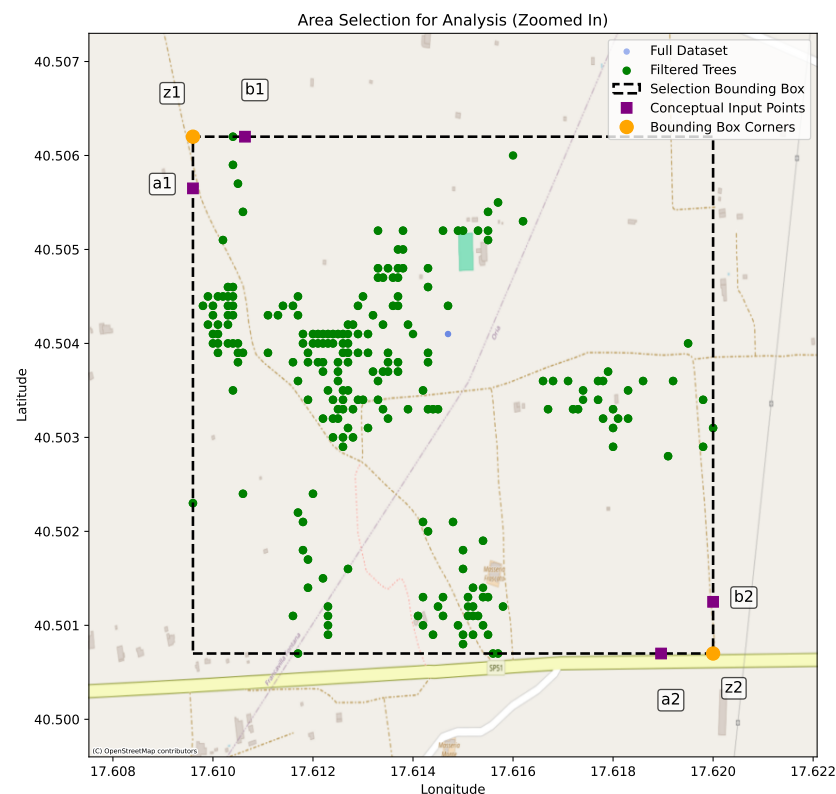


Figure 2. A zoomed-in view of the selected area.

3.2. Graph Construction and Representation

We replicate the olive orchard as a proximity graph $G = (V, E)$, where the set of vertices $V = \{v_1, v_2, \dots, v_n\}$ represents n individual olive trees derived from georeferenced field recordings that can be found in [32]. Each node is characterized by its geographic coordinates (lat_i, lon_i) and a dynamic health state $H_i(t) \in \{0, 1\}$, where there are 2 distinct states, NEGATIVO for negative and POSITIVO for positive result samples, respectively. Given that Xf infection is chronic and currently incurable in the study area, the model follows $N \rightarrow P$ transition dynamics without recovery. As Figure 2 shows the a zoomed-in view of the selected area and represents the selected trees as the green points. The purple squares represent the conceptual input points (a1, b1, a2, b2), and the orange circles show the calculated bounding box corners (z1, z2). The resulting graph is treated as an undirected and unweighted proximity graph, where edges merely signify a potential pathway for transmission.

Edges (yellow lines in Figure 3) represent potential transmission pathways mediated by the insect vector, while (red dots also in Figure 3) show individual olive trees. To capture the multi-scale nature of the spread, we utilize a literature-based threshold approach [3,7,24], where short-range edges ($d_{ij} < 35$ m): Representing canopy-to-canopy transmission and local vector hopping, consistent with the median daily dispersal distance observed in field studies. In addition, an edge e_{ij} is established between nodes i and j if their Euclidean distance d_{ij} satisfies $d_{ij} \leq d_{max}$, where $d_{max} = 2.5$ km represents the maximum dispersal range under wind-assisted conditions [33].

The resulting spatial network structure, overlaid on high-resolution satellite imagery, is illustrated in Figure 3. The hierarchical visualization distinguishes between dense local clusters and the broader connectivity of the landscape.

Graph-based Transmission Network Structure
Maximum Interaction Distance: 2500m



Figure 3. Representative proximity graph of the study area overlaid on high-resolution satellite imagery (Esri World Imagery [34]), note that the points correspond to the vertices of the graph and the yellow lines correspond to the edges of the graph.

3.3. Network Properties and Metrics

The connectivity of the graph is formally represented by the adjacency matrix A , where $A_{ij} = 1$ if $e_{ij} \in E$, and 0 otherwise. To characterize the structural complexity of the transmission network, we compute key topological metrics:

- **Degree centrality (k_i):** Represents the number of neighbors connected to node i within the d_{max} radius, serving as a primary indicator of local exposure risk.
- **Local clustering coefficient (C_i):** Quantifies the tendency of trees to form tightly connected groups (cliques), which facilitates rapid localized saturation.

These metrics are used to find “epidemiological hubs” well-linked trees that magnify and amplify the disease. The resulting degree distribution and spatial density of edges (visualized in Figure 3) determine the balance between localized, diffusive spread and long-range stochastic jumps. These properties are integrated into the simulation engine to calculate the probability of multi-step infection pathways across the landscape.

3.4. Infection Probability Model

The infection model we developed consists of three different components. The first describes the spatial structure, the second the seasonality, and the third presents the stochastic propagation mechanism, which combines them. The distance buckets are grounded in the distinct dispersal modes of the vector *Philaenus spumarius*.

3.4.1. Spatial Kernel and Transmission Buckets

Furthermore, we use a multi-level decay function $K(d_{ij})$ based on Haversine distances to describe the movement range of the vector, which we call buckets. This formulation offers a substantial benefit over binary or uniform transmission models; it notably enables the simulation to mimic both rapid neighborhood saturation and the infrequent, long-distance ‘leaps’ that initiate fresh outbreaks. This piecewise function is a semi-quantitative heuristic. The short-range bucket (<35 m) corresponds to the vector’s active flight capability within the canopy, based on findings presented in [7,33], designed to create a computationally efficient proxy for the complex, multi-modal dispersal of the vector. The long-range bucket (up to 2.5 km) captures wind-assisted passive dispersal (“leap-frogging”), consistent with EFSA reports [3]; here is how we define the kernel structure:

$$K(d) = \begin{cases} 1.0 & \text{if } d \leq 35 \text{ m (High risk, canopy-to-canopy)} \\ 0.15 & \text{if } 35 \text{ m} < d \leq 200 \text{ m (Local dispersal)} \\ 0.05 & \text{if } 200 \text{ m} < d \leq 1000 \text{ m (Medium range)} \\ 0.005 & \text{if } 1000 \text{ m} < d \leq 2500 \text{ m (Wind-assisted long range)} \end{cases} \quad (1)$$

3.4.2. Modeling Seasonal Dynamics

The seasonal forcing term $S(t)$ is empirically informed by integrating independent datasets on vector flight performance [33] and bacterial titer dynamics [35]. Although the mathematical combination (the product of these factors) is phenomenological, the resulting peak transmission window (June–July) shows high temporal congruence with field-observed infection surges in Southern Italy. To assess the robustness of $S(t)$, we performed a sensitivity test by shifting the seasonal peak by ± 30 days; the results showed that while the timing of the epidemic curve shifts, the final spatial distribution of high-risk nodes remains consistent, indicating that the structural connectivity of the host network is the primary driver of the spatial pattern, while $S(t)$ acts as a necessary temporal modulator.

To provide a precise mathematical description as required for stochastic simulation, $S(t)$ is defined as a periodic, piecewise constant function (step function) with a period of $T = 365$ days. For any given time step t (day), the value of $S(t)$ is uniquely determined by the calendar month m and is calculated as the product of two dimensionless normalized indices $[0, 1]$:

$$S(t) = B_{load}(m) \times V_{flight}(m), \quad \text{for } t \in \text{month } m \quad (2)$$

where $B_{load}(m)$ represents the Pathogen Availability Index and $V_{flight}(m)$ the Vector Mobility Index. This formulation ensures that $S(t)$ maintains its periodic behavior ($S(t) = S(t + 365)$) while correctly modulating the instantaneous transmission rate, as we will discuss later as shown in Table 1.

The temporal evolution of $S(t)$ is illustrated in Figure 4. As shown, the transmission risk peaks in early summer (June, $S(t) = 0.72$) when bacterial titers reach maximum levels, while it remains moderate in autumn due to the resurgence of vector flight activity.

- **Vector Mobility (V_{flight}):** Based on flight-mill experiments [33], female spittlebugs exhibit peak performance in spring and autumn. Activity is reduced in July and August (0.50–0.60) due to high temperatures and estivation, which constrains the transmission potential during peak heat.
- **Pathogen Availability (B_{load}):** Bacterial concentration in leaf petioles varies by phenological stage [35]. Isolation success is negligible during spring vegetative growth, peaks during the hull split stage (80–95%), and declines during senescence (10%).

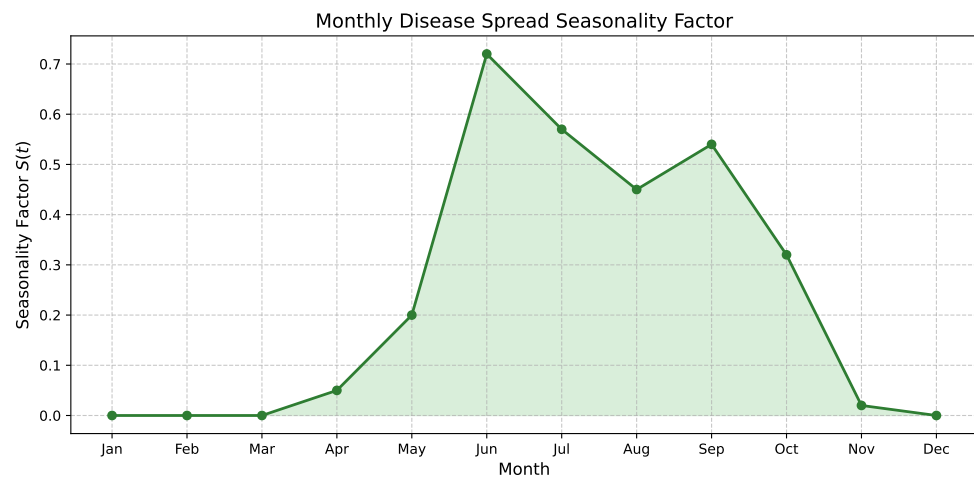


Figure 4. Monthly variation in the disease spread seasonality factor $S(t)$. The values reflect the synergy between pathogen availability in the host plant and vector flight performance throughout the year.

Table 1. Parameterization of the seasonality factor $S(t)$ based on vector flight performance and pathogen availability in the host plant.

Month	Phenological Stage	Pathogen Availability (B_{load})	Vector Flight (V_{flight})	Seasonality Factor ($S(t)$)
January–March	Dormancy/Nymphs	0.00	0.00	0.00
April	Vegetative Growth	0.05	1.00	0.05
May	Nut Growth	0.20	1.00	0.20
June	Hull Split (Start)	0.80	0.90	0.72
July	Hull Split (Peak)	0.95	0.60	0.57
August	Harvest	0.90	0.50	0.45
September	Leaf Fall	0.60	0.90	0.54
October	Late Leaf Fall	0.45	0.70	0.32
November	Senescence	0.10	0.20	0.02
December	Dormancy	0.00	0.00	0.00

3.4.3. Stochastic Propagation Model

The transmission of the pathogen from an infected node i to a susceptible neighbor j is modeled as a stochastic Poisson process operating on the proximity graph, which is a standard approach for event-based phenomena in epidemiology. The instantaneous rate of transmission, $\lambda_{ij}(t)$, from i to j at time t is defined as the product of three key components:

$$\lambda_{ij}(t) = \beta \cdot K(d_{ij}) \cdot S(t) \quad (3)$$

where β is the baseline transmission rate (a calibrated parameter representing the intrinsic infectivity), $K(d_{ij})$ is the dimensionless spatial kernel that accounts for distance (i.e., distance between i and j), and $S(t)$ is the dimensionless seasonality factor.

Given this rate, the probability $P_{ij}(t, \Delta t)$ that node j becomes infected by node i during a discrete time interval Δt (in our simulation, $\Delta t = 1$ day) is derived from the cumulative distribution function of the exponential distribution:

$$P_{ij}(t, \Delta t) = 1 - \exp(-\lambda_{ij}(t) \cdot \Delta t) \quad (4)$$

This formulation correctly ensures that the probability is bounded between 0 and 1. During the simulation, for each susceptible neighbor j of an infected node i , a random number $r \sim U(0, 1)$ is drawn at each daily time step. Infection occurs if $r < P_{ij}(t, \Delta t)$.

3.5. Simulation Framework

The simulation is initialized with a single infection source. Based on the official monitoring records from the Apulia region [32], Node 72 is identified as the earliest confirmed case in the study area and is therefore set as “Patient Zero” at $t = 0$ (21 March 2017) to trigger the epidemic process. The simulation then proceeds in discrete daily time steps ($\Delta t = 1$ day). For each day, the algorithm iterates through all currently infected nodes. For each susceptible neighbor of an infected node, a stochastic event is triggered based on the infection probability $P_{ij}(t, \Delta t)$ defined in Equation (4). A random number $r \sim U(0, 1)$ is generated, and if $r < P_{ij}(t, \Delta t)$, the neighboring node transitions to the infected state for the next time step.

Given the model’s stochasticity, a single run is insufficient to capture the complete risk landscape. We therefore employ a Monte Carlo approach, executing $N = 1000$ independent simulations for each parameter set. This allows us to calculate the *Infection Probability* for each node v :

$$P_{inf}(v) = \frac{\sum_{k=1}^N \mathbb{I}_k(v)}{N} \quad (5)$$

where $\mathbb{I}_k(v)$ is an indicator function that is 1 if node v is infected in simulation run k . Each run terminates either when $t = 365$ days or when 100% of nodes are infected. The date of infection for every node is logged for subsequent spatiotemporal analysis. In Figure 5, the methodological workflow of the system is illustrated.

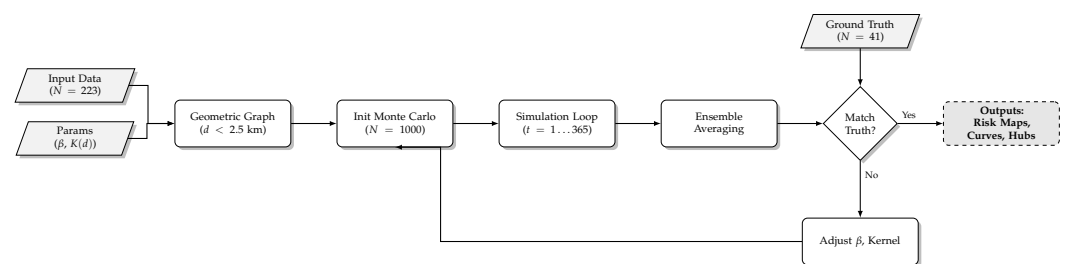


Figure 5. Methodological workflow of the simulation engine.

3.6. Identifying Critical Nodes and Super-Spreaders

While static network metrics like topological betweenness centrality can identify structural bridges, the dynamic nature of disease spread requires a more functional approach. In this study, we use a “Functional Centrality” derived directly from the simulation outcomes.

Nodes that are infected early and frequently across the Monte Carlo runs are identified as critical transmission hubs, or “Super-Spreaders”. These nodes act as crucial bridges, facilitating the jump of the pathogen between distinct spatial clusters of trees. Identifying these hubs allows for targeted control strategies. Prioritizing the treatment or removal of these high-centrality nodes can theoretically fragment the transmission network far more effectively than random interventions.

3.7. Model Calibration and Validation

To calibrate the model and evaluate its predictive accuracy, we validate it against a ground-truth set of 41 confirmed cases from late 2017, with the goal of identifying parameters that reproduce this specific spatial pattern. We utilize the Jaccard Similarity Index (J) to quantify the spatial overlap between the simulated outbreak area (S_{sim}) and the ground-truth infection cluster (S_{real}), and the Count Error ($\Delta N = N_{sim} - N_{real}$) to assess systematic over- or under-estimation of the epidemic’s size. It is important to note that, due to the limited availability of high-resolution datasets with comparable monitoring efforts, the use of a single regional dataset for both calibration and validation, differentiated only by a temporal split, which does not fully attest the model’s predictive power. Thus,

the current validation should therefore be interpreted as a measure of the model's ability to self-consistently reproduce the observed epidemic dynamics, rather than as a proof of its predictive accuracy in a different context. However, it is not excluded that, when calibrated on a particular olive orchard, the proposed framework provides an estimation of the dynamics of the spread based on the positioning of the trees inside the olive orchard.

To find the optimal parameter configuration, a systematic sensitivity analysis was performed using a Grid Search to explore the parameter space of the transmission rate (β) and the kernel dispersal multiplier (K). The parameter space was defined as follows:

- **Transmission Rate (β):** Tested values ranged from 0.05 (low-intensity, cryptic spread) to 0.8 (rapid outbreak), to identify the threshold between silent spread and network saturation.
- **Kernel Multiplier (K):** Tested values ranged from 1.0 to 3.0, assessing the impact of extending the vector's dispersal range beyond the baseline distance decay probabilities.

For each combination of (β, K) , the simulation was executed for $N = 50$ iterations to smooth out stochastic variations. The model's performance was evaluated using the Jaccard Similarity Index (J), which measures the spatial congruence between the simulated infected set (S_{sim}) and the ground-truth infected set (S_{real}):

$$J(\beta, K) = \frac{|S_{sim} \cap S_{real}|}{|S_{sim} \cup S_{real}|} \quad (6)$$

Figure 6 presents the results of our parameter tuning process through a systematic grid search. The heatmaps visualize the model's performance, measured by the Jaccard Similarity Index, across a range of values for the transmission rate (β) and the kernel multiplier (K).

Two distinct validation scenarios were evaluated:

- **Scenario A (Figure 6a):** This panel shows the model's performance against the ground-truth data from only the first 30 days of the outbreak. It tests the model's ability to capture the initial, localized spread.
- **Scenario B (Figure 6b):** This panel validates the model against the complete, year-end dataset of infections. It serves as the primary benchmark for the model's ability to replicate the overall epidemic progression.

From these results, we can draw a key conclusion for selecting the optimal parameters. In both scenarios, we observe that for high values of the transmission rate ($\beta \geq 0.2$), the Jaccard index plateaus to a uniform score. This indicates a "saturation regime," where the transmission is so high that the model unrealistically infects most of the network, losing its ability to discriminate between high and low-risk areas.

A crucial insight from this analysis is the trade-off between maximizing the Jaccard score and ensuring model stability. While the absolute highest score might occur at slightly higher values, those regions are on the edge of the saturation regime, where the model is over-fitted. Therefore, we selected our optimal parameters not by simply picking the single "brightest" cell, but by identifying a region that offers a high Jaccard score while remaining safely below the saturation threshold. The combination of $\beta = 0.1$ and $K = 1.5$ represents this optimal balance. It achieves a near-maximal Jaccard score in the long-term scenario (Scenario B) while ensuring the model produces selective, structured outbreaks, thus providing more robust and credible simulation dynamics. The model correctly identifies nodes that will become infected later in the year, which are not yet positive in the short-term dataset, thus demonstrating its utility as an early warning tool that can be used beyond the immediate visible spread. This is further discussed in the validation results in Section 4.2.

Finally, the use of a single regional dataset for both calibration and validation was deemed the most rigorous given the rarity of high-resolution datasets containing documented N-to-P transitions. By validating the model's performance across two different time horizons (30-day vs. Full Year), we demonstrate strong internal consistency and the model's ability to replicate the complex local transmission dynamics. The robustness of the subsequent centrality-based conclusions is supported by the high Jaccard scores achieved in the long-term scenario ($J = 0.659$), confirming the model's reliability as a decision-support tool for prioritizing interventions in high-density host landscapes where similar vector-host interactions occur.

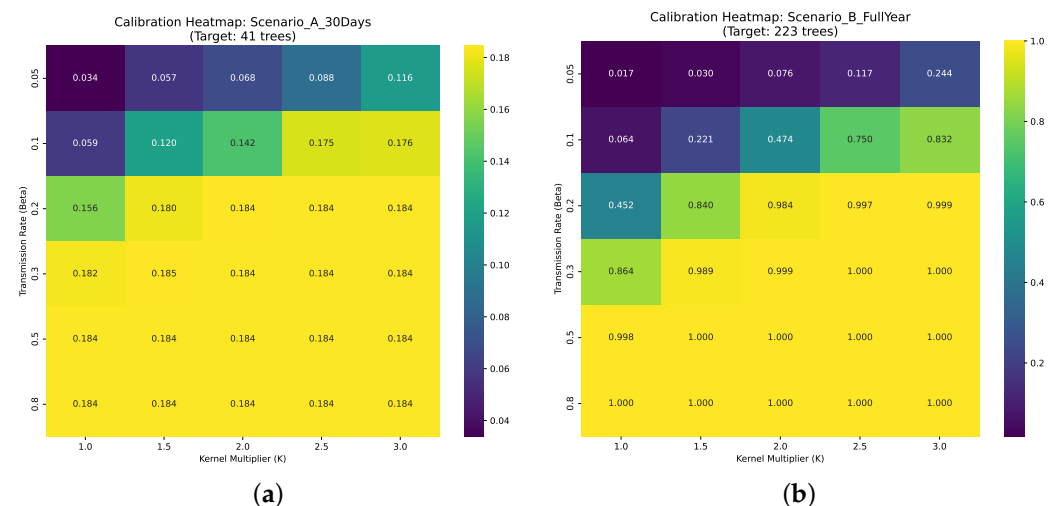


Figure 6. Sensitivity analysis heatmaps showing the Jaccard Similarity Index for varying Transmission Rates (β) and Kernel Multipliers (K) on Scenario A (a) and Scenario B (b).

4. Experimental Evaluation

The geometric graph construction resulted in a spatial network comprising $N = 223$ nodes. Using a cutoff distance of $d_{max} = 2.5$ km, the resulting graph is highly connected, reflecting the dense planting patterns typical of olive monocultures in Apulia. The connectivity analysis indicates that the network forms a single giant component, implying that without intervention, a pathogen introduced at any node has the theoretical topological path to infect the entire orchard.

4.1. Spatial Spread Dynamics

The simplified model was developed as a baseline for comparison, simulating a Susceptible-Infected (SI) process using a Monte Carlo method over 1000 iterations. This baseline assumes a constant infection probability ($p = 0.001$) for all susceptible hosts within a fixed radius of 400 m from an infectious source, without incorporating seasonal or other environmental factors.

The results in Figure 7 illustrate that the simplified distance-only model, while following a general logistic growth pattern, significantly deviates from the empirical data, predicting a more rapid initial outbreak. In contrast, the proposed simulation model demonstrates a much closer correspondence with the real-world progression of the epidemic. This comparison highlights the critical importance of the nuanced parameters integrated into our proposed framework such as the distance-decay kernel and seasonal variability in capturing accurately the complex spatio-temporal dynamics of the disease spread.

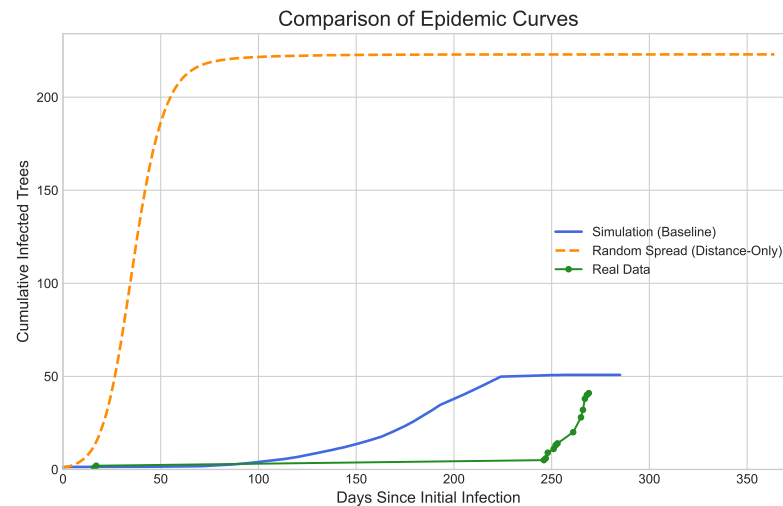


Figure 7. Comparative analysis of epidemic progression curves. The figure presents the cumulative number of infected trees over a 365-day period from three distinct sources: (i) the proposed epidemiological model (Simulation—Baseline), (ii) a simplified distance-only model (Random Spread), and (iii) field observations (Real Data).

Figure 8 illustrates the temporal evolution of the epidemic over the simulated annual cycle ($T = 365$ days) using the calibrated parameters ($\beta = 0.1$, $K = 1.5$). The model was executed for $N = 1000$ Monte Carlo iterations to ensure statistical robustness. The cumulative infection curve exhibits a sigmoidal but flattened trajectory. Unlike explosive outbreaks often modeled with high β , the simulation reveals a “Silent Spread” dynamic. The infection progresses slowly during the spring months, accelerates moderately during the peak vector activity season (Summer, Days 200–300), and plateaus at approximately 52 infected trees (mean value). This pattern suggests that while the disease is persistent, its effective transmission rate is constrained by environmental factors, preventing the rapid saturation of the network.

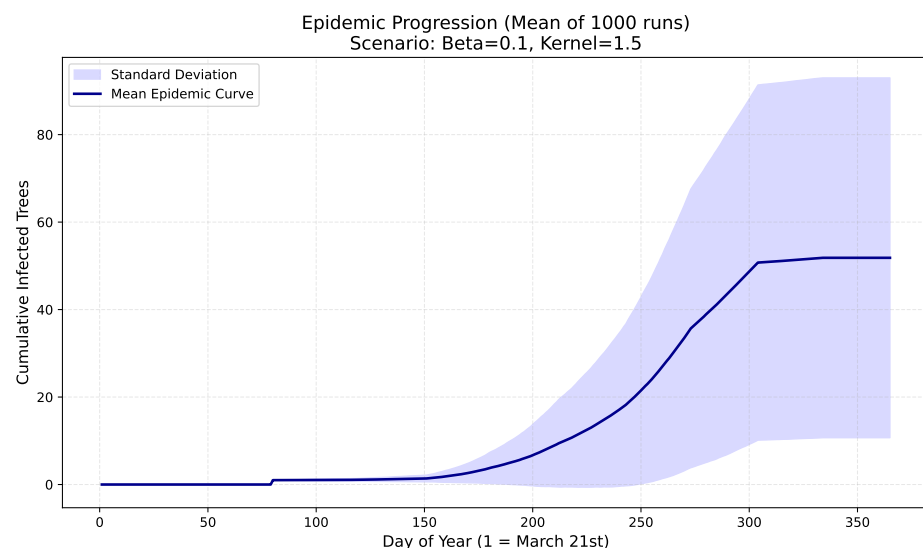


Figure 8. Cumulative epidemic curve of X_f over one year, based on the mean of 1000 Monte Carlo simulations. The shaded region represents the standard deviation across all runs, illustrating the stochastic variability in the epidemic’s progression. The curve plateaus at approximately 52 infected trees, confirming the “Silent Spread” dynamic.

4.2. Model Validation

The Monte Carlo simulations ($N = 1000$ iterations) provided a probabilistic risk assessment for each tree. Instead of a binary outcome, the model output defines an infection probability map. As shown in the probabilistic risk map (Figure 9, Right Panel), the risk is not uniformly distributed. The simulation effectively filters out stochastic noise: approximately 80% of the orchard exhibits a low infection probability ($P_{inf} < 20\%$), indicating that transmission chains in these areas frequently go extinct. Conversely, distinct high-risk clusters emerge in the central and northwestern sectors, where P_{inf} consistently exceeds 30%, marking these areas as established infection hotspots.

A deeper analysis of Figure 9 reveals several key insights. The left panel shows the ground truth, the actual locations of the 41 trees confirmed positive in 2017. These infections are not randomly scattered but are concentrated in specific geographic clusters. The right panel, our simulated risk map, shows the model's output as the probability of infection for every tree in the network after 1000 simulations. The color gradient, from blue (low risk) to red (high risk), visualizes this probability.

The crucial finding is the strong spatial correspondence between the two panels. The areas colored yellow, orange, and red in the simulated map, representing probabilities from 30% to over 60%, align almost perfectly with the locations of the true positive cases. This demonstrates that the model, driven by the unweighted undirected graph and the seasonal transmission kernel, successfully learned and reproduced the underlying factors that made these specific areas vulnerable to infection. It did not simply predict a random spread but identified the same corridors and clusters of infection that occurred in reality. The model's ability to recreate these hotspots, even when initialized from a single “patient zero”, underscores its capacity to capture the essential dynamics of the local spread. The model's validity was assessed by comparing the simulation outputs against the ground-truth dataset of 41 trees confirmed to be infected in late 2017.

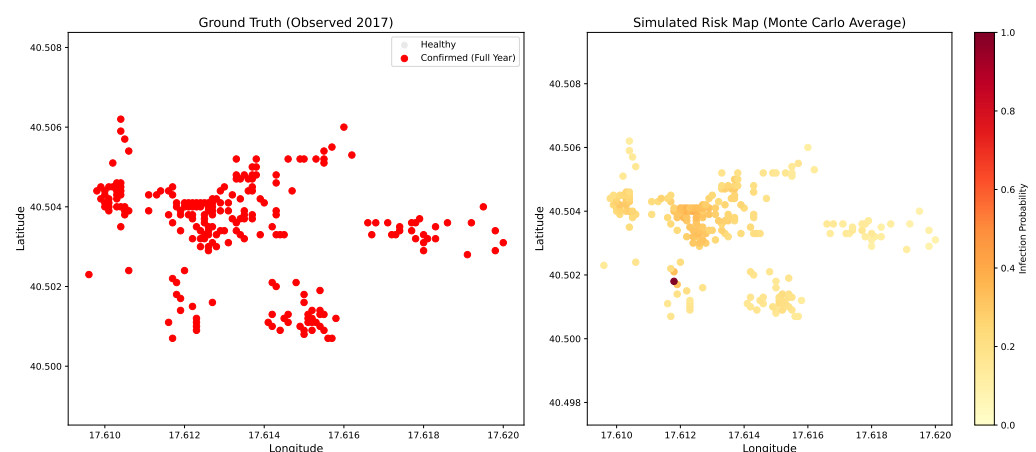


Figure 9. Spatial model validation.

Spatial Validation: Figure 9 presents the side-by-side comparison of the actual infection status (Left) and the simulated risk (Right). The model successfully reproduces the spatial distribution of the real cases. The high-risk clusters predicted by the model spatially coincide with the locations of the confirmed infected trees. The left subfigure refers to the Ground Truth with red dots indicating the confirmed cases, while on the right (Simulated Risk Map), the infection probability for each is represented with gradient color. The model correctly identifies the location of the main infection clusters.

Quantitative Metrics: The calibration performance was quantified using the Jaccard Similarity Index against two validation sets (Figure 10). The model achieved a high

Jaccard score of **0.659** when compared against the full-year 2017 dataset, confirming strong agreement between the predicted hotspots and the final observed state.

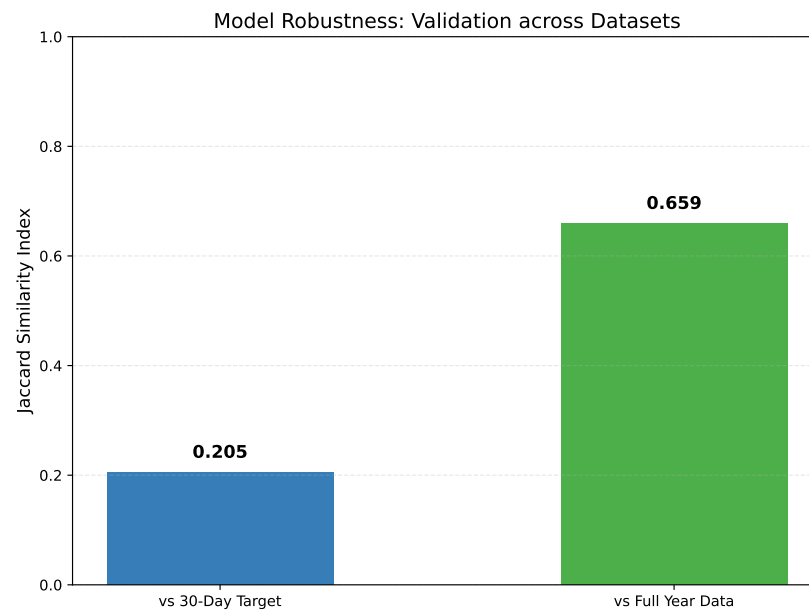


Figure 10. Validation Metrics across datasets.

Comparison against the initial 30-day target yielded a score of 0.205. This quantitative difference between the initial and final validation scores (0.205 vs. 0.659) reflects the temporal progression of the epidemic simulation over the annual cycle, capturing infections that developed after the first month of observation.

4.3. Centrality-Based Analysis

Through the analysis of recurrent infection patterns across the simulation runs, specific nodes were identified as functional hubs or “Super-Spreaders”. Nodes located in the high-density areas of the geometric graph (e.g., IDs 129, 178, and 180) demonstrated the highest frequency of acting as transmission bridges. These nodes facilitate the jump of the pathogen between the primary infection cluster and secondary zones. Their identification is crucial, as they represent the most effective targets for prioritized sampling or removal to break the transmission chain.

The analysis systematically varies the number of removed nodes, denoted by Δ (Delta), to observe the corresponding changes in epidemic outcomes. For each scenario, we present two key visualizations.

- **The Epidemic Curve:** Illustrates the number of newly infected nodes over time, providing insight into the speed and magnitude of the spread.
- **The Geographic Risk Map:** Displays the spatial distribution of infection risk, highlighting high-risk and low-risk zones post-intervention.

This first strategy focuses on removing nodes with the highest Betweenness Centrality scores. The second strategy implements an alternative approach by removing nodes with the highest Closeness Centrality. The results of each strategy are depicted in Figures 11 and 12 deploying the first strategy and in Figures 13 and 14 deploying the second strategy, where in both cases it is investigated the impact of key node removal (Δ) from the baseline (no removal) to the removal of 175 key nodes.

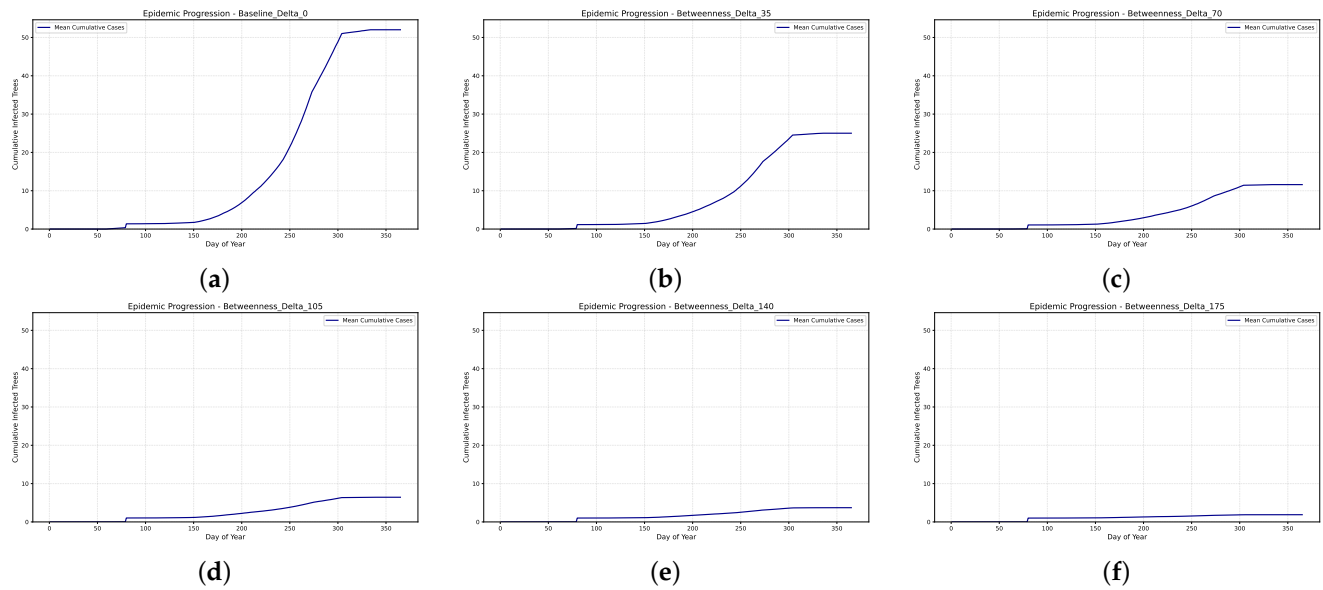


Figure 11. Epidemic Curves for the Betweenness Centrality strategy at varying levels of node removal (Δ), $\Delta = 0$ (a), $\Delta = 0.35$ (b), $\Delta = 0.70$ (c), $\Delta = 105$ (d), $\Delta = 140$ (e), and $\Delta = 175$ (f).

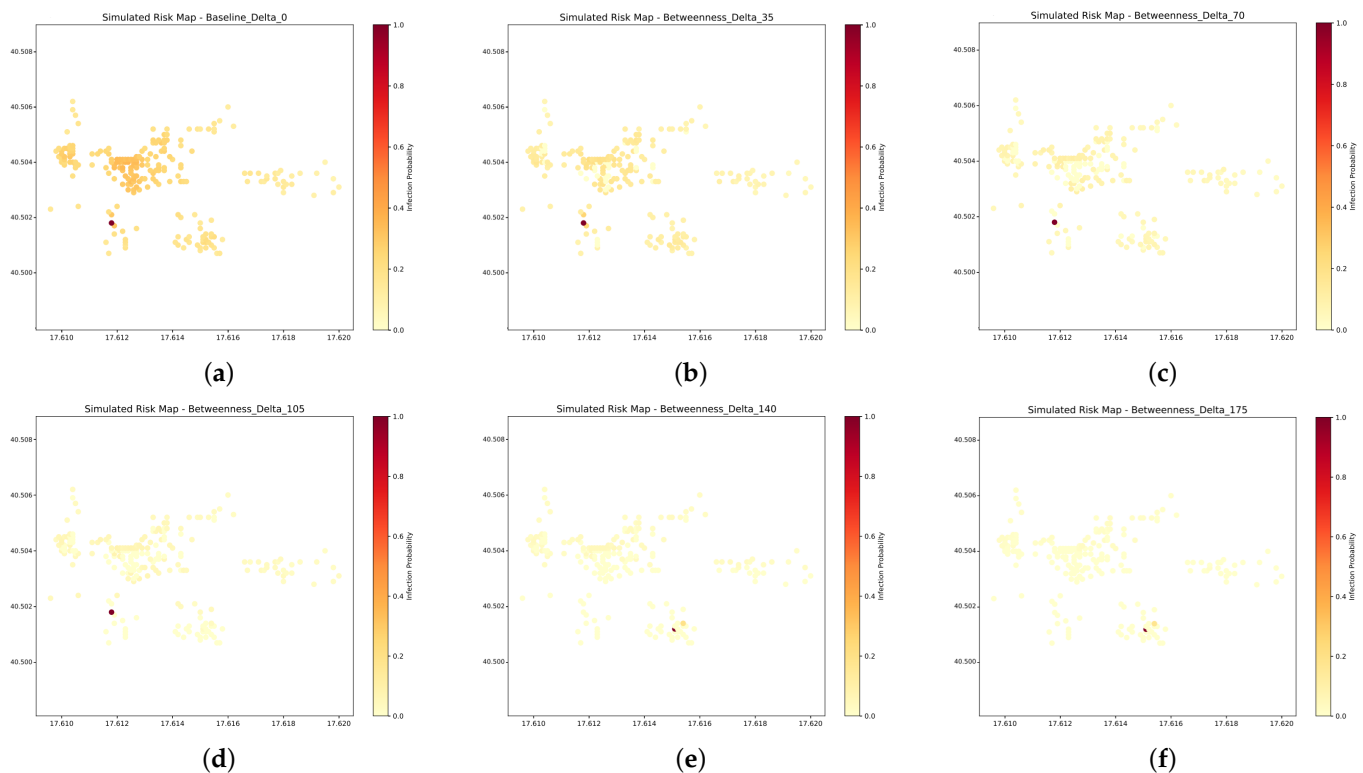


Figure 12. Geographic Risk Maps for the Betweenness Centrality strategy at varying levels of node removal (Δ), $\Delta = 0$ (a), $\Delta = 0.35$ (b), $\Delta = 0.70$ (c), $\Delta = 105$ (d), $\Delta = 140$ (e), and $\Delta = 175$ (f).

Comparative Assessment

A direct comparison of the two strategies reveals that targeting nodes by Betweenness Centrality offers a more effective mitigation outcome. As shown by contrasting Figure 11 with Figure 13, the reduction in the epidemic's final size is consistently greater under the Betweenness strategy for any given number of removed nodes (Δ). This finding is spatially corroborated by the risk maps (Figures 12 and 14), where the fragmentation of high-risk clusters is more pronounced when removing high-betweenness nodes.

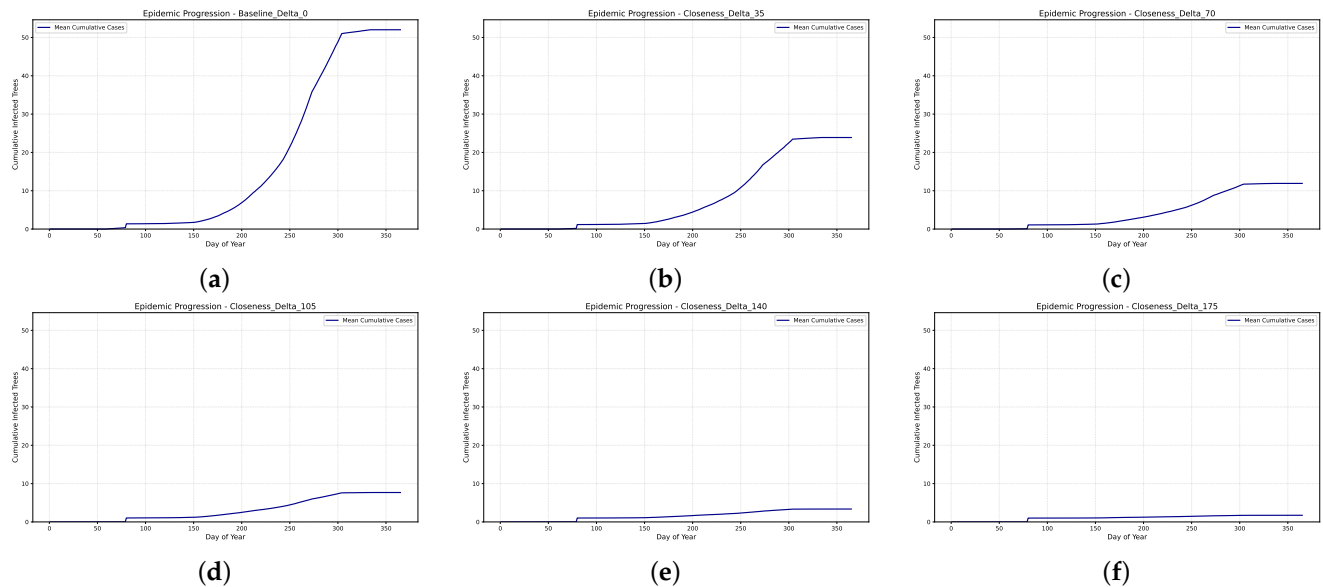


Figure 13. Epidemic Curves for the Closeness Centrality strategy at varying levels of node removal (Δ), $\Delta = 0$ (a), $\Delta = 0.35$ (b), $\Delta = 0.70$ (c), $\Delta = 105$ (d), $\Delta = 140$ (e), and $\Delta = 175$ (f).

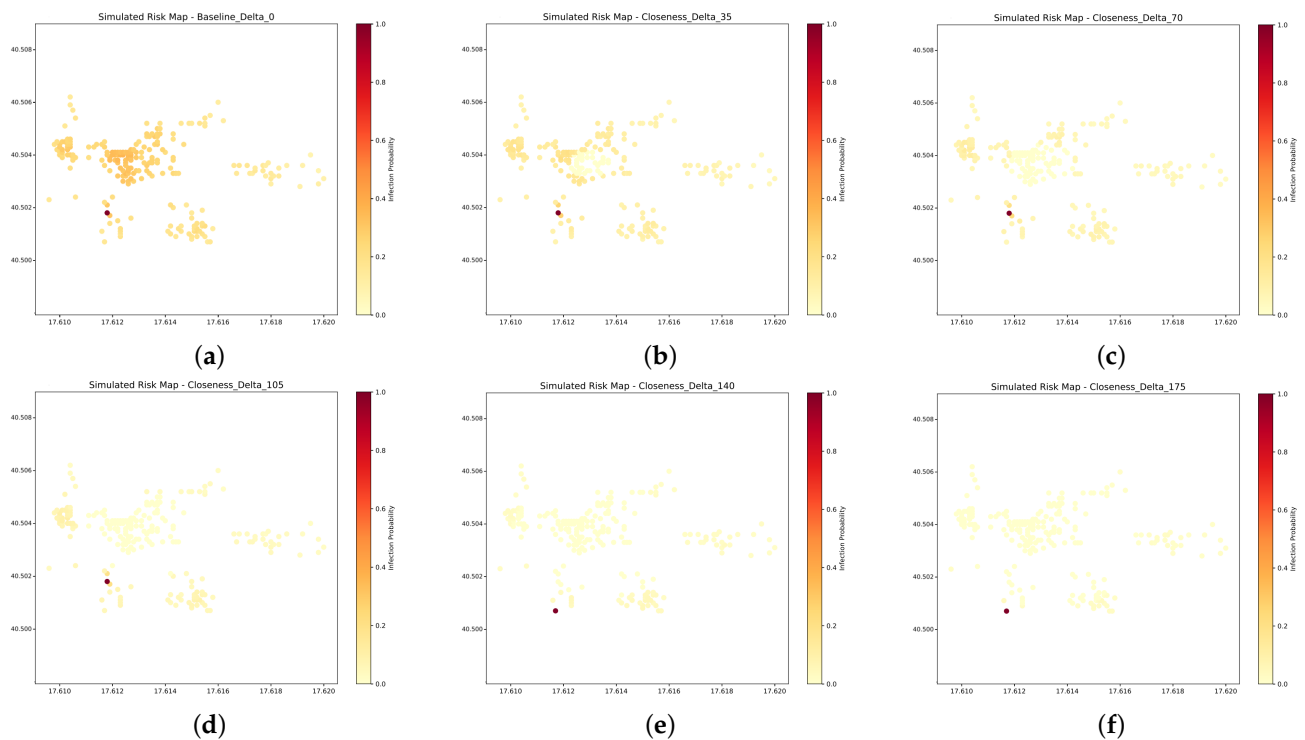


Figure 14. Geographic Risk Maps for the Closeness Centrality strategy at varying levels of node removal (Δ), $\Delta = 0$ (a), $\Delta = 0.35$ (b), $\Delta = 0.70$ (c), $\Delta = 105$ (d), $\Delta = 140$ (e), and $\Delta = 175$ (f).

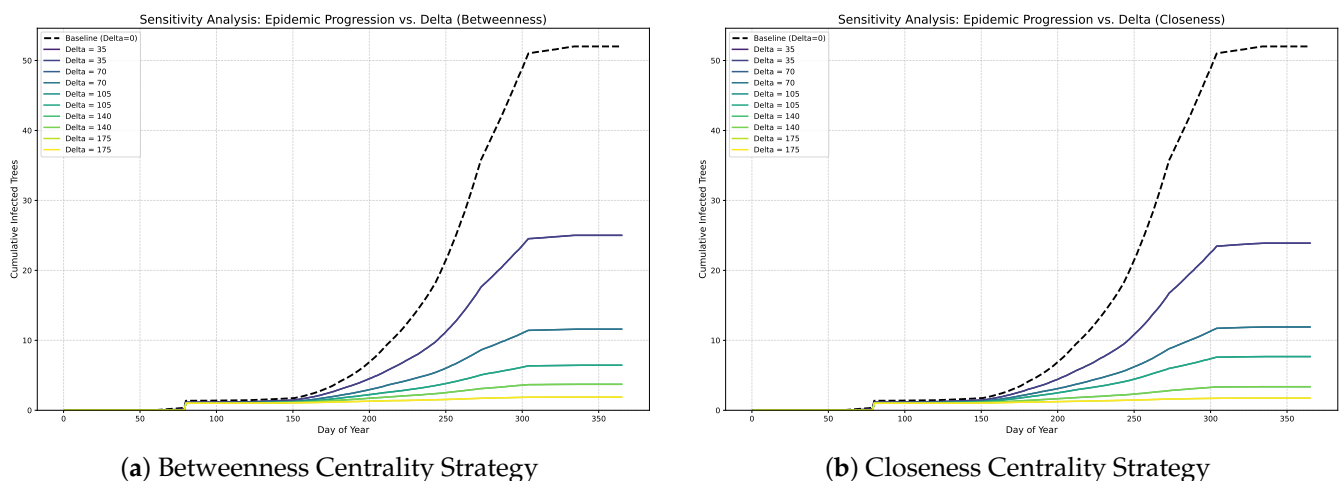
Concluding, cumulative epidemic figures have been produced to summarize the simulated mitigation results in the comparative plots in Figure 15. These plots show the average final size of the epidemic as a function of the number of removed nodes (Δ) for each strategy.

Table 2 provides a quantitative evaluation of the two targeted node removal strategies, which are based on betweenness and closeness centrality. The results reveal that both methods achieve a substantial reduction in the final epidemic size compared to the baseline scenario. A clear dose–response relationship is observed, where the efficacy of both strategies scales with the number of removed nodes (Δ).

Table 2. Quantitative Comparison of Mitigation Strategies.

Removed Nodes (Δ)	Betweenness Centrality		Closeness Centrality	
	Final Size (Mean)	Reduction (%)	Final Size (Mean)	Reduction (%)
0 (Baseline)	52.012	0.0%	52.012	0.0%
35	25.011	51.91%	23.889	54.07%
70	11.596	77.70%	11.909	77.10%
105	6.442	87.60%	7.678	85.24%
140	3.715	92.86%	3.369	93.52%
175	1.885	96.38%	1.736	96.66%

A comparative analysis of the two strategies uncovers a non-linear relationship in their relative performance. For a low level of intervention ($\Delta = 35$), the closeness-based strategy is marginally more effective, yielding a 54.07% reduction compared to 51.91% for the betweenness-based strategy. Conversely, at intermediate intervention levels ($\Delta = 70$ and $\Delta = 105$), the betweenness-based strategy exhibits superior performance, achieving a reduction of 87.60% for $\Delta = 105$. Notably, at high levels of node removal ($\Delta = 140$ and $\Delta = 175$), the closeness-based strategy regains a slight advantage, culminating in the highest overall performance with a 96.66% reduction at the maximum intervention level. In summary, both approaches demonstrate the remarkable potential of centrality-based measures for epidemic control, as they are capable of reducing the final epidemic size by over 96% at the highest levels of intervention studied.

**Figure 15.** Comparative performance of mitigation strategies. Both plots show the final epidemic size as a function of the number of removed nodes (Δ), illustrating diminishing returns.

5. Conclusions

This study has demonstrated that the Xf spreading in olive orchards follows a dual dispersal mechanism. Initially, the pathogen spreads locally through short-range, tree-to-tree interactions, leading to a localized saturation. However, as the epidemic progresses, it transitions into a regional outbreak driven by long-distance “jumps”. These long-distance dispersals, facilitated by key nodes in the proximity graph, help to expand the epidemic beyond the initially affected area. This combination of local and regional dynamics highlights the complexity of pathogen spread in olive orchards. Furthermore, the validation results support this, with a moderate Jaccard index of 0.205 in the short-term scenario, which improves to 0.659 in the long-term. This suggests that the model can replicate the cryptic spread of the pathogen, confirming its potential as a risk-assessment and early warning tool to identify areas vulnerable to future outbreaks.

One of the main contributions of this study is the use of graph centrality metrics to identify “super spreaders”. Our analysis shows that a small group of strategically placed trees plays a key role in disease transmission. These high-centrality nodes act as amplifiers, facilitating the spread of *Xf* across the orchard. Targeting these trees for early intervention or surveillance could be a more effective and cost-efficient strategy compared to traditional area-wide approaches, such as buffer zones. This targeted approach not only reduces the socio-economic impact of large-scale interventions but also helps to contain the disease more effectively in high-risk areas.

5.1. Temporal Validation Analysis

A critical insight from the dual validation process is the divergence between the short-term (30-day) and long-term (Full Year) similarity scores. The moderate Jaccard index of 0.205 against the 30-day target is not an indicator of poor performance, but rather a reflection of the model’s *predictive horizon*.

Since the simulation projects the risk over a full annual cycle, it identifies trees that are currently asymptomatic or healthy in the initial 30-day window but are highly probable to become infected later in the season. When validated against the full-year dataset, the score rises significantly to 0.659. This trajectory confirms that the model correctly anticipated the latent spread of the pathogen that was not yet visible in the first month of observation, thereby validating its utility as an early warning tool for future infection risks.

5.2. Centrality as a Decision-Support Metric

A key contribution of this study is the application of graph centrality metrics to identify “epidemiological hubs”. Our analysis reveals that a small subset of trees, due to their strategic geographic position, acts as a primary conduit for disease amplification.

Prioritizing these high-centrality nodes for intensive surveillance or early-stage interventions could theoretically fragment the transmission network more effectively than random culling. This transition from area-wide management to centrality-based control could mitigate the loss of cultural heritage while maintaining high containment efficacy in high-risk zones.

Observing Section 4.3, several key conclusions can be drawn. First, both Betweenness and Closeness Centrality are viable and effective metrics for identifying critical nodes for epidemic mitigation. Second, the relationship between the number of removed nodes and the reduction in spread is non-linear, exhibiting diminishing returns. This observation is critical for policy-making, as it suggests that an optimal, cost-effective intervention point exists. Concluding, for this specific network, both strategies perform comparably, with neither showing a consistent and significant advantage over the other across all intervention levels.

5.3. Parameter Interpretation

A key aspect of our modeling approach is the use of a distance-based, multi-level transmission kernel. The specific distance buckets (0–35 m, 35–200 m, etc.) and their corresponding transmission multipliers are intended to serve as proxies for different modes of vector dispersal. The shortest range ($d \leq 35$ m) represents the high-probability, canopy-to-canopy spread and local hopping within a dense orchard block. The intermediate ranges reflect the typical daily dispersal behavior of *P. spumarius*, while the longest range ($d > 1000$ m) accounts for rare, wind-assisted or accidental long-distance jumps that are crucial for establishing new satellite infection clusters.

Furthermore, to fully quantify the added value of our structured approach, a comparison against simpler baseline models would be beneficial. For instance, a null model assuming random spread (where any node can infect any other with uniform probability, regardless of

distance) would likely fail to reproduce the observed spatial clustering. A more informative baseline would be a simple distance-dependent SI model, where the transmission probability is a continuous, decaying function of distance (e.g., an exponential or power-law kernel) without the seasonal component $S(t)$ or the multi-level bucket structure. Comparing our model's validation score (Jaccard index) against such a baseline would explicitly demonstrate the predictive gain obtained from incorporating vector seasonality and a multi-modal dispersal kernel. While this quantitative comparison is beyond the scope of the current manuscript, it remains an important direction for future work to rigorously establish the value of each model component.

5.4. Comparison with Other Studies

In contrast to continuous reaction-diffusion models [14,18], our geometric graph method offers the essential spatial resolution needed to assess risk at the level of individual trees. Although statistical models like those of Cendoya et al. [23] effectively identify major environmental influences, our mechanistic simulation enables hypothetical scenario exploration in regions before epidemics occur. Incorporating pathogen load dynamics, as proposed by Chatterjee et al. (2008) [8], guarantees that the simulation aligns with the true biological context of the Xf infection cycle, offering a more accurate early warning tool compared to solely topological models. We must not overlook that traditional containment strategies, such as uniform buffer zones, are often criticized for their high socio-economic costs [16], our framework offers a “precision-targeting” alternative.

5.5. Model Sensitivity

A key aspect of our modeling approach is the use of a distance-based, multi-level transmission kernel. The specific distance buckets (0–35 m, 35–200 m, etc.) and their corresponding transmission multipliers are intended to serve as proxies for different modes of vector dispersal. The shortest range ($d \leq 35$ m) represents the high-probability, canopy-to-canopy spread and local hopping within a dense orchard block. The intermediate ranges reflect the typical daily dispersal behavior of *P. spumarius*, while the longest range ($d > 1000$ m) accounts for rare, wind-assisted or accidental long-distance jumps that are crucial for establishing new satellite infection clusters. The Grid Search presented in Section 3.7 provides a preliminary sensitivity analysis for the overall transmission rate (β) and kernel multiplier (K), revealing that the model is highly sensitive to β , with a clear transition to network saturation. The model is less sensitive to the kernel multiplier K in the tested range, suggesting the fundamental bucket structure is a primary driver of the spatial pattern. Finally, to fully quantify the added value of our approach, a comparison against a simpler baseline (i.e., the basic approach that assumes random spread where any node can infect any other with uniform probability using also standard distance over SI epidemic model properties) failed to reproduce the observed spatial clustering.

5.6. Future Research Directions

Despite the promising results, there are limitations to the model. It assumes a static host landscape and does not consider real-time changes in tree health or human-mediated actions, like tree removal. Additionally, while the model includes seasonal dynamics, it does not account for microclimatic factors, such as extreme weather events, which can impact vector behavior and pathogen transmission. The model also assumes that long-distance dispersal is limited to 2500 m, but human-mediated movement of infected plant material could extend this range, making the spread even more unpredictable. Future improvements could focus on incorporating these factors to make the model more accurate and reflective of real-world dynamics.

For future work, it is important to address several limitations of the current model. Incorporating almost real-time data on tree health, as we developed in a previous work [36], and tree removal would make the model more dynamic and responsive. Additionally, including microclimatic variables and accounting for extreme weather events could improve the model's accuracy in predicting vector behavior. Additionally, the model could be used as an inverted tool to calculate the optimal distance to plant new crops so that they maintain a low probability of spread, or, if the crop already exists, suggest which trees to cut down. Finally, since the model has been calibrated, it is emphasized that this novel approach can be used in other geographical areas in the future.

Author Contributions: Conceptualization, N.M.P., I.P., I.K. and M.A.; Methodology, N.M.P., I.P., I.K. and M.A.; Software, N.M.P., I.P., C.S. and I.K.; Validation, N.M.P., I.P., C.S. and I.K.; Formal analysis, N.M.P., I.P., I.K. and M.A.; Investigation, N.M.P., I.P., C.S., I.K. and M.A.; Resources, C.S.; Data curation, N.M.P., I.P. and I.K.; Writing—original draft, N.M.P. and M.A.; Writing—review & editing, N.M.P., I.P., C.S. and I.K.; Visualization, N.M.P. and I.P.; Supervision, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are openly available in the Emergenza Xylella: Dati Monitoraggio at http://www.emergenzaxylella.it/portal/portale_gestione_agricoltura/Download/mon_xf (accessed on 30 January 2026). The processed data in this study are openly available at GitHub in https://github.com/nickpol29/xylella_spreading (accessed on 30 January 2026).

Acknowledgments: We are deeply saddened by the sudden passing of our supervisor and co-author, Markos Avlonitis. We dedicate this work to him in recognition of his inspiring guidance and his pivotal role in conceiving the idea for this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wells, J.M.; Raju, B.C.; Hung, H.Y.; Weisburg, W.G.; Mandelco-Paul, L.; Brenner, D.J. *Xylella fastidiosa* gen. nov., sp. nov: Gram-Negative, Xylem-Limited, Fastidious Plant Bacteria Related to *Xanthomonas* spp. *Int. J. Syst. Evol. Microbiol.* **1987**, *37*, 136–143. [\[CrossRef\]](#)
2. Bansal, K.; Kumar, S.; Kaur, A.; Singh, A.; Patil, P.B. Deep phylo-taxono genomics reveals *Xylella* as a variant lineage of plant associated *Xanthomonas* and supports their taxonomic reunification along with *Stenotrophomonas* and *Pseudoxanthomonas*. *Genomics* **2021**, *113*, 3989–4003. [\[CrossRef\]](#) [\[PubMed\]](#)
3. EFSA Panel on Plant Health (PLH); Bragard, C.; Dehnen-Schmutz, K.; Di Serio, F.; Gonthier, P.; Jacques, M.A.; Jaques Miret, J.A.; Justesen, A.F.; MacLeod, A.; Magnusson, C.S.; et al. Update of the Scientific Opinion on the risks to plant health posed by *Xylella fastidiosa* in the EU territory. *EFSA J.* **2019**, *17*, e05665. [\[CrossRef\]](#)
4. EFSA; Gibin, D.; Gutierrez Linares, A.; Fasanelli, E.; Pasinato, L.; Delbianco, A. Update of the *Xylella* spp. host plant database—Systematic literature search up to 30 June 2023. *EFSA J.* **2023**, *21*, e8477. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Saponari, M.; Boscia, D.; Nigro, F.; Martelli, G.P. Identification of dna sequences related to *Xylella fastidiosa* in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (Southern Italy). *J. Plant Pathol.* **2013**, *95*, 668.
6. Cornara, D.; Cavalieri, V.; Dongiovanni, C.; Altamura, G.; Palmisano, F.; Bosco, D.; Porcelli, F.; Almeida, R.P.; Saponari, M. Transmission of *Xylella fastidiosa* to different crops by *Philaenus spumarius* (Hemiptera: Aphrophoridae) in Apulia, Italy. *J. Econ. Entomol.* **2017**, *110*, 863–871. [\[CrossRef\]](#)
7. Bodino, N.; Cavalieri, V.; Dongiovanni, C.; Simonetto, A.; Saladini, M.A.; Plazio, E.; Gilioli, G.; Molinatto, G.; Saponari, M.; Bosco, D. Dispersal of *Philaenus spumarius* (Hemiptera: Aphrophoridae), a Vector of *Xylella fastidiosa*, in Olive Grove and Meadow Agroecosystems. *Environ. Entomol.* **2020**, *50*, 267–279. [\[CrossRef\]](#)
8. Chatterjee, S.; Almeida, R.P.P.; Lindow, S. Living in two Worlds: The Plant and Insect Lifestyles of *Xylella fastidiosa*. *Annu. Rev. Phytopathol.* **2008**, *46*, 243–271. [\[CrossRef\]](#)
9. Salih, R.I.; Jawad, S.; Dehingia, K.; Das, A. The effect of a psychological scare on the dynamics of the tumor-immune interaction with optimal control strategy. *Int. J. Optim. Control Theor. Appl.* **2024**, *14*, 276–293. [\[CrossRef\]](#)

10. Loumou, A.; Giourga, C. Olive groves: “The life and identity of the Mediterranean”. *Agric. Hum. Values* **2003**, *20*, 87–95. [CrossRef]
11. Besnard, G.; Terral, J.F.; Cornille, A. On the origins and domestication of the olive: A review and perspectives. *Ann. Bot.* **2018**, *121*, 385–403. [CrossRef]
12. Kaniewski, D.; Campo, E.V.; Boiy, T.; Terral, J.F.; Khadari, B.; Besnard, G. Primary domestication and early uses of the emblematic olive tree: Palaeobotanical, historical and molecular evidence from the Middle East. *Biol. Rev.* **2012**, *87*, 885–899. [CrossRef]
13. Kottelenberg, D.; Hemerik, L.; Saponari, M.; van der Werf, W. Shape and rate of movement of the invasion front of *Xylella fastidiosa* spp. *pauca* in Puglia. *Sci. Rep.* **2021**, *11*, 1061. [CrossRef]
14. White, S.M.; Bullock, J.M.; Hooftman, D.A.; Chapman, D.S. Modelling the spread and control of *Xylella fastidiosa* in the early stages of invasion in Apulia, Italy. *Biol. Invasions* **2017**, *19*, 1825–1837. [CrossRef] [PubMed]
15. Serio, F.; Imbriani, G.; Girelli, C.R.; Miglietta, P.P.; Scortichini, M.; Fanizzi, F.P. A Decade after the Outbreak of *Xylella fastidiosa* subsp. *pauca* in Apulia (Southern Italy): Methodical Literature Analysis of Research Strategies. *Plants* **2024**, *13*, 1433. [CrossRef]
16. Schneider, K.; van der Werf, W.; Cendoya, M.; Mourits, M.; Navas-Cortés, J.A.; Vicent, A.; Lansink, A.O. Impact of *Xylella fastidiosa* subspecies *pauca* in European olives. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9250–9259. [CrossRef]
17. Hellenic Ministry of Rural Development and Food. National Contingency Plan for the Bacterium *Xylella fastidiosa* (Wells et al.). Official Document, Directorate of Plant Produce Protection. 2024. Available online: https://www.minagric.gr/images/stories/docs/agrotis/Georgika_Farmaka/Fytoeigionomikos_Elegxos/SxedioXylellafastidiosa_090724.pdf (accessed on 29 January 2026).
18. Anița, S.; Capasso, V.; Scacchi, S. Controlling the Spatial Spread of a *Xylella* Epidemic. *Bull. Math. Biol.* **2021**, *83*, 32. [CrossRef]
19. Gilioli, G.; Simonetto, A.; Colturato, M.; Bazarra, N.; Fernández, J.R.; Naso, M.G.; Donato, B.; Bosco, D.; Dongiovanni, C.; Maiorano, A.; et al. An eco-epidemiological model supporting rational disease management of *Xylella fastidiosa*. An application to the outbreak in Apulia (Italy). *Ecol. Model.* **2023**, *476*, 110226. [CrossRef]
20. Surano, A.; Abou Kubaa, R.; Nigro, F.; Altamura, G.; Losciale, P.; Saponari, M.; Saldarelli, P. Susceptible and resistant olive cultivars show differential physiological response to *Xylella fastidiosa* infections. *Front. Plant Sci.* **2022**, *13*, 968934. [CrossRef]
21. Soubeyrand, S.; de Jerphanion, P.; Martin, O.; Saussac, M.; Manceau, C.; Hendriks, P.; Lannou, C. Inferring pathogen dynamics from temporal count data: The emergence of *Xylella fastidiosa* in France is probably not recent. *New Phytol.* **2018**, *219*, 824–836. [CrossRef]
22. Chapman, D.; Occhibove, F.; Bullock, J.M.; Beck, P.S.; Navas-Cortes, J.A.; White, S.M. Modelling plant disease spread and containment: Simulation and approximate Bayesian Computation for *Xylella fastidiosa* in Puglia, Italy. *PLoS Comput. Biol.* **2025**, *21*, e1013539. [CrossRef]
23. Cendoya, M.; Martínez-Minaya, J.; Dalmau, V.; Ferrer, A.; Saponari, M.; Conesa, D.; López-Quílez, A.; Vicent, A. Spatial Bayesian Modeling Applied to the Surveys of *Xylella fastidiosa* in Alicante (Spain) and Apulia (Italy). *Front. Plant Sci.* **2020**, *11*, 1204. [CrossRef]
24. Strona, G.; Carstens, C.J.; Beck, P.S. Network analysis reveals why *Xylella fastidiosa* will persist in Europe. *Sci. Rep.* **2017**, *7*, 71. [CrossRef]
25. Martinetti, D.; Soubeyrand, S. Identifying lookouts for epidemio-surveillance: Application to the emergence of *Xylella fastidiosa* in France. *Phytopathology* **2019**, *109*, 265–276. [CrossRef]
26. Fierro, A.; Liccardo, A.; Porcelli, F. A lattice model to manage the vector and the infection of the *Xylella fastidiosa* on olive trees. *Sci. Rep.* **2019**, *9*, 8723. [CrossRef]
27. Bajocco, S.; Raparelli, E.; Bregaglio, S. Assessing the driving role of the anthropogenic landscape on the distribution of the *Xylella fastidiosa*-driven “olive quick decline syndrome” in Apulia (Italy). *Sci. Total Environ.* **2023**, *896*, 165231. [CrossRef]
28. Brunetti, M.; Capasso, V.; Montagna, M.; Venturino, E. A mathematical model for *Xylella fastidiosa* epidemics in the Mediterranean regions. Promoting good agronomic practices for their effective control. *Ecol. Model.* **2020**, *432*, 109204. [CrossRef]
29. Ali, H.M.; Ameen, I.G.; Gaber, Y.A. The effect of curative and preventive optimal control measures on a fractional order plant disease model. *Math. Comput. Simul.* **2024**, *220*, 496–515. [CrossRef]
30. Boters Pitarch, J.; Signes Pont, M.T.; Szymański, J.; Mora Mora, H. Application of a stochastic compartmental model to approach the spread of environmental events with climatic bias. *Ecol. Inform.* **2023**, *77*, 102266. [CrossRef]
31. Polymenakos, N.M. *Xylella fastidiosa* Stochastic Simulation Framework. Private GitHub Repository. 2025. Available online: https://github.com/nickpol29/xylella_spreading (accessed on 29 January 2026).
32. Regione Puglia. Emergenza Xylella: Dati Monitoraggio. Official Monitoring Datasets of the Apulia Region Phytosanitary Service. 2017. Available online: http://www.emergenzaxylella.it/portal/portale_gestione_agricoltura/Download/mon_xf (accessed on 4 January 2025).
33. Lago, C.; Garzo, E.; Moreno, A.; Barrios, L.; Martí-Campoy, A.; Rodríguez-Ballester, F.; Fereres, A. Flight performance and the factors affecting the flight behaviour of *Philaenus spumarius* the main vector of *Xylella fastidiosa* in Europe. *Sci. Rep.* **2021**, *11*, 17608. [CrossRef]

34. Esri. World Imagery. ArcGIS Online. 2026. Available online: <https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9> (accessed on 29 January 2026).
35. Zecharia, N.; Miri, V.; Dror, O.; Hatib, K.; Holland, D.; Dani, S.; Bahar, O. Seasonal Dynamics and Distribution of *Xylella fastidiosa* in Infected Almond Trees. *Phytopathology* **2024**, *114*, 1186–1195. [[CrossRef](#)] [[PubMed](#)]
36. Polymenakos, N.M.; Vlachos, I.; Karydis, I.; Avlonitis, M. A Novel Low-Cost System for *Xylella fastidiosa* Early Detection Using Electrical Conductivity Probes. *Proceedings* **2025**, *117*, 9. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.