Computers and Electronics in Agriculture 237 (2025) 110561

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Improved pest and disease spread model for simulating the spread of invasive species: A case study of pine wilt disease

Zhuoqing Hao [a,b], Wenjiang Huang [a,b,*], Biyao Zhang [a,*], Yifan Chen [c,d], Guofei Fang [c,d], Jing Guo [a,b], Yanru Huang [a,b], Xiangzhe Cheng [a,b], Bohai Hu [a,b]

[a] Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences. Beijing 100101, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China
[c] Center for Biological Disaster Prevention and Control, National Forestry and Grassland Administration, Shenyang 110034, China
[d] Key Laboratory of National Forestry and Grassland Administration on Forest and Grassland Pest Monitoring and Warning, Shenyang 110034, China

ARTICLE INFO

Keywords:
Biological invasion
Pine wilt disease
Boundary distance
Ecological modelling
Disaster prevention
Control

ABSTRACT

Biological invasion is a major environmental issue facing the world today. Controlling invasive species has become a key focus of governments globally. Pine wilt disease (PWD) is a significant invasive species in China. This study examines the invasion and expansion processes of PWD and provides control strategies. Three regions in China with differing climatic, environmental, and socio-economic conditions were selected. Using real disaster records from 2018 to 2022, the K-means clustering method was applied to analyse the mechanisms of disease spread. Epidemic patches were classified into two groups: invasion-type and expansion-type. By considering both natural and human factors driving the invasion and expansion of PWD, the PoPS model was improved. The model was then used to analyse the effectiveness of control measures against PWD spread in China. The results indicate that the calculated boundary distances and the improved model effectively explained new outbreaks caused by the invasion and expansion processes. Simulation results, combined with China's current control strategies, provide recommendations for disaster prevention. This study describes PWD spread as two processes: invasion and expansion. It reveals the transmission dynamics of PWD in China, offers a predictive model, and proposes control strategies. The findings provide scientific support for disaster prevention and research on invasive species.

1. Introduction

Invasive alien species refer to organisms that have been introduced, established, and spread beyond their native habitats (Diagne et al., 2021; Elton, 2020). The rapid growth of international trade and tourism, coupled with global economic integration and climate change, has accelerated the occurrence of biological invasions. These species are now recognized as one of the top five global environmental challenges of the 21st century (Chen et al., 2020; Seebens et al., 2021), and their prevention and control have become a key focus of governments worldwide (Doherty et al., 2016). Species invasions can introduce new internal feedback mechanisms or disrupt the existing feedback balance within ecosystems, posing one of the most significant threats to global biodiversity, second only to habitat loss (Xu et al., 2004). They are widely acknowledged for causing substantial damage to biodiversity, ecosystem functions, and associated services, while also threatening human health, societal welfare, and economic stability. Globally, invasive alien species are estimated to cause annual losses of up to $1.4 trillion, accounting for nearly 5 % of the global gross national product (GNP) (Pimentel et al., 2005). In China, invasive species have caused severe harm to agriculture, industry, ecosystems, wildlife, and genetic resources. Notable examples include the *Bursaphelenchus xylophilus* (pine wood nematode), *Hemiberlesia pityosophila* (pine needle scale), *Oracella acuta* (loblolly pine mealybug), *Eichhornia crassipes* (water hyacinth), *Mikania micrantha* (mile-a-minute weed) (Xu et al., 2004).

Interactions between organisms and abiotic drivers influence the

* Corresponding authors at: Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences. Beijing 100101, China.
E-mail addresses: haozhuoqing20@mails.ucas.ac.cn (Z. Hao), huangwj@aircas.ac.cn (W. Huang), zhangby@aircas.ac.cn (B. Zhang), chenyifan960510@163.com (Y. Chen), fgfly@163.com (G. Fang), guojing211@mails.ucas.ac.cn (J. Guo), huangyanru19@mails.ucas.ac.cn (Y. Huang), chengxiangzhe22@mails.ucas.ac.cn (X. Cheng), hubohai22@mails.ucas.ac.cn (B. Hu).

https://doi.org/10.1016/j.compag.2025.110561
Received 12 December 2024; Received in revised form 24 April 2025; Accepted 14 May 2025
Available online 22 May 2025
0168-1699/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

spatial distribution and residence time of invaders. The drivers of biological invasions include ecological determinants and anthropogenic factors such as climate change, global trade, and human disturbances (Roura-Pascual et al., 2021). Invasive alien species are introduced into new regions, undergoing a lag phase and naturalization phase before reproducing and expanding. Once conditions become favorable, they may proliferate uncontrollably. A key factor in the potential risk of species expansion is their dispersal ability, such as the capacity of the species or its vectors to disperse actively or passively through wind (Shigesada, 1997). Increasing evidence suggests that landscape fragmentation facilitates the establishment of invasive species, as native species in fragmented landscapes are more vulnerable to invasion (Malavasi et al., 2014). Although species invasions and range expansions have been part of Earth's history before human presence, human activities have accelerated the rate of expansion (Xu et al., 2004), often overcoming natural limitations (Jiang, 2018). Numerous studies indicate that human-mediated dispersal plays a crucial role in the long-distance spread of invasive terrestrial invertebrates (Gilbert et al., 2004; Robinet et al., 2011). Trade activities, in particular, have direct and indirect impacts on biological invasions and interact with other drivers (Hulme, 2021). The combination of long-distance jumps and short-distance movements, known as stratified dispersal, can significantly enhance the rate of expansion (Shigesada, 1997).

Model construction for invasive species can be broadly divided into two categories: habitat-based Species Distribution Models (SDMs) and spread dynamics models based on dispersal processes. Species distribution models have been widely used in invasion biology research (Guisan et al., 2014), primarily to assess the invasion risk and potential distribution of potential invaders using records mainly from their native ranges (Hui, 2023). Bioclimatic models using presence-absence data for climate matching are often employed to determine the potential distribution of species (Elith* J, H. Graham* C, P. Anderson R, 2006). Convolutional neural networks can enhance species distribution models by capturing the spatial structure of the environment, and the integration of both models can achieve significant improvements (Deneu et al., 2021). Linking population genomics, landscape ecology, and species distribution models into a common framework allows for the study of invasion processes (Sherpa and Després, 2021). Liu Tingting et al. conducted a diffusion risk analysis of vegetation pests and diseases using Random Forest (RF) and Support Vector Machine (SVM), two machine learning methods (Liu et al., 2023). Robinet proposed that, due to the low probability and unpredictability of long-distance jumps, the most suitable model to describe this type of spread is stochastic, potentially requiring hundreds of repeated simulations to obtain a representative estimate of potential spread (Robinet et al., 2009). Dynamic network models (DNM) simulate the spread of invasive species, providing a risk assessment framework consistent with the observed spatial scale, while retaining the fundamental biological characteristics of the organisms and spatial data related to habitat resources (Ferrari et al., 2014). Petrasova et al. used the PoPS spread model to analyze the spread of sudden oak death (SOD) in Oregon and evaluated different management strategies (Jones et al., 2021; Petrasova et al., 2020).

Pine wilt disease (PWD) is a major invasive species in China, caused by *Bursaphelenchus xylophilus*. It poses a significant threat to the ecological and biological safety of conifer forests in Asia and Europe. Native to the eastern coasts of the United States and Canada, it has now spread as an invasive species in China, South Korea, Japan, Portugal, and Spain (Hao et al., 2022). Since its introduction to China in 1982, PWD has rapidly spread across the country. To date, it has affected pine trees in 18 provinces, covering 633 county-level administrative regions (https://www.forestry.gov.cn/).The spread of PWD relies on two main mechanisms: natural dispersal and human-mediated dispersal. Natural dispersal occurs when vector insects transfer *B. xylophilus* from infected dead trees to nearby healthy pines, causing new infections. In China, the primary vector is the *Monochamus alternatus* beetle, which typically has one generation per year and a life cycle synchronized with *B. xylophilus*. Its natural dispersal range is usually under 200 m, with reports of maximum distances up to 3.3 km (Robinet et al., 2011; Ning et al., 2004). Human-mediated dispersal occurs when infected pine materials are transported from infected areas to uninfected regions. Once transported, vector insects can emerge from the infected wood and spread the nematode to healthy pine trees, causing new infections. This mode of dispersal can occur over large distances and is primarily driven by trade and transportation of infected materials, including logs and wooden products, along railways and highways (Ning et al., 2004). The spread of PWD in several countries exceeds the natural dispersal range of *M. alternatus* beetle. In China, the average annual spread distance is estimated at 7.5 km (Robinet et al., 2009; Togashi and Shigesada, 2006).

This study uses PWD as a case study to explore the processes of invasion, expansion, and control of invasive species. As an invasive species with a combination of short- and long-distance dispersal mechanisms, the dynamic coupling of the *B. xylophilus*, vectors, host tree species, and the stochasticity of human activities introduce significant complexity to its invasion and expansion. SDMs lack the capability to simulate the process of disaster propagation, making it difficult to achieve dynamic simulations of disaster spread. Methods such as machine learning or deep learning, due to their low interpretability, struggle to effectively dissect the underlying causes of disasters. Therefore, understanding the transmission mechanisms of PWD and constructing mathematical models based on propagation process for prediction and control represents a critical challenge currently faced. This study thoroughly considers the complex mechanisms underlying the spread of PWD. By employing the K-means Clustering method, the concept of boundary distance between two transmission modes is introduced to segment samples from newly infected areas. A data-driven modeling approach focused on the spread process is employed, where two transmission processes of invasive species—forward propagation of the disaster and reverse deduction—are used to construct models for the expansion and invasion of PWD. These models are further applied to disaster control scenario simulations. The dynamic sample segmentation and process-oriented modelling approach proposed in this study have significant implications for early prevention and outbreak management, enabling predictions of future spread and evaluations of containment strategies for newly invaded areas. This also provides a foundational modelling framework for analysing the dynamic processes of invasion and expansion of other invasive species.

2. Data and method

2.1. Study area

China, located in the eastern part of the Eurasian continent, spans a vast territory and hosts a wide variety of pine species with broad distributions. Under natural conditions, at least 17 pine species in China are known to be susceptible to *B. xylophilus*. Existing climate-based studies on suitable habitats indicate that most provinces in China fall within high or moderate suitability zones for PWD. Combined with the strong adaptability of *B. xylophilus*, the disease has continuously expanded beyond its climatic boundaries in recent years. Consequently, areas with pine trees in China are at potential risk of PWD outbreaks (Ye, 2019). This study focuses on three regions in China affected by PWD (Fig. 1). Based on the transmission characteristics of PWD, we consider the differences in the natural environment (elevation variation, absolute altitude, coastal vs. inland regions) and human activities (local and regional timber use, timber export, timber for landscape value). The variations in both natural and human factors lead to differences in the driving processes of PWD transmission. Developing invasion and expansion models for PWD in different regions allows the assessment of the model's applicability across diverse areas of China. This approach aims to extend the model for broader application, enabling the prediction of PWD outbreaks on a national scale. The three study areas selected in this research encompass diverse conditions that reflect these natural and

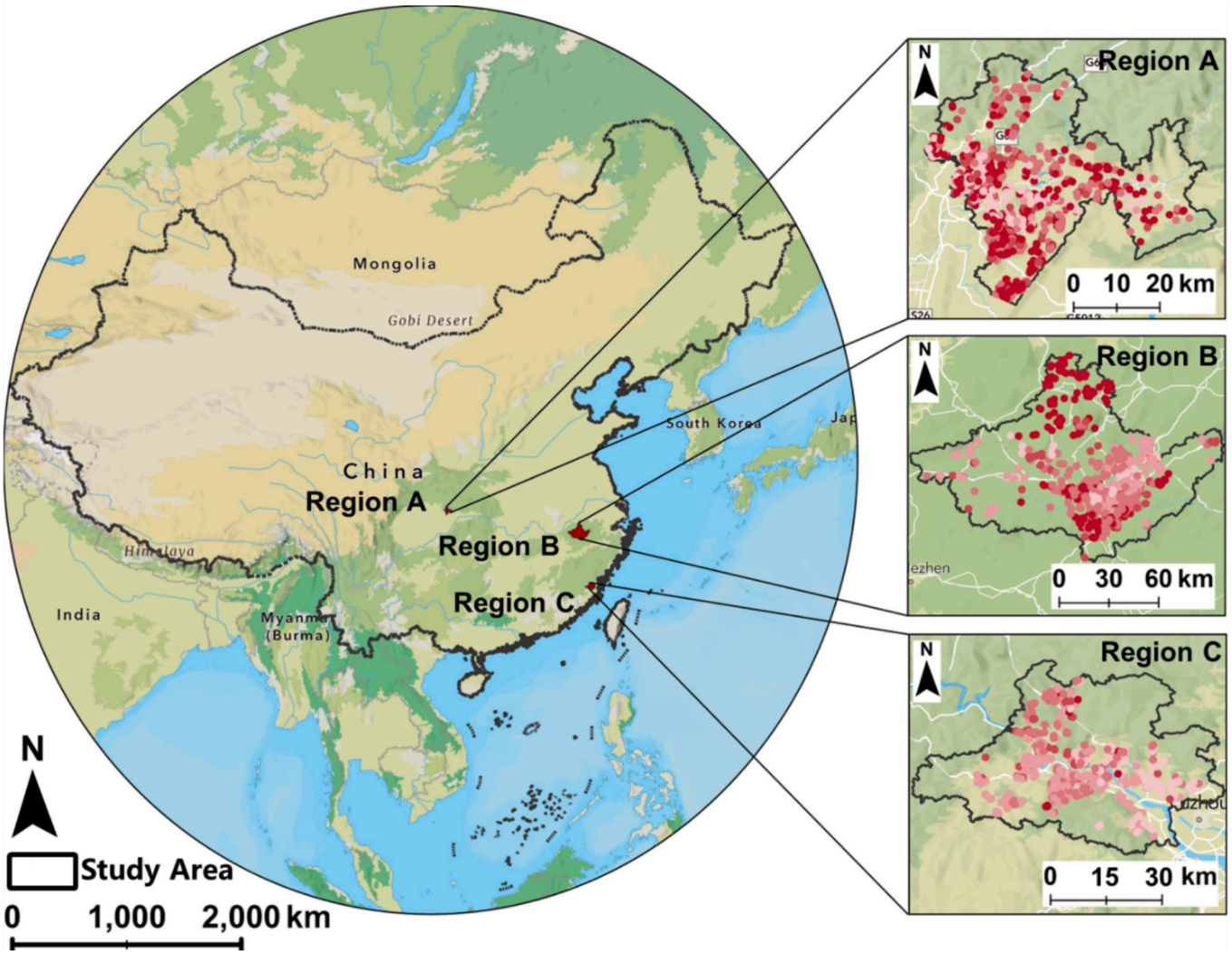


Fig. 1. Study Area.

human environmental differences. The specific environmental conditions of the three regions are as follows:

(1) Region A

Region A encompasses 21 townships within Xuanhan County, Dazhou City, Sichuan Province, with geographic coordinates ranging from 31.11°N to 31.63°N and 107.49°E to 108.10°E, covering a total area of 1,533.90 km^2. The first recorded outbreak of PWD in this area occurred in 2018. Xuanhan County is located in the northeastern part of Sichuan Province, characterized by mountainous terrain with significant topographic variation and abundant vegetation. Pine forests in this region are primarily distributed in high-altitude mountainous areas. Study Area A is situated inland, with an economy predominantly based on traditional agriculture and forestry. The region has limited pathways for external transmission, and wood transportation activities are mainly localized within the Sichuan-Chongqing area, catering to local and nearby demands for construction and timber processing.

(2) Region B

Region B is located in Huangshan City, at the southernmost tip of Anhui Province, with geographic coordinates ranging from 29.40° to 30.51°N and 117.20° to 119.90°E, covering a total area of 9,807 km^2. The first recorded outbreak of PWD in this area occurred in 2018. Huangshan City is predominantly mountainous, featuring interwoven mountains and rivers, with its highest peak reaching an elevation of 1,864.8 m. The forest coverage rate in the study area is 82.9 %. *Pinus hwangshanensis* (Huangshan pine) is a representative species of the region and a key feature of the Huangshan Scenic Area. Renowned for its aesthetic and cultural value, *P. hwangshanensis* (e.g., the iconic Guest-Greeting Pine) is widely celebrated in tourism and is not extensively used for commercial transportation. As a major hotspot for PWD outbreaks, Huangshan City faces a high risk of disease spread. Consequently, prevention and control policies are strictly implemented.

(3) Region C

Region C is located in Fuzhou City, Fujian Province, encompassing 22 townships within Jin'an District, Minhou County, and Minqing County. The geographic coordinates range from 25.98° to 26.48°N and 118.56° to 119.36°E, with a total area of 2,444.46 km^2. The first recorded outbreak of PWD in this area occurred in 2017. Fuzhou City is characterized by a typical estuarine basin topography, with elevations mostly ranging between 600 and 1,000 m. The current forest coverage rate is 57.8 %, and the greening rate reaches 88.6 %. In terms of tree species composition, coniferous forests significantly outnumber broad-leaved forests, with a ratio of approximately 75:25. Among the conifers, *P. massoniana* (Masson pine) and *Cunninghamia lanceolata* (Chinese fir) dominate, with Masson pine being particularly prominent. The

region has abundant pine resources, with timber transport primarily serving Fujian Province and the broader East China region. Due to its developed export-oriented forestry economy and extensive industrial chains, transportation activities are large-scale, exposing the region to multiple risks associated with domestic and international timber movements.

2.2. Data and data processing

(1) PWD data

The outbreak data used in this study consists of epidemic plot data from 2018 to 2022 within the study area. These data were obtained from the Forest and Grassland Pest Control Station under the Center for Biological Disaster Prevention and Control, National Forestry and Grassland Administration (NFGA) (https://www.forestry.gov.cn/). The data are reported annually following spring and autumn surveys conducted by local forestry and grassland bureaus. The dataset includes information such as the central coordinates of each epidemic plot, the year of first occurrence, plot area, land use type, topography, and slope.

(2) Host Data

Due to the lack of continuous temporal data on pine forest distribution in China and the widespread distribution of pine trees across the country, this study uses forest distribution as a substitute for pine forest distribution. Forest distribution data were obtained from the ESRI | Sentinel-2 Land Cover (2017–2023) dataset (https://livingatlas.arcgis.com/). The dataset includes six land cover categories: Water, Trees, Crops, Built, Bare, and Range. Data from 2018 to 2022 were acquired annually, and areas with the attribute value "Trees" were extracted to represent the distribution of *B. xylophilus* hosts.

(3) Predictor Variables

Topographic data were sourced from NASA's SRTM Digital Elevation 30 m dataset, providing ground elevation information. Slope and aspect data were derived through slope calculations based on the elevation dataset.

Fractional Vegetation Cover (FVC) were calculated using Sentinel-2 satellite imagery and SRTM topographic data. The red (B4) and near-infrared (B8) bands were chosen to calculate vegetation indices. Slope and aspect data were used to apply Minnaert correction for terrain shading, reducing the impact of topography on reflectance values. Composite imagery of the red and near-infrared bands was used to calculate the Normalized Difference Vegetation Index (NDVI), which was then used to derive FVC. $NDVI_{soil}$ and $NDVI_{veg}$ represent the NDVI values of bare soil and pure vegetation, respectively. Based on previous studies, this research sets $NDVI_{soil} = 0.2$(Montandon and Small, 2008) and $NDVI_{veg} = 0.8$(Yue et al., 2020). The resulting FVC data had a spatial resolution of 30 m.

$$NDVI = \frac{B8 - B4}{B8 + B4} \tag{1}$$

$$FVC = \frac{NDVI - NDVI_{soil}}{NDVI_{veg} - NDVI_{soil}} \tag{2}$$

Data on timber processing plants were obtained through web scraping, providing vector point data for all timber processing plants as of 2022. The dataset includes attributes such as geographic coordinates, plant names, and registration years (https://www.qcc.com/). The annual distribution of Timber Processing Plants (TPP), as well as the distribution of New Timber Processing Plants (NTPP), was extracted based on registration years.

Road network data were sourced from the 2019 China Road Vector Data, provided by the Geographic Data Sharing Infrastructure, College of Urban and Environmental Science, Peking University (https://geodata.pku.edu.cn).

The Forest Proximate People (FPP) dataset provides an estimate of the number of people living in or within 1 km of forests (forest-proximate people) for the year 2019 with a pixel size of 100 m at a global level (Newton et al., 2022; Fao, 2022).

Data on the Expansion of Built-up Areas (EBA) were obtained from the ESRI | Sentinel-2 Land Cover (2017–2023) dataset (https://livingatlas.arcgis.com/) with a resolution of 10 m. Built-up areas were extracted for each year from 2018 to 2022. Using the 2018 initial built-up areas as a baseline, connected region analysis was applied to identify all potential expansion areas by cumulatively marking newly added pixels of built-up land annually. Additionally, entirely new expansion areas originating from previously non-built-up land were identified.

2.3. Method

2.3.1. Spread mechanism Identification: K-means clustering

This study utilizes clustering algorithms to analyse and identify the spread mechanisms of newly affected plots of PWD. The human-mediated spread distance of PWD exceeds its natural spread distance. Therefore, the breakpoint between these distances represents the boundary between the two types of spread mechanisms. The goal of clustering algorithms is to partition data into several groups or clusters based on proximity, grouping the data using Euclidean distance. K-means clustering is a classic algorithm and a prominent example of unsupervised learning. It is widely applied across various fields due to its simplicity, efficiency, and ease of implementation. K-means algorithm uses distance as a metric to divide a dataset into K non-overlapping clusters, calculating the average distance and initializing centroids. Each cluster is represented by its centroid. The objective is to partition a sample set $D = \{x_1, x_2, \cdots, x_n\}$ into $K(K < n)$ clusters $C = \{C_1, C_2, \cdots, C_k\}$. The metric E indicates the compactness of samples within a cluster relative to the cluster mean vector. A smaller E value indicates higher similarity among samples within a cluster. The mean vector μ_i of cluster C_i is calculated as follows:

$$\boldsymbol{E} = \sum_{i=1}^{k} \sum_{x \in C_i} \|\boldsymbol{x} - \boldsymbol{\mu_i}\|_2^2 \tag{3}$$

$$\boldsymbol{\mu_i} = \frac{1}{|\boldsymbol{C_i}|} \sum_{x \in C_i} \boldsymbol{x} \tag{4}$$

Newly affected epidemic plots of PWD may originate from the spread of the previous year's outbreak or retransmission from the current year's outbreak. Considering the spread distance over two years allows for a more accurate calculation of the boundary distance and sample segmentation. For each newly affected epidemic plot, the nearest distance to the outbreak plots of the current year (NDC) and the nearest distance to the outbreak plots of the previous year (NDP) were calculated using the Euclidean distance method, resulting in two datasets. Based on the transmission mechanisms of PWD, newly affected epidemic plots can be classified into two types: invasion and expansion. The two datasets are sorted in ascending order, and a significant gap in transmission distance is considered an indicator of a shift in the dominant driving factor from natural to human-induced (Jung et al., 2024). Therefore, this study sets $k = 2$, and dividing them into two categories and extracting the values at the boundary. The larger value between the two datasets was taken as the Boundary Distance. The calculated Boundary Distance was used to classify all epidemic plot based on the transmission mechanisms and was subsequently incorporated into the model as a parameter (the flight distance of vector insects). The rules for distinguishing transmission mechanisms of epidemic plot are as follows:

- Expansion Type − Natural Spread (NS):$NDP < Boundary\,Distance$

- Invasion Type – Human-Mediated Spread (HMS):$(NDC > Boundary\ Distance) \wedge (NDP > Boundary\ Distance)$
- Invasion Type – Human-Mediated Spread with Natural Spread (NFHMS):$NDC < Boundary\ Distance) \wedge (NDP > Boundary\ Distance)$

2.3.2. Construction of the spread model: pest or pathogen spread model

This study utilizes the PoPS model to simulate the transmission process of PWD. The PoPS (Pest or Pathogen Spread) Model is specifically designed to predict the spread of pests and diseases and to evaluate control strategies (Jones et al., 2021). It is a stochastic, spatially explicit, susceptible simulation model capable of predicting the spatio-temporal spread of target species. Existing studies have applied this model in pest and disease prediction research (Petrasova et al., 2020). The model can be described as follows in the context of pest and disease spread research: Beta (β) is the starting point. The model calculates spread potential by incorporating environmental factors (X,P,T), and the number of already infected hosts (I). The dispersal kernel function (d_{ij}) is determined by short-range (α_1) and long-range (α_2) dispersal patterns. The establishment probability at a new location depends on environmental conditions (X,P,T) and the proportion of suitable hosts (the ratio of susceptible hosts (S) to total potential hosts (N). The model can be expressed by the following Eq. (5):

$$\varphi_{ijt} = \beta X_{it} P_{it} T_{it} I_{it} {}^{*} K(d_{ij}; \alpha_1, \alpha_2)^{*} \frac{X_{it} P_{it} T_{it} S_{jt}}{N_j} \tag{5}$$

To make the model applicable to PWD, it needs to be adapted according to the specific spread process of PWD. PWD involves a complex cycle among *B. xylophilus*, vector beetles, and host pine trees. Based on the biological characteristics of the nematode and its vector beetle, as well as the current status of wood transportation in China, the coupled invasion and expansion processes of PWD can be illustrated as shown in Fig. 2. The spread of the disease can be described as an expansion process (①) and an invasion process (②, ③), where processes ① and ③ are dominated by natural factors, while process ② is driven by anthropogenic factors. According to the two spread processes of PWD, we used the PoPS model to construct both the invasion model and the expansion model.

The model was built using the open-source PoPS library, written in modern C++ language, with R interface (Jones et al., 2021). This study utilized the R package of the PoPS model to implement the model in R Studio. The model construction, inputs, outputs, and accuracy evaluation are as follows:

(1) Construction of the expansion process model and accuracy evaluation

The spread process of PWD refers to the dissemination from existing outbreak locations, influenced by natural environmental factors, spreading outward. The natural factors driving the spread of PWD are affected by environmental variables such as temperature, precipitation, host distribution, and topography. Since PWD continues to surpass its suitable ecological boundaries in China (Zheng et al., 2021), the risk of PWD exists wherever pine trees are found. Given that the study area is relatively small with minimal meteorological variation, this study excluded the effects of temperature and precipitation during the diffusion process. The focus of this study is on host distribution and topographic factors as key drivers in the spread process (Hao et al., 2022). Host distribution, extracted as the distribution of susceptible tree species, was incorporated into the model. The model uses these three factors to construct a resistance surface for disease spread, which serves as the environmental basis for the propagation of the disease.

This study used Species Distribution Models (SDMs) to construct the resistance surface. Previous studies have used SDMs to construct transmission resistance surfaces (Liu et al., 2018). Therefore, the resistance surface for PWD spread in this study was derived using Maxent (Maximum Entropy Model), the most widely used and effective tool in SDM applications, based on FVC, DEM, Slope (Phillips et al., 2004). The objective function of the Maxent model can be described by equation (6), where $P(x)$ is the predicted probability distribution, X represents all possible locations in the environmental space, and $H(P)$ is the entropy, which quantifies the uncertainty of the system. The study randomly sampled 100 epidemic plots across three regions, extracting their geographic coordinates, with 30 % of the points used as validation samples. The average of 100 repetitions of the Maxent model was used as the resistance surface for natural PWD spread.

$$\mathrm{H(P)} = -\sum_{x \in X} P(x) log P(x) \tag{6}$$

The model construction for the expansion process of PWD utilized epidemic plot data from 2018 to 2021 as the starting point input into the

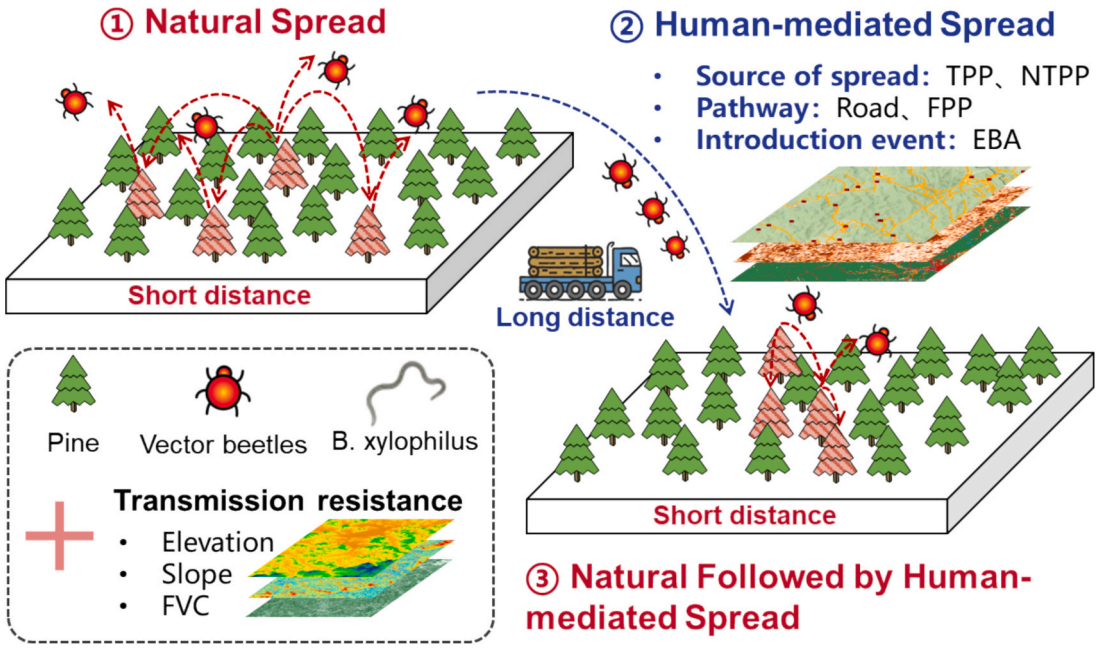


Fig. 2. Flowchart of Spread Mechanisms.

model to predict the occurrence of PWD in the following year. A power-law dispersal kernel and boundary distances calculated using K-means clustering, accounting for the flight tendencies of the vector beetles (Dimitrijevic and Bacic, 2013). The model was constructed based on the distribution of host pine trees, using the transmission resistance surface calculated by Maxent as an environmental factor. The model was run 500 times, and the output probability of PWD occurrence was taken as the expansion risk of the disease.

(2) Construction of the invasion process model and accuracy evaluation

The anthropogenic factor-driven spread process is analysed from three perspectives: the sources of spread, pathway, and introduction event. The invasion process differs from the expansion process in that it is a form of "originless" spread. Timber processing plants handle infected wood and wood products; if infected wood is not processed in a timely manner, outbreaks are more likely to spread around these plants. Changes in construction land use signify major infrastructure projects, which are typically associated with the inflow of large amounts of timber. In the invasion process of PWD, timber processing plants can be assumed as the point of origin for the spread, with land-use changes serving as the endpoint. Regardless of the origin or endpoint, the movement of the disease between these points is facilitated by road networks and human activities (FPP). Therefore, this study considers the disaster spread process from both the forward and backward propagation perspectives. Given that both the introduction and diffusion of the disease are dynamic processes, once the human-driven invasion occurs, natural diffusion quickly follows in the surrounding infected areas. Thus, in addition to considering the invasion risk, the natural re-spread of the disease must also be accounted for.

Therefore, the model construction uses wood processing plants (TPP, NTPP) and Expansion of Built-up Areas (EBA) from 2019 to 2022 as the starting points, with FPP as environmental factors. Anthropogenic spread is modelled using the road network structure, based on the distribution of host pine trees. The model was run 500 iterations to generate preliminary results on the invasion of new areas driven by human activities. The invasion risk results were used as the starting points for the dispersal model, and the model was iterated 500 times. The output probability of PWD occurrence was taken as the invasion risk of the disease.

(3) Model Validation and Performance Assessment.

This study employed multiple evaluation metrics to assess model performance, with recall (sensitivity) and the Receiver Operating Characteristic (ROC) curve being the primary indicators. Recall is defined as the proportion of true positive samples that are correctly identified by the model. It reflects the model's ability to capture the actual occurrence of PWD within the study area. A higher recall indicates a lower risk of omission, ensuring that regions with actual infestations are effectively detected. The ROC curve, which performs a threshold-independent evaluation by scanning across all possible thresholds, provides a comprehensive assessment of the model's ability to distinguish between positive and negative classes. It is particularly suitable for scenarios with imbalanced class distributions and serves as a complementary measure to recall by offering a more holistic evaluation of model performance.

TP represents the number of true positives, and FN represents the number of false negatives. Recall is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

The ROC curve plots the false positive rate (FPR) on the x-axis and the true positive rate (TPR) on the y-axis. The area under the ROC curve (AUC) serves as a quantitative indicator of the model's discriminative ability. The AUC ranges from 0.5 to 1, where a value of 0.5 indicates performance equivalent to random guessing, and a value of 1 represents perfect classification. The AUC is calculated using the non-parametric trapezoidal integration method, which accumulates the areas between adjacent points through linear interpolation.

$$AUC = \frac{1}{2}\sum_{i=1}^{n}(FPR_i - FPR_{i-1})(TPR_i - TPR_{i-1}) \tag{8}$$

The natural spread of PWD typically follows an annual cycle. Therefore, the natural spread model was validated using expansion- type newly affected epidemic plots of the next year (NS). Predicted occurrences were identified by extracting the predicted risk probabilities at sample points, from which the recall rate was calculated to reflect the model's ability to explain the actual new infections in the following year. Given that infected compartments are difficult to eradicate once the disease appears, the combined distribution of infected compartments from the first year and the newly infected compartments from natural spread in the second year was used to represent the true distribution of natural spread. This combined dataset was employed to construct the ROC curve and calculate the global AUC value, providing a comprehensive evaluation of the natural spread model.

Human-mediated spread is typically characterized by lag effects and randomness, and its outcomes are greatly influenced by wood quarantine and control measures. Therefore, the invasion model uses invasion-type newly affected epidemic plots over four consecutive years (HMS + NFHMS) as validation data. Instead of using the ROC curve, the model is evaluated using a combination of recall rate and risk area. This metric reflects the model's ability to explain the actual new epidemic plots caused by human-mediated spread of PWD within the region.

(4) Scenario simulation of control and prevention.

Considering that the actual occurrence of disease spread is closely related to disease control measures, this study conducted prevention and control simulations based on the prediction results for the two spread processes of PWD. For the expansion process, 10 control scenarios were set with disease removal ratios ranging from 0 % to 90 % in increments of 10 %. For the invasion process, isolation zones ranging from 0 to 500 m, with a 50-meter step size, resulted in 11 different prevention and control scenarios. All models were repeated 500 times, and the effectiveness of the disease control measures was evaluated.

3. Results

3.1. Discriminant analysis of spread mechanisms and factor response analysis

The boundary distances of PWD in the three study areas were extracted using K-means clustering, enabling the classification of sample points based on their underlying spread mechanisms. Response curve analyses were conducted for two groups of samples with respect to natural dispersal factors and human-driven factors. The results are presented in Fig. 3.

K-means clustering was applied to newly affected PWD outbreak points in the three regions from 2019 to 2022 (Fig. 3-a). The calculated boundary distances between human-mediated and natural spread for the three regions were 2,806.84 m, 3,655.57 m, and 3,325.87 m, respectively. Based on the boundary distances, PWD spread was classified into three types: NS, HMS, and NFHMS. In Region A, invasion-type newly affected epidemic plots (HMS+NFHMS) were concentrated in 2020, with NFHMS accounting for 96.88 % of all invasion-type newly affected epidemic plots (HMS+NFHMS). In Region B, invasion-type newly affected epidemic plots (HMS+NFHMS) were more evenly distributed across the years, accounting for 37.15 % of all newly affected epidemic plots, with invasion spread (HMS+NFHMS) in 91.16 % of NFHMS. In

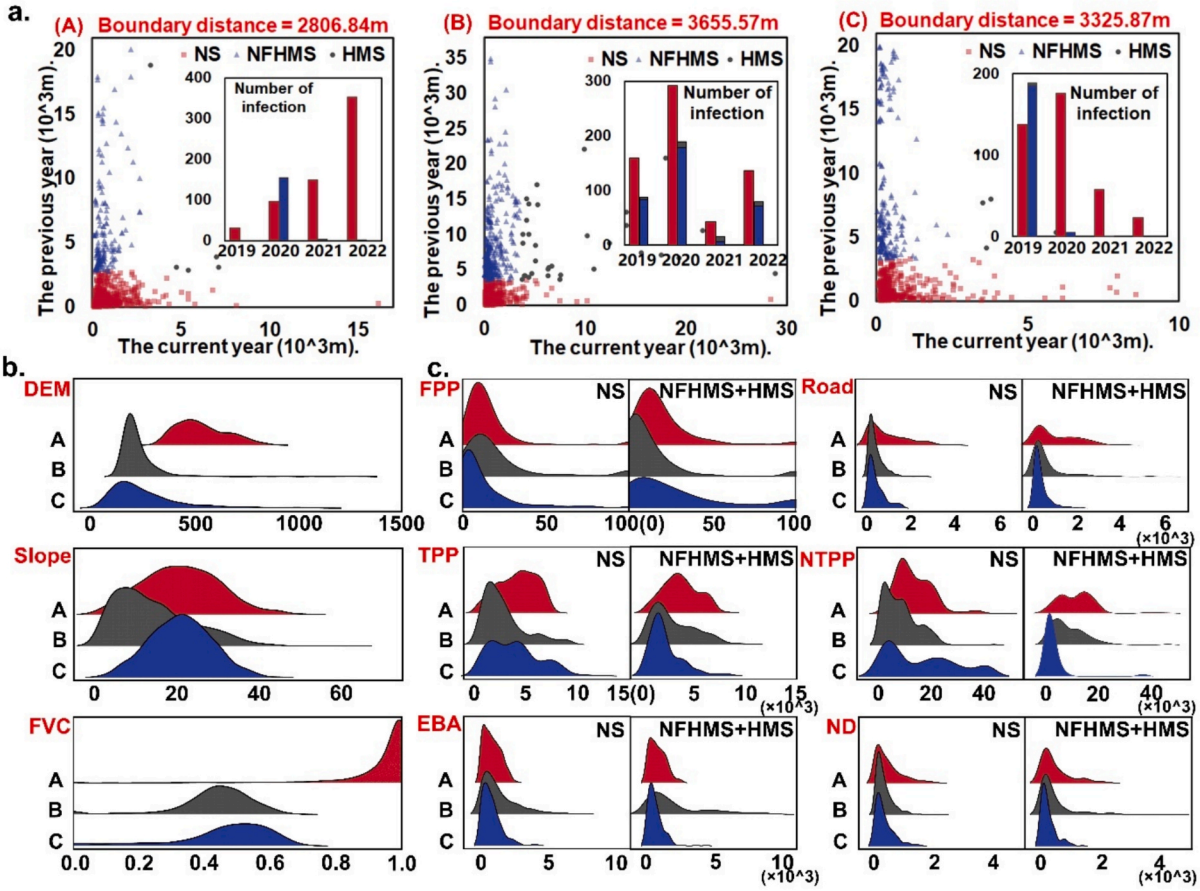


Fig. 3. K-means Clustering Results and Factor Response Curves.

Region C, invasion-type newly affected epidemic plots (HMS+NFHMS) were concentrated in 2019, with NFHMS accounting for 97.54 % of all invasion-type newly affected epidemic plots (HMS+NFHMS).

Response curves were plotted for different spread types against natural spread factors (Fig. 3-b). Epidemic plots in all three regions were concentrated in areas with lower elevations, gentler slopes, and denser vegetation. The majority of epidemic plots were at elevations below 1,000 m, with slopes concentrated between 0° and 40°.The newly affected epidemic plots in Region A were located at higher elevations compared to those in Region B and C. In Region B and C, most newly affected epidemic plots were concentrated at elevations below 500 m, though some were found at elevations exceeding 1,000 m. Despite the differences in elevation distribution among the three study areas, newly affected epidemic plots in all regions were primarily located in areas with gentle slopes, where elevation variations were minimal. FVC varied across the regions. In Region A, epidemic plots were concentrated in areas with dense vegetation, while in Regions B and C, FVC ranged primarily between 0.4 and 0.6.

Response curves were plotted to analyse the distribution characteristics of epidemic plot samples under the two spread mechanisms in relation to human activity-driven factors (Fig. 3-c). The results indicate that both expansion-type and invasion-type newly affected epidemic plots are generally located near human activity events. Among them, newly affected PWD plots had the shortest distances to Road and EBA, with the nearest distances concentrated between 0 and 4,000 m for Road and 0–5,000 m for EBA. The distances between newly affected epidemic plots and TPP as well as NTPP were the most variable. Within the same region, the distances between these two human activities and expansion-type newly affected epidemic plots (NS) were generally greater than those between them and invasion-type newly affected epidemic plots (HMS+NFHMS). Considering that human activities do not exhibit the same synergistic effects as natural factors, the study further examined the Nearest Distance (ND) of introduction factors and plotted response curves for the distance between epidemic plots and the nearest human activity. The results show that, in all three study areas, the ND between human activity and newly affected PWD plots remained within the boundary distance (2,806.84 m, 3,655.57 m, and 3,325.87 m). Only two newly affected epidemic plots in Region B exceeded the boundary distance.

3.2. *Modeling the expansion process of PWD*

Epidemic plots from 2018 to 2021 were used as input, while expansion- type newly affected epidemic plots for 2019 to 2022 were used as validation data, the results are shown in Fig. 4. The study findings indicate that the model achieved high accuracy in predicting the spread of PWD (Fig. 5 and Table 1). However, prediction accuracy tended to decrease in years with significant human-mediated spread. Additionally, when a large number of human-mediated newly affected epidemic plots occurred in one year, the risk area for the following year expanded rapidly. This pattern was more pronounced in Region A and C, whereas Region B consistently demonstrated better performance. The likely reason for this discrepancy is that human-mediated spread in Region A and C was concentrated in a single year and involved fewer epidemic plots, making the boundary distances less accurate compared to those in Region B.

In Region A, the risk area was 225 km^2 in 2019, 252 km^2 in 2020, 606 km^2 in 2021, and 687 km^2 in 2022. The rapid increase in the risk area in 2021 was attributed to a significant number of human-mediated newly affected epidemic plots in 2020. The model achieved AUC values greater than 0.8 in Region A, indicating high performance. However, the recall rate in 2020 was lower than in other years, at 72.63 %, suggesting

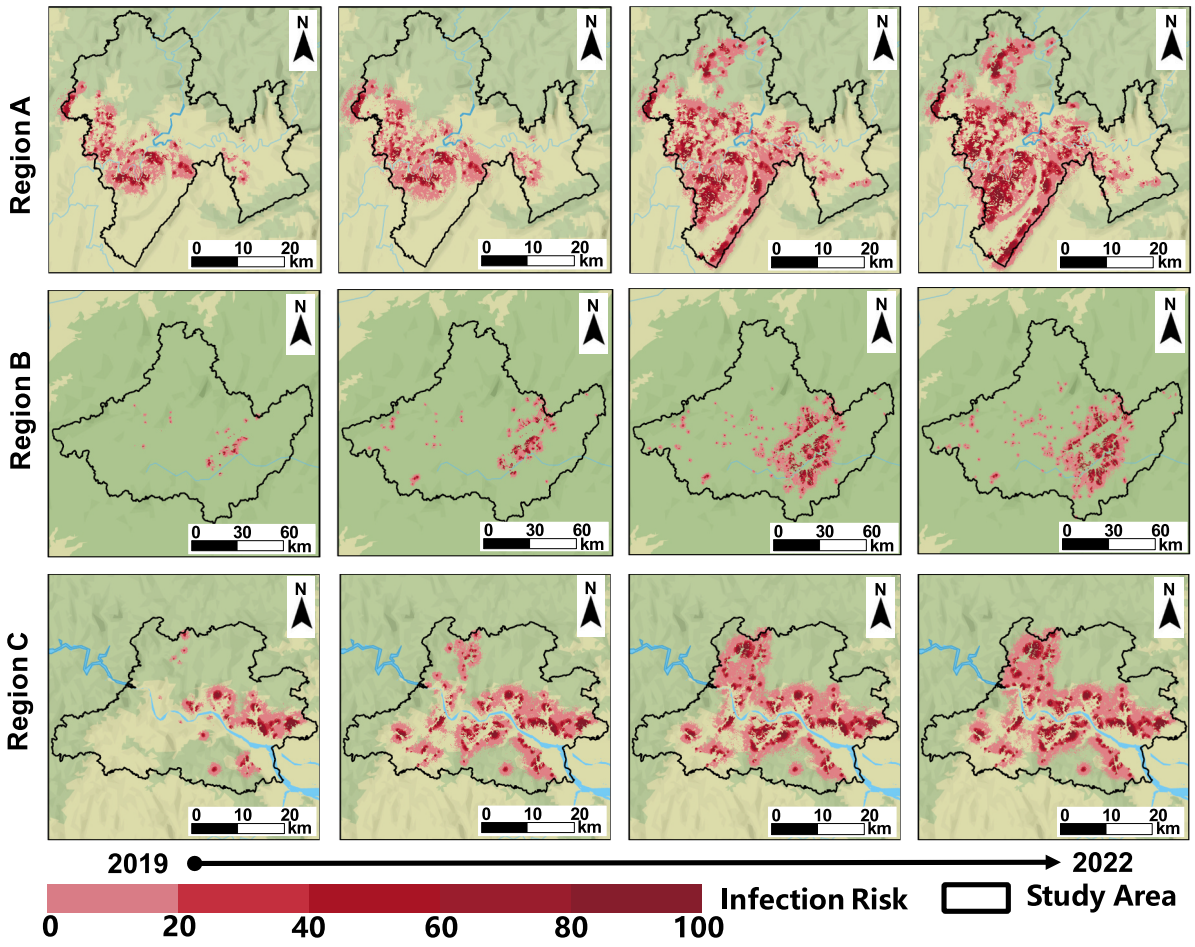


Fig. 4. Predicted Expansion Risk Results (2019–2022).

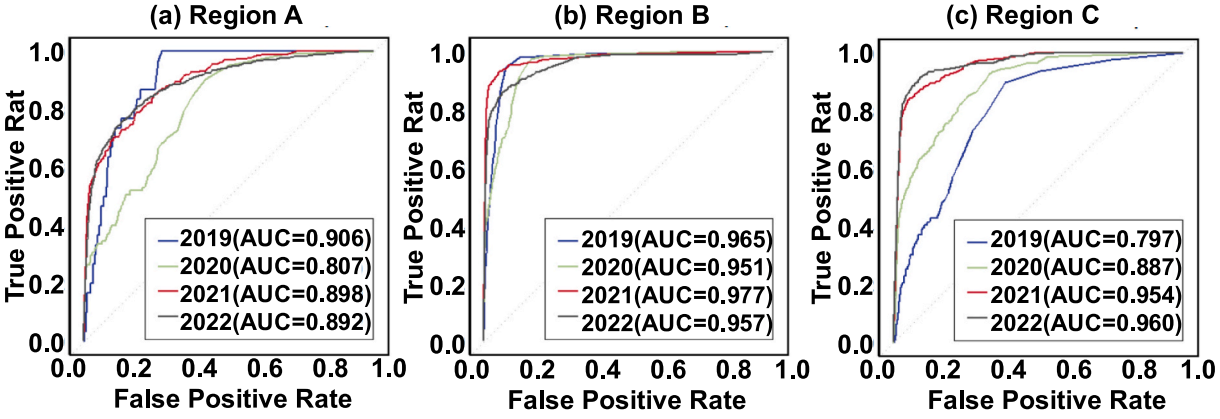


Fig. 5. ROC Curve of the Natural Spread Risk Prediction Results.

Table 1
Predicted Expansion Risk Results and Accuracy (2019–2022).

Region	2019		2020		2021		2022	
	Recall	Area(km^2)	Recall	Area(km^2)	Recall	Area(km^2)	Recall	Area(km^2)
A	100	225	72.63 %	252	97.84 %	606	95.50 %	687
B	94.24 %	160	97.2 %	529	90.24 %	1238	95.52 %	1369
C	71.31 %	263	94.87 %	621	100	815	100	861

weaker explanatory power of the model for that particular year. In Region B, the risk area was 160 km2 in 2019, 529 km2 in 2020, 1,238 km2 in 2021, and 1,369 km2 in 2022. The model performed better in Region B than in Region A, with both AUC and recall values showing favourable results. In Region C, the risk area was 263 km2 in 2019, 621 km2 in 2020, 815 km2 in 2021, and 861 km2 in 2022. The rapid increase in the risk area in 2020 was due to a large number of human-mediated newly affected epidemic plots in 2019. The model's performance in Region C varied across different years, with relatively low AUC and recall values in 2019, recorded at 0.797 and 71.31 %, respectively.

3.3. Modeling the invasion process of PWD

The invasion model consists of two components: predicting human-mediated introduction and modelling natural secondary spread. The results of invasion-type spread predictions from 2019 to 2022 are shown in Fig. 6 and Table 2. The introduction model performed well in this process, effectively explaining the long-distance introduction of PWD. Since the prediction of invasion-type newly affected epidemic plots relies on static pathways such as FPP and road networks, the temporal variation of the disease is primarily determined by the establishment of new timber processing plants and the continuous expansion of construction projects. Over four consecutive years, the total risk area from human-mediated spread in the three regions was generally larger than that from expansion-type spread, but the changes in risk area were relatively small.

In Region A, risk areas were distributed across all townships within the region, with an average risk area of 742 km^2 over four years. The risk area showed fluctuating changes during this period, and the model explained 93.88 % of the invasion-type newly affected epidemic plots over four years. In Region B, risk areas were concentrated in the eastern part of the region, particularly in forest areas surrounding urban centres, where a large number of zones had risk levels exceeding 80 %. The risk area gradually increased over four years, with an overall growth rate of approximately 1.79 % and an average risk area of 3,281 km^2. The model explained 92.17 % of the invasion-type newly affected epidemic plots in this region over four years. In Region C, risk areas were concentrated along the forest edges near urban centre and around road networks in forest interiors. The risk area exhibited fluctuating growth, with an average risk area of 455 km^2 over four years. The model explained 92.05 % of the invasion-type newly affected epidemic plots in this region during the same period.

3.4. Disaster management scenario simulation

Disaster management scenarios were developed for both the invasion and expansion processes to study the impact of different control measures on disease spread. Given that multiple regions in China have experienced outbreaks of PWD, the level of disaster control depends on the resources allocated for its management. Therefore, this study incorporates two control strategies: management intensity and epidemic isolation zones. The results of the simulations under the two control strategies are shown in Fig. 7.

The expansion-type spread relies on the locations of epidemic plots from the previous year, making the level of epidemic control a critical factor influencing disease spread. Epidemic control levels were set from 0 % to 90 %, in increments of 10 %. Using the 2022 epidemic plot data, natural spread models under different control scenarios were

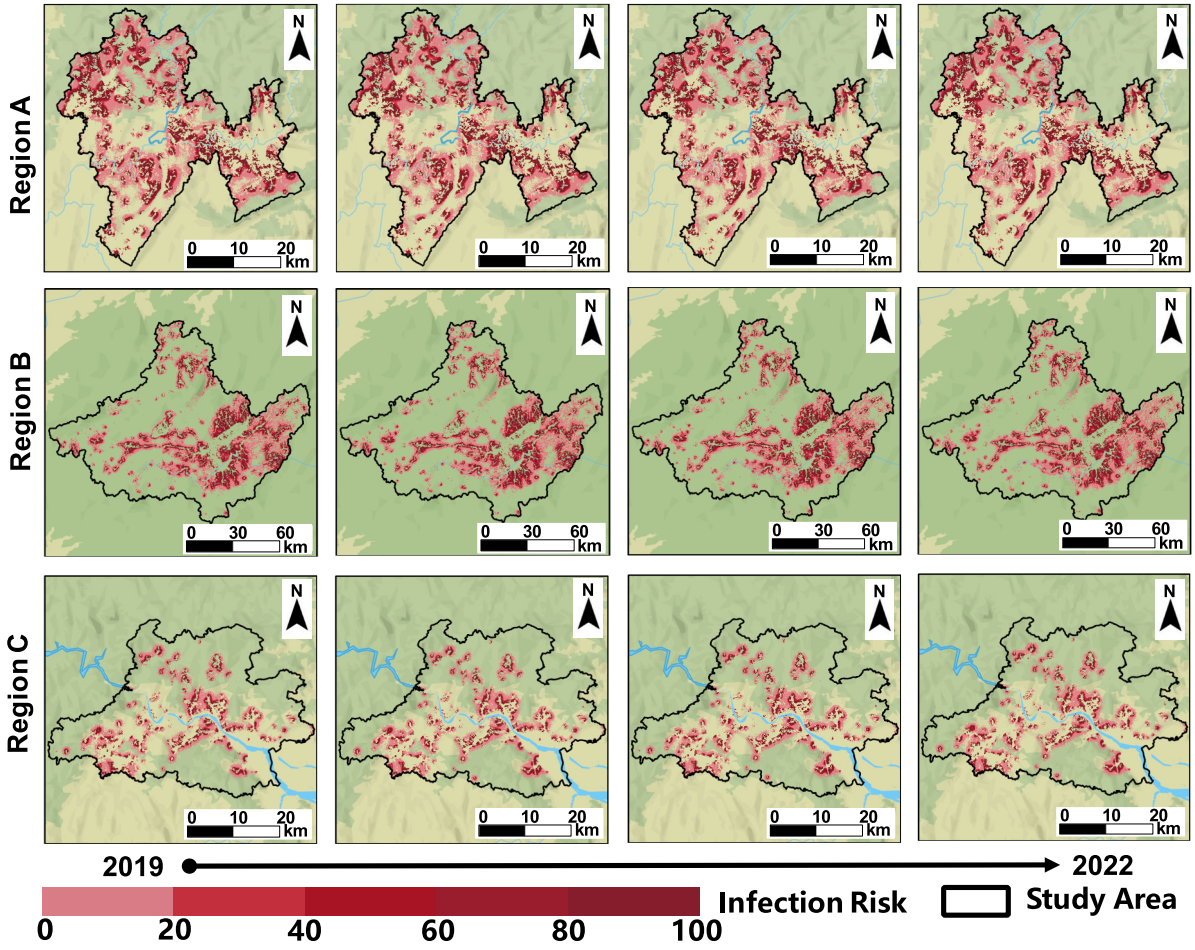


Fig. 6. Predicted Invasion Risk Results (2019–2022).

Table 2
Predicted Invasion Risk Results and Accuracy (2019–2022).

Region	Risk Area(km^2)				The Change of Risk Area	Total Accuracy
	2019	2020	2021	2022		
A	743	746	728	752		93.88 %
B	3,251	3,277	3,286	3,309		92.17 %
C	442	459	454	464		92.05 %

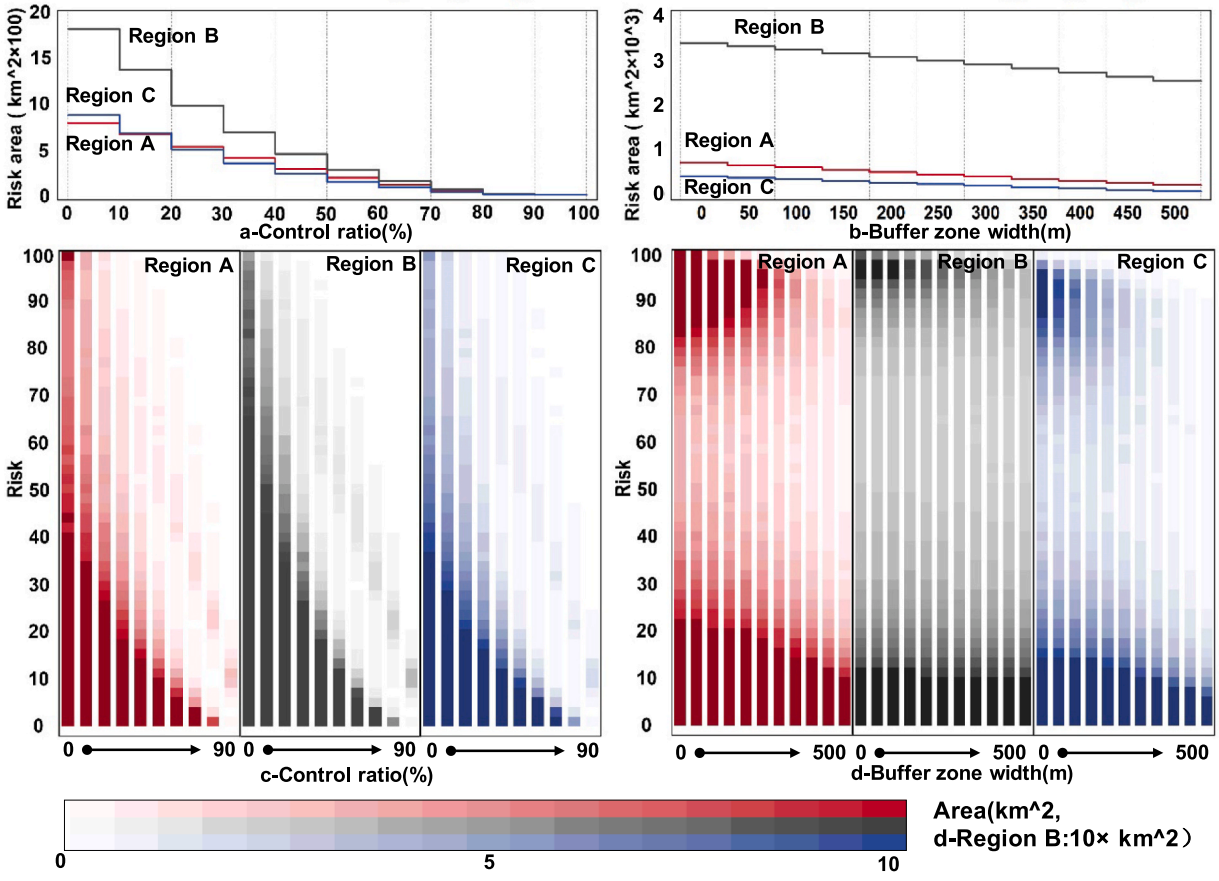


Fig. 7. Simulation of Epidemic Prevention and Control Measures.

constructed, with each model simulated 500 times. The results are shown in Fig. 6-a and c. The findings indicate that the level of epidemic control has a significant impact on the spread of the disease in the following year. Compared to scenarios without control measures, the epidemic risk area decreases annually with increasing control intensity. When the control level exceeds 40 %, the epidemic risk values begin to decline, and high-risk areas gradually diminish. Invasion-type spread is driven by human activities, particularly the distribution and expansion of timber processing plants and built-up areas. To prevent disease spread, isolation zones ranging from 0 to 500 m, in increments of 50 m, were introduced. Using the 2022 introduction risk points, introduction risk diffusion models were constructed under different control scenarios, with each model simulated 500 times. The results are shown in Fig. 6-b and d. The results show that, compared to scenarios without control measures, the risk area decreases gradually with the expansion of isolation zones. However, the reduction in risk area is less significant for introduction-type spread than for diffusion-type spread. Nevertheless, changes in risk values are notable, as regions with 80–100 % risk values gradually decrease as the width of the isolation zone increases.

4. Discussion

The PoPS model demonstrated strong performance in simulating the invasion and expansion processes of the disease across the three study regions. The expansion model effectively predicted the temporal and spatial dynamics of the disease. The three study areas exhibit significant differences in natural environmental conditions, particularly in elevation and host distribution, which may explain the variation in the flight boundary distances of the vector beetles among the regions. Previous studies have shown that the flight of *M.alternatus* is limited by elevation barriers, which is consistent with the result of the shortest spread boundary distance observed in Region A (Zhao et al., 2008). The differences in boundary distances between Region B and Region C are likely due to varying intensities of pest control efforts. Since this study

uses region-specific boundary distances as parameters for the vector beetles' flight capacity in the model, the model maintains high accuracy despite environmental heterogeneity across the three study areas. The human-mediated introduction model provided valuable insights into why PWD has erupted in multiple regions in China and worldwide. The human-mediated spread model achieved high recall rates for newly affected epidemic areas across three regions with different timber circulation directions, indicating its ability to explain the outbreaks of PWD in various regions of China and even globally. Although regional variations in human activities and control policies introduce randomness into the spread process, the underlying transmission mechanism exhibits a degree of universality. The predictions offer an estimation of the potential spread range over a defined future period.

The model's strong performance across all three regions demonstrates its adaptability. However, two issues regarding the model require further explanation:

(1) Randomness

PWD is spread through the random movement of vectors carrying *B. xylophilus* eggs (Dimitrijevic and Bacic, 2013). This process follows the ecological cycles of *B.* xylophilus, vector beetles, and pine trees and can be described as a biased random walk based on directional tendencies during the spread process (Riascos and Mateos, 2021). The results of Section 3.2 in this study indicate that for expansion-type newly affected epidemic plots (NS), the model's explanatory power is weaker for plots located near the boundary distance. This is due to the random walk process of PWD, which results in a lower occurrence probability at the boundary compared to areas surrounding existing epidemic plots. Therefore, in this study, a power-law diffusion kernel was used to model short-distance dispersal, which captures the long-tail effect (Clauset et al., 2009). In addition, the model mitigated the randomness-induced variability in the prediction results through repeated experiments. Nevertheless, this issue persists in the model, which is consistent with the ecological process of PWD spread. The randomness in the invasion process stems from the intensity and location of human activities, which do not follow the ecological cycles of the disease and are inherently more uncertain compared to the random movement of insects (Robinet et al., 2009). The randomness of human activities is mitigated in the modelling process by incorporating activity-related information. This is specifically addressed by extracting data on the ongoing expansion of construction projects, which reflects human activity over a given period. This also explains the temporal dynamics of both the invasion and expansion models.

(2) Scope of the results

The predicted scope of the two models may be larger than the actual occurrence of the disease. For the expansion process, this is because actual disease occurrence is often a result of control measures implemented during the outbreak. The expansion model does not account for PWD control measures. In practice, China has implemented rigorous PWD prevention and control strategies, which explicitly emphasize selective removal of dead and dying trees. These trees are to be promptly shredded or incinerated either on-site or nearby (Ye, 2019). The effectiveness of PWD control is closely tied to the level of investment by local governments. In the case of the invasion process, the predicted spread area is often much larger than the actual newly affected areas. This discrepancy is due to strict quarantine measures in timber transportation and China's diverse natural environments and climate. Despite the expanding suitable range of *B. xylophilus*, the significant environmental differences between source and destination regions often require a period of adaptation (Jiang et al., 2022). The influx of PWD is closely linked to trade activities (Xie et al., 2009). However, not all human activities related to pine trees lead to the entry of infected timber. Cross-regional spread is primarily caused by illegal timber trade that escapes quarantine (Wang, 2015; Zhou, 2019), a process that is inherently difficult to predict.

The expansion model can be used to assess the spatial extent and rate of natural spread, providing a more accurate estimate of the disaster risk propagation range. This result helps forest managers identify high-risk areas and optimize resource allocation. Under uncontrolled conditions, the simulation results of the model can visually present the spatiotemporal dynamic distribution of high-risk areas, thus providing scientific evidence for early monitoring and precise prevention and control. Regarding the spread of PWD, this model fully accounts for the boundary distance of disaster spread and the driving factors, which play a more significant role at the local scale than temperature and precipitation (Hao et al., 2022). The research results indicate that outbreaks are mostly concentrated in areas with lower elevations, higher tree densities, and gentler topographies, where disaster prevention and control measures are relatively mature. However, it is evident that outbreaks also occur in areas with higher elevations and steeper terrains. Although such cases are less frequent, their spread can result in significant damage. The prevention and control of PWD in these regions require the integration of both domestic and international advanced technologies. Various strategies have been explored, including the use of parasitoid wasps, entomopathogenic nematodes (EPN), and insect-pathogenic fungi for biological control of the vector species. Novel approaches have also been studied, such as using kairomones and pheromones to trap *M. alternatus* beetle, thereby monitoring and suppressing vector populations (Back et al., 2024). For ecologically important regions and historic or culturally significant pine trees, a variety of trunk-injection protectants have been developed domestically and internationally to prevent and halt infection in pine trees (Ye and Wu, 2022). However, differing levels of investment in disease control across regions in China, combined with the steep and rugged terrain of forested areas where infected trees are located, make the treatment of infected wood particularly challenging. Effective eradication of the disease in outbreak areas remains difficult, as confirmed by the disease control policy simulations conducted in this study.

The invasion process model can effectively identify the risk of disease introduction over large areas and has predictive capability for the coming years. This predictive ability can be used to guide the development of prevention and control policies. In situations with limited resources, efficient management can be achieved by accurately identifying high-risk areas, which helps optimize prevention and control costs. The high-risk areas identified by the model in this study allow managers to proactively plan isolation zones, effectively curbing long-distance spread during the early stages of disease introduction. This finding aligns with real-world experiences with isolation zones in Portugal (Back et al., 2024). Another effective control measure is to strengthen quarantine and monitoring efforts in these areas. This model clearly reveals the process of disaster spread and allows for the formulation of prevention and control measures from multiple angles for different stages of the spread process. Human-mediated spread primarily involves the movement of timber and wood products, including logs, sawn timber, processed wood, wood chips, harvest residues, and wood packaging. In Asia, the main carriers of human-mediated spread are infected wood, packaging crates, cable reels, and optical cable reels (Back et al., 2024; Ye and Wu, 2022). The findings of this study, along with existing disease control measures, indicate that disease quarantine consists of two components: origin quarantine and transport quarantine. Origin quarantine focuses on timber processing plants and marketplaces in outbreak areas, while transport quarantine prohibits the movement of infected pine wood and associated products out of outbreak areas or into uninfected areas (Ye and Wu, 2022). Proper management of infected wood, ensuring no outflow, and strictly prohibiting the movement of infected wood are effective measures to prevent disease spread along transmission pathways.

The invasion model and the expansion model developed in this study demonstrate significant predictive capability and provide important

scientific support for the prevention and control of PWD. The results from the invasion model can be used to identify new regions where outbreaks may occur, while the expansion model further quantifies the potential pathways and rates of disease expansion within already infected areas. By combining these two models, the study offers a more comprehensive assessment of the disease spread risk and supports the formulation of integrated control strategies. In the future, combining the predictions from these two models will enable a full-chain prediction of PWD, from long-distance invasion to regional spread, providing a technical reference for the study of spread dynamics of similar invasive species. However, there are still limitations. This study relies on temporal forest distribution data as host data for PWD, without incorporating specific host tree species data and lacking field-based geographic survey data. As a result, the predicted infected range larger than the actual, and it is difficult to account for the susceptibility and resistance of different pine species. Therefore, integrating the raster results obtained in this study with forest resource survey data can enhance the accuracy of the predictions. Additionally, due to the current lack of detailed timber trade data with specific introduction points and dissemination months, the research is based on actual timber usage in China. Considering that timber transportation is often associated with major infrastructure projects, the expansion of built-up areas was used as a proxy for the introduction of the disease. However, this approach may not be easily transferable to other countries where PWD occurs. Future research on PWD and other invasive species will need to address challenges related to host species distribution, timber trade volume, and the timing of timber trade introductions.

5. Conclusion

This study uses the process of PWD invasion and expansion as an example to construct a model of biological invasion. Based on the biological characteristics and spread mechanisms of PWD, discriminant analysis was conducted on the spread mechanisms of newly affected epidemic areas. Models were built using newly affected epidemic plot samples and spread driving factors for the two types of spread mechanisms, with both models showing good explanatory power for the new outbreaks of the disease. In relation to the disaster spread process, this study also conducted research on the implementation and effectiveness of epidemic prevention and control measures, providing guidance on current PWD control strategies based on real disaster prevention policies and control effectiveness. The results of this study contribute to the modeling of other biological invasion processes and provide scientific and technical support for the prevention and control of PWD in China.

CRediT authorship contribution statement

Zhuoqing Hao: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Wenjiang Huang:** Writing – review & editing, Supervision, Funding acquisition. **Biyao Zhang:** Writing – review & editing, Supervision. **Yifan Chen:** Data curation. **Guofei Fang:** Data curation. **Jing Guo:** Formal analysis. **Yanru Huang:** Methodology. **Xiangzhe Cheng:** Formal analysis. **Bohai Hu:** Data curation.

Funding

This research was funded by National Natural Science Foundation of China (42201355), Sino-UK Crop Pest and Disease Forecasting & Management Joint Laboratory, SINO-EU Dragon 6 proposal (ID 95250), Global Vegetation Pest and Disease Dynamic Remote Sensing Monitoring and Forecasting (GEO-PDRS).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

Diagne, C., Leroy, B., Vaissière, A.-C., et al., 2021. High and rising economic costs of biological invasions worldwide. *Nature* 592, 571–576. https://doi.org/10.1038/s41586-021-03405-6.

Elton C S. *The ecology of invasions by animals and plants.* In: Springer Nature, 2020.

Chen, B., Sun, Y., Han, Z., et al., 2020. Challenges in preventing and controlling invasive alien species in China. *J. Biosafety* 29 (03), 157–163. https://doi.org/10.3969/j.issn.2095-1787.2020.03.001.

Seebens, H., Bacher, S., Blackburn, T.M., et al., 2021. Projecting the continental accumulation of alien species through to 2050. Glob. Chang. Biol. 27, 970–982. https://doi.org/10.1111/gcb.15333.

Doherty, T.S., Glen, A.S., Nimmo, D.G., et al., 2016. Invasive predators and global biodiversity loss. Proc. Natl. Acad. Sci. 113, 11261–11265. https://doi.org/10.1073/pnas.1602480113.

Xu, H., Qiang, S., Han, Z., et al., 2004. The distribution and introduction pathway of alien invasive species in China. *Biodivers. Sci.* 626–638. CNKI:SUN:SWDY.0.2004-06-007.

Pimentel, D., Zuniga, R., Morrison, D., 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecol. Econ.* 52, 273–288. https://doi.org/10.1016/j.ecolecon.2004.10.002.

Roura-Pascual, N., Leung, B., Rabitsch, W., et al., 2021. Alternative futures for global biological invasions. *Sustain. Sci.* 16, 1637–1650. https://doi.org/10.1007/s11625-021-00963-6.

Shigesada, N., 1997. *Biological Invasions: Theory and Practice*. Oxford University Press.

Malavasi, M., Carboni, M., Cutini, M., et al., 2014. Landscape fragmentation, land-use legacy and propagule pressure promote plant invasion on coastal dunes: a patch-based approach. Landsc. Ecol. 29, 1541–1550. https://doi.org/10.1007/s10980-014-0074-3.

Jiang, J., 2018. Analysis of invasion of alien plants in China. *Value Eng.* 37, 289–293. https://doi.org/10.14018/j.cnki.cn13-1085/n.2018.21.125.

Gilbert, M., Grégoire, J.C., Freise, J.F., et al., 2004. Long-distance dispersal and human population density allow the prediction of invasive patterns in the horse chestnut leafminer Cameraria ohridella. *J. Anim. Ecol.* 73, 459–468. https://doi.org/10.1111/j.0021-8790.2004.00820.x.

Robinet, C., Van Opstal, N., Baker, R., et al., 2011. Applying a spread model to identify the entry points from which the pine wood nematode, the vector of pine wilt disease, would spread most rapidly across Europe. Biol. Invasions 13, 2981–2995. https://doi.org/10.1007/s10530-011-9983-0.

Hulme, P.E., 2021. Unwelcome exchange: International trade as a direct and indirect driver of biological invasions worldwide. *One Earth* 4, 666–679. https://doi.org/10.1016/j.oneear.2021.04.015.

Guisan, A., Petitpierre, B., Broennimann, O., et al., 2014. Unifying niche shift studies: insights from biological invasions. *Trends Ecol. Evol.* 29, 260–269. https://doi.org/10.1016/j.tree.2014.02.009.

Hui, C., 2023. The dos and don'ts for predicting invasion dynamics with species distribution models. *Biol. Invasions* 25, 947–953. https://doi.org/10.1007/s10530-022-02976-3.

Elith, J., Graham*, H.C., Anderson, P.R., et al., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x.

Deneu, B., Servajean, M., Bonnet, P., et al., 2021. Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. PLoS Comput. Biol. 17, e1008856. https://doi.org/10.1371/journal.pcbi.1008856.

Sherpa, S., Després, L., 2021. The evolutionary dynamics of biological invasions: a multi-approach perspective. Evol. Appl. 14, 1463–1484. https://doi.org/10.1111/eva.13215.

Liu, T., Yang, J., Zhou, R., et al., 2023. Prediction of spread risk of pine wilt disease based on geographic raster variables and machine learning models. J. Zhejiang a&f Univ. 40, 617–626. https://doi.org/10.11833/j.issn.2095-0756.20220470.

Robinet, C., Roques, A., Pan, H., et al., 2009. Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One* 4, e4646.

Ferrari, J.R., Preisser, E.L., Fitzpatrick, M.C., 2014. Modeling the spread of invasive species using dynamic network models. *Biol. Invasions* 16, 949–960. https://doi.org/10.1007/s10530-013-0552-6.

Jones, C.M., Jones, S., Petrasova, A., et al., 2021. Iteratively forecasting biological invasions with PoPS and a little help from our friends. *Front. Ecol. Environ.* 19, 411–418. https://doi.org/10.1002/fee.2357.

Petrasova, A., Gaydos, D.A., Petras, V., et al., 2020. Geospatial simulation steering for adaptive management. *Environ. Model. Softw.* 133, 104801. https://doi.org/10.1016/j.envsoft.2020.104801.

Hao, Z., Huang, J., Li, X., et al., 2022. A multi-point aggregation trend of the outbreak of pine wilt disease in China over the past 20 years. *For. Ecol. Manage.* 505, 119890. https://doi.org/10.1016/j.foreco.2021.119890.

Ning, T., Fang, Y., Tang, J., et al., 2004. Advances in research on Bursaphelenchus xylophilus and its key vector Monochamus spp. *Entomol. Knowl.* 97–104. https://doi.org/10.3969/j.issn.0452-8255.2004.02.001.

Togashi, K., Shigesada, N., 2006. Spread of the pinewood nematode vectored by the Japanese pine sawyer: modeling and analytical approaches. *Popul. Ecol.* 48, 271–283. https://doi.org/10.1007/s10144-006-0011-7.

Ye, J., 2019. Epidemic status of pine wilt disease in china and its prevention and control techniques and counter measures. *Scientia Silvae Sinicae* 55, 1–10. https://doi.org/10.11707/j.1001-7488.20190901.

Montandon, L.M., Small, E.E., 2008. The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI. *Remote Sens. Environ.* 112, 1835–1845. https://doi.org/10.1016/j.rse.2007.09.007.

Yue, J., Tian, Q., Dong, X., et al., 2020. Using broadband crop residue angle index to estimate the fractional cover of vegetation, crop residue, and bare soil in cropland systems. *Remote Sens. Environ.* 237, 111538. https://doi.org/10.1016/j.rse.2019.111538.

Newton, P., Castle, S., Kinzer, A., et al., 2022. *The number of forest-and tree-proximate people*. FAO.

FAO T. The state of the world's forests 2022. Forest pathways for green recovery and building inclusive, resilient and sustainable economies. *Rome, FAO*, 2022.

Jung, J.-M., Yoon, S., Hwang, J., et al., 2024. Analysis of the spread distance of pine wilt disease based on a high volume of spatiotemporal data recording of infected trees. For. Ecol. Manage. 553, 121612. https://doi.org/10.1016/j.foreco.2023.121612.

Zheng, Y., Liu, P., Shi, Y., et al., 2021. Difference analysis on pine wilt disease between Liaoning Province of northeastern China and other epidemic areas in China. *J. Beijing Forest. Univ.* 43, 155–160. https://doi.org/10.12171/j.1000−1522.20200108.

Hao, Z., Fang, G., Huang, W., et al., 2022. Risk prediction and variable analysis of pine wilt disease by a maximum entropy model. *Forests* 13. https://doi.org/10.3390/f13020342.

Liu, C., Newell, G., White, M., et al., 2018. Identifying wildlife corridors for the restoration of regional habitat connectivity: a multispecies approach and comparison of resistance surfaces. *PLoS One* 13, e0206071. https://doi.org/10.1371/journal.pone.0206071.

Phillips, S.J., Dudík, M., Schapire, R.E., 2004. A maximum entropy approach to species distribution modeling. In: *Proceedings of the Proceedings of the Twenty-First International Conference on Machine Learning*. https://doi.org/10.1145/1015330.1015412.

Dimitrijevic, D.D., Bacic, J., 2013. Mathematical analysis of dynamic spread of Pine Wilt disease. *Commun. Agric. Appl. Biol. Sci.* 78, 389–399.

Zhao B G, Futai K, Sutherland J R, et al. Pine wilt disease. In: Springer, 2008.

Riascos, A.P., Mateos, J.L., 2021. Random walks on weighted networks: a survey of local and non-local dynamics. *J. Complex Networks* 9, cnab032. https://doi.org/10.1093/comnet/cnab032.

Clauset, A., Shalizi, C.R., Newman, M.E.J., 2009. Power-law distributions in empirical data. *SIAM Rev.* 51, 661–703. https://doi.org/10.1137/070710111.

Jiang, S., Wu, H., Li, D., et al., 2022. Analysis on disaster characteristics of pine wood nematode in Northeast China. *Forest Pest and Disease* 41, 9–15. https://doi.org/10.19688/j.cnki.issn1671-0886.20220012.

Xie B Y, Cheng X Y, Shi J, et al. The formation and expansion mechanism of the invasive population of pine wood nematode–Progress of the National Key Basic Research and Development Program "Basic Research on the Invasion Mechanism and Control of Dangerous Agricultural and Forestry Biology. *ScienceinChinaSeriesC:life Sciences*, 2009, 39: 333-341. doi: CNKI:SUN:JCXK.0.2009-04-001.

Wang C. Study on the Harm of Pine Wood Nematode Disease and Comprehensive Control Countermeasures. *Anhui Agricultural Science Bulletin*, 2015, 21: 108-109. doi: CNKI: SUN: AHNB.0.2015-08-049.

Zhou, H., 2019. A Review of Major Prevention and Control Technologies for Pine Wilt Disease in China. *Jilin Agriculture* 94–95. https://doi.org/10.14025/j.cnki.jlny.2019.04.048.

Back, M.A., Bonifácio, L., Inácio, M.L., et al., 2024. Pine wilt disease: A global threat to forestry. *Plant Pathol.* 73, 1026–1041. https://doi.org/10.1111/ppa.13875.

Ye, J., Wu, X., 2022. Research progress of pine wilt disease. *Forest Pest and Disease* 41, 1–10. https://doi.org/10.19688/j.cnki.issn1671-0886.20220026.