Journal of Plant Pathology (2024) 106:1697–1713
https://doi.org/10.1007/s42161-024-01704-1

ORIGINAL ARTICLE

Modelling the front dynamics of invasive plant pathogens through the analysis of spatial gradients

Guglielmo Lione[1,2] · **Marianna Giraudo**[1] · **Paolo Gonthier**[1,2]

Received: 30 December 2023 / Accepted: 16 July 2024 / Published online: 1 October 2024

Abstract
Variables of phytopathological interest correlated to the impact of plant diseases, such as incidence and severity, may display a spatial pattern resulting from an underlying, yet unknown gradient. Along the main direction of the gradient the variable assessed at the site level either increases, or decreases. Spatial gradients may also arise because of the movement of a front of invasion, an imaginary moving contour separating areas already infested by a plant pathogen from those still pathogen-free. Adequate geostatistical tools may shed light on gradients directional properties, as well as on the direction the front of invasion is coming from or moving to. Tools currently available for that may be impractical due to the advanced computational and programming skills required for their application. Hence, the goals of this study were: (I) to develop, test and validate a new user-friendly geostatistical tool named DirGrad (*Dir*ection of *Grad*ient) aimed at analyzing spatial gradients resulting from the impact of plant diseases; (II) to build an algorithm able to run DirGrad on R, one the most widespread open source software for statistics; and (III) to apply DirGrad for the *ex post* modelling of the invasion front dynamics. The designed algorithm was successfully validated both in silico and in the field by using data from real case studies such as those of the invasive fungal pathogens *Heterobasidion irregulare* and *Ophiostoma novo-ulmi* in a forest stand of Central Italy and across the Swedish island of Gotland, respectively. The algorithm is released as a user-friendly open-source script.

Keywords Algorithm · Biological invasions · Epidemiology · Monte Carlo method · Theoretical plant pathology

Introduction

Biological invasions following entry, establishment and spread of alien plant pathogens are on the rise as a consequence of massive international and intercontinental trade of plant commodities, climate change, and human activity altering ecosystems structure and functioning (Bradley et al. 2010; Liebhold et al. 2017; IPPC Secretariat 2021). The last two factors may also lead to the resurgence of endemic diseases caused by native plant pathogens by contributing to their further spread and by increasing their impact, geographic distribution and host range (Anderson et al. 2004; Raza and Bebber 2022).

Once plant pathogens invade newfound areas, the spatial progression of the disease may be represented by an imaginary moving contour (i.e. invasion front) dividing the space in two regions. In the first region, the pathogen is present and the disease is affecting the invaded ecosystems, while in the other the pathogen is still absent (Azimzade et al. 2020). Although for native plant pathogens the term "invasion front" may be technically improper to define their local spatial spread, the same concept of a moving contour separating colonized and non-colonized areas may apply. Indeed, within the overall borders of their current distribution areas, native pathogens may occupy new areas or spread across ecosystems previously not infested (Labbé et al. 2017; Numminen and Laine 2020). Hence, in the frame of this study, the notion of invasion front will be used in a broad sense, including both alien and native pathogens.

✉ Guglielmo Lione
guglielmo.lione@unito.it

[1] Department of Agricultural, Forest and Food Sciences (DISAFA), University of Torino, Largo Paolo Braccini 2, I-10095 Grugliasco, Italy

[2] Interdepartmental Centre for Innovation in the Agro-environmental Sector (AGROINNOVA), University of Torino, Largo Paolo Braccini 2, I-10095 Grugliasco, Italy

The real-time mapping of the invasion front is possible by conducting phytosanitary surveys over successive years across a target geographic area, in combination with field samplings followed by the diagnosis of the pathogen and the quantification of its abundance and impact (Kottelenberg et al. 2021). The displacement of the invasion front may also be assessed through the follow-up of the disease symptom progression analyzed with the support of aerial or satellite imagery acquired over time series, or at given time-points (Prospero and Cleary 2017). However, the real-time mapping of the invasion front displacement may be hardly achievable due to practical constraints, and often the invasion history of a given plant pathogen needs to be reconstructed *ex post* from existing records, provided that they are available. For instance, first/new reports usually include information about the site coordinates and the date of detection of plant pathogens, thus providing the data needed to appraise their progressive spread across a region. Just to mention some relevant examples, real-time mapping or *ex post* reconstruction of the invasion history of plant pathogens were applied to different case studies targeting:

- the bacterium *Xylella fastidiosa* subsp. *pauca* in Apulia (Italy), the causal agent of the quick decline syndrome of olive (*Olea europaea* L.) during the early 2000' (Kottelenberg et al. 2021);
- the fungal ascomycete *Cryphonectria parasitica* (Murrill) M.E. Barr in USA, the chestnut blight pathogen leading to the virtual extinction of American chestnut (*Castanea dentata* (Marsh.) Borkh.) at the beginning of the 1900' (Gravatt 1949);
- *Hymenoscyphus fraxineus* (T. Kowalski) Baral, Queloz & Hosoya, an invasive ascomycete causing dieback on ash (*Fraxinus excelsior* L.) since the first years of the 1990' in Europe, where it has triggered one of the most detrimental epidemics in forest ecosystems (Timmermann et al. 2011; Queloz et al. 2017);
- *Ophiostoma novo-ulmi* Brasier, a fungus responsible of the Dutch elm disease, associated with high mortality rates on *Ulmus* spp. in North America and Europe since the first half of the 1900'. The invasion of *O. novo-ulmi* across the pathogen-free zone of the Gotland island (Sweden) was monitored extensively since the onset of the first disease foci in 2005 until 2013, when most of the territory was infested (Menkis et al. 2015).

The reconstruction of the spatial and temporal dynamics of the invasion front may shed light not only on the history of biological invasions or epidemic resurgences, but it may also allow to make predictions on possible future scenarios (Harwood et al. 2011). Modelling the spread of both alien and native plant pathogens is a key topic for theoretical plant pathologists (Van Maanen and Xu 2003; Cunniffe et al. 2015). Not surprisingly, an increasing body of literature is devoted to the development of geostatistical tools to analyze invasion front dynamics (Azimzade et al. 2020; Kottelenberg et al. 2021). However, a relevant issue hampering the modelling of the invasion front is represented by the lack of spatial and temporal explicit data about the presence/absence of the pathogen (i.e. the data used to model the spread of *X. fastidiosa*, *C. parasitica*, *H. fraxineus* and *O. novo-ulmi* in the previously cited studies). Nonetheless, in many cases, the progressive colonization of new ecosystems by plant pathogens may result in a gradient, namely a spatial trend of the pathogen's incidence and impact increasing along a given direction (Gregory 1968). Although many biological, ecological, epidemiological and evolutionary factors may influence the spread of plant diseases (Garbelotto et al. 2010), sites where the pathogen's arrival and establishment occurred earlier are those more likely to display higher levels of disease-related impact. For instance, Gonthier et al. (2014) reported a reconstruction of the invasion history of *Heterobasidion irregulare* Garbel. & Otrosina, a basidiomycete accidentally introduced in the 1940' from US and currently invasive in central Italy, where it causes root rots and wood decay in conifer stands, in association with high mortality rates of its hosts. This reconstruction was based on previous studies connecting possible introduction pathways (Gonthier et al. 2004), the current distribution area of *H. irregulare* (Gonthier et al. 2007), and the allelic richness of the pathogen's population (Garbelotto et al. 2013). Noteworthy, in sites were *H. irregulare* had established first, canopy gaps associated with host mortality were substantially larger, diminishing their average extent towards the putative location of the invasion front (Gonthier et al. 2014). Comparable findings were obtained from a study targeting a completely different model system and spatial scale (Parker et al. 1997). Following artificial inoculations conducted in a tomato (*Solanum lycopersicum* L.) field, the foliar pathogen *Septoria lycopersici* Speg. induced symptoms of leaf spot characterized by a spatial and temporal severity gradient (Parker et al. 1997). Indeed, disease severity raised with time, while its magnitude displayed a decreasing trend at increasing distances from the inoculation point (Parker et al. 1997).

If spatial gradients resulting from the impact of plant diseases arise, adequate geostatistical tools may shed light on the directional properties of their underlying cause, such as the density of the pathogen's infectious inoculum, the occurrence of suitable environmental conditions, as well as the movement of the front of invasion (Gregory 1968; Campbell and Noe 1985). Although the notion of gradient is well-known in calculus, its assessment may be impractical since it may involve vector algebra, manipulation of partial

differential equations, or complex spatial interpolations, requiring the support of specialized proprietary software or high-level skills in computing and programming (Mitchell 2005, 2013; Crawley 2013; Oliver and Webster 2014; Awange and Paláncz 2016; Lione and Gonthier 2016). In addition, plant pathologists analyzing disease gradients may be often interested in finding the answer to a basic question: *which is the likely overall compass direction of the gradient, putatively resulting from the movement of the front of invasion?* Hence, the goals of this study were: (I) to develop, test and validate a new user-friendly geostatistical tool named DirGrad aimed at analyzing spatial gradients resulting from the impact of plant diseases assessed in the form of continuous variables (e.g. incidence, severity, yield losses, and so on); (II) to build an algorithm able to run DirGrad on one the most widespread open source software available for statistics (i.e. R); and (III) to apply and validate DirGrad for the *ex post* modelling of the invasion front dynamics.

Materials and methods

Design of the DirGrad general algorithm

The general algorithm of DirGrad (*Dir*ection of *Grad*ient) was designed based on a series of biological and mathematical assumptions and conditions, as described below. For a given number N of sites S indexed by the subscript integer i ranging from 1 to N, a continuous quantitative variable of phytopathological interest λ is assessed at the site level. For each site, the corresponding spatial Cartesian coordinates $(x,\ y)$ in a projected metric system are known. Since the reference system translation is invariant for the purpose of this algorithm, without loss of generality, site coordinates are set to values $x \geq 0$ and $y \geq 0$. An arbitrary vertical line perpendicular to the x-axis is defined outside the x-range of sites and located towards the positive x semi-axis, as follows: $x_E = max\left(\sqrt{x_i^2 + y_i^2}\right) + 0.5\sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N-1}}$. In the previous equation $max\left(\sqrt{x_i^2 + y_i^2}\right)$ is the maximal distance separating sites coordinates from the origin of the reference Cartesian system $(0,0)$, while $0.5\sqrt{\frac{\sum_{i=1}^{N}\left(x_i - \bar{x}\right)^2}{N-1}}$ is half the standard deviation of the sites coordinates (Crawley 2013). The front of invasion is assumed linear in shape, parallel to x_E (i.e. with generic equation $x = x_F$), and progressing with time from x_E towards the origin of the reference system, so that $x_F < x_E$ and, for any $\Delta t > 0$, $x_{F(t+\Delta t)} < x_{F(t)}$.

New coordinates (x', y') of sites are calculated resulting from the anticlockwise rotation of the reference system based on the angle α, according to the classical equations $x' = xcos\alpha + ysin\alpha$ for the x-axis and $y' = -xsin\alpha + ycos\alpha$ for the y-axis. Although the spatial system of the Cartesian coordinates is arbitrary, without loss of generality, the spatial directions corresponding to angles α are labelled as octants, following the windrose (i.e. compass) direction as follows: $\alpha = 0°$ (East, E); $\alpha = 45°$ (Northeast, NE); $\alpha = 90°$ (North, N)); $\alpha = 135°$ (Northwest, NW); $\alpha = 180°$ (West, W); $\alpha = 225°$ (Southwest, SW); $\alpha = 270°$ (South, S); and $\alpha = 315°$ (Southeast, SE). The above angles expressed in sexagesimal degrees correspond to radians $0;\ \frac{\pi}{4};\frac{\pi}{2};\frac{3}{4}\pi\ ;\pi\ ;\frac{5}{4}\pi\ ;\frac{3}{2}\pi$; and $\frac{7}{4}\pi$.

It is assumed that for a given, yet unknown angle α^* the distance $d = |x' - x_E|$ determines the onset of a linear decreasing gradient of the variable of phytopathological interest λ in the form $\lambda(d) = md + q$, with the constraint $m < 0$ and $q = -m \cdot\ max(d_i)$. The gradient is also implicitly time-dependent, resulting from the passage of the moving front of invasion of the pathogen across the location of each site. Hence, it is also assumed that λ is an unknown differentiable function f of time (t), for which the property $\frac{\partial f(d,t)}{\partial t} > 0$ applies. In other terms, if the front of invasion $x_{F(t)}$ of a plant pathogen moving from direction α^* crossed a site s_j prior than another site s_k, this implies that $d_j < d_k$ and that $\lambda(d_j) > \lambda(d_k)$, thus implying a monotonically decreasing pattern of the spatial gradient.

The core of the DirGrad method is finding which unknown octant α^* determined the onset of the gradient of the variable of phytopathological interest λ, thus reconstructing the direction from which the front of invasion is more likely to come from. Since the gradient $\lambda(d)$ is a linear and decreasing function of d, the Pearson's linear correlation coefficient R (Crawley 2013) calculated between λ and d is expected to be minimal and negative for the value of α^* from which the front of invasion is coming from (i.e. R_{α^*}), while maximal and positive at radians $\alpha^* + \pi$ where the front is moving towards (i.e. $\mathrm{R}_{\alpha^*+\pi}$). Since, by definition, d varies depending on the initial arbitrary location of x_E, the calculation of R is standardized and made comparable among different sets of sites S by ranking the values of d (Conover and Iman 1981). The value of α that minimizes the Pearson's linear correlation coefficient R between the observed values of λ and $rank(d)$ is the outcome of the algorithm assessing the most likely value of α^*. The gradient directions are consequently assigned by accounting for the sign and the magnitude of the index R, as follows: the direction of a decreasing gradient from an octant O_1 ($R < 0$) to the opposite O_2 ($R > 0$) is expressed as $O_1 \rightarrow O_2$ ($\mp R$) and, similarly, the direction of an increasing gradient from an

octant O_1 ($R>0$) to the opposite O_2 ($R<0$) is expressed as $O_1 \rightarrow O_2$ ($\pm R$).

The geostatistical approach proposed by Lione and Gonthier (2016) was fine-tuned to build an algorithm to assess the performance of the DirGrad fitting. The bootstrap distribution of R was obtained for each octant by sampling with replacement the column-vector λ from the original dataset through 10^4 iterations, running the DirGrad algorithm on all bootstrapped column-vectors λ^* to derive the bootstrap series of R^* values (Tibshirani and Efron 1993; Carsey and Harden 2013; Crawley 2013). From the above bootstrap distributions of R, the 0.05 and 0.95 quantiles were calculated to build a one-tailed test comparing the observed values of R to the corresponding critical thresholds $R_{0.05}$ and $R_{0.95}$ (Tibshirani and Efron 1993; Carsey and Harden 2013; Crawley 2013; Lione and Gonthier 2016). Positive R values were compared with $R_{0.95}$, while negative ones were contrasted with $R_{0.05}$, retaining R_{α^*} and $R_{\alpha^*+\pi}$ if fulfilling the conditions $R_{\alpha^*} < R_{0.05}$ and $R_{\alpha^*+\pi} > R_{0.95}$ (Tibshirani and Efron 1993; Carsey and Harden 2013; Crawley 2013; Lione and Gonthier 2016).

The resulting DirGrad algorithm was embedded in a user-friendly R language script in order to be available as open source tool for plant pathologists (Lione and Gonthier 2016; R Core Team 2023).

In silico application

The DirGrad algorithm was tested in silico by conducting three separate runs: the first was conducted by applying the algorithm to a simulated spatial gradient complying with the underlying assumptions of the model, the second run was carried out on a gradient violating such assumptions, while the third run was performed on a gradient meeting only part of the assumptions. Hence, the goal of the first runs was to verify whether DirGrad could successfully reconstruct the most likely compass direction of the gradient, putatively resulting from the movement of a linear front of invasion. Conversely, the goal of the second run was to check if the algorithm could exclude the presence of a linear front of invasion in case the front was circular. Finally, the third run was aimed at testing if the DirGrad algorithm may be more robust against violations of linearity than of monotonicity in the increasing/decreasing pattern. All single-run applications of the DirGrad algorithm were conducted through the R language script (see section "Results") that was compiled by following the theoretical approach outlined in the previous section. Spatial gradients constructed to conduct the three runs were built as described below.

In the first run, a linear decreasing gradient of the variable of phytopathological interest was built with the equation of $\lambda(d)$ parametrized by setting $m = -3$ and plotting the resulting xy graph of the function within an x range set from 0 to 10^3. The above equation was used to simulate a front of invasion coming from East and moving towards West E→W ($\mp R$). Overall, $N = 20$ sites were randomly located in a Cartesian plane by drawing the corresponding coordinates from a continuous uniform distribution (Carsey and Harden 2013). The seeds for the single run simulation (Carsey and Harden 2013) were fixed at 16,393 for x and 16,394 for y. A subsequent translation of the above coordinates was conducted by adding a positive constant attaining twice the minimum value extracted for the x-and y-coordinate, respectively. To simulate a $\lambda(d)$ gradient from East (0°) to West (180°), the equation following the application of $max\left(\sqrt{x_i^2+y_i^2}\right)+0.5\sqrt{\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2}{N-1}}$ for computing x_E was used, and the corresponding distances d were calculated accordingly prior to their inclusion in the $\lambda(d)$ equation. The resulting values of λ were assigned to the simulated site locations. Sites were labelled numerically with numbers increasing with increasing values of d. Subsequently, for the preselected values of α corresponding to each compass octant, site coordinates rotation was calculated along with the new corresponding distances from x_E, while original λ values assigned were maintained unaltered. At any rotation, the underlying gradient was rendered by using an ordinary spatial kriging interpolation fitted with the spherical option at 7 lags and 300 pixels raster resolution, corresponding to 8 λ heatmaps (Awange and Paláncz 2016; Olmedo 2022). A visual assessment of the raster heatmaps was conducted to check the consistency between the kriging interpolation and the orientation of the simulated underlying gradient decreasing from East to West. The consistency check was based on the principle that, by construction, the heatmap color ramp from right to left must correspond to the decreasing gradient associated to the angle α set for the coordinate system rotation. The Pearson's linear correlation coefficients R between λ and the ranked distances gathered from each value of α were calculated. The correctness of outcome resulting from the DirGrad algorithm was quantitatively assessed by verifying that the value of α^* minimizing R corresponded to the $\lambda(d)$ gradient from East to West set to conduct the simulation (i.e. minimum and negative R at 0°, maximum and positive R at 180°). The performance of the DirGrad fitting was assessed by testing whether the decreasing gradient E→W ($\mp R$) fulfilled the condition $R_{\alpha^*} < R_{0.05}$.

In the second run, a circular spatial gradient violating the assumptions of linearity and with a pattern not monotonically increasing/decreasing along any given direction was built prior to the application of the DirGrad algorithm. The circular spatial gradient was built from the probability density of a bivariate normal distribution parametrized with

$\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$ (Krishnamoorthy 2006) and cropped within a squared Cartesian window with the x and y range set from −3 to 3. A total of 50 points were selected by randomly drawing their Cartesian coordinates (x, y) from a uniform distribution with lower and upper limits set to −3 and 3, respectively (Carsey and Harden 2013). The 50 corresponding values of the probability density obtained from the bivariate normal distribution were assigned to λ. The spatial gradient and the point locations were rendered and analyzed with the DirGrad algorithm as described for the previous run.

In the third run, a circular spatial gradient was built by following the same process described for run 2, but only the sector included in the first Cartesian quadrant and 13 point locations were considered. This gradient violated the assumptions of linearity (i.e. the front was circular and the pattern decreased non-linearly, deriving from a Gaussian curve), was monotonically decreasing along any given direction departing from the axes center (i.e. S→N, SW→NE, W→E), but not from NW→SE or SE→NW where decreasing/increasing monotonicity was not met.

In silico validation

DirGrad was validated in silico through a series of Monte Carlo simulations (Carsey and Harden 2013; Lione and Gonthier 2016; Lione et al. 2016) conducted by mirroring the approach described for the first run application of the algorithm, although with some adjustments described below.

The single scenario with the equation of $\lambda(d)$ parametrized by setting $m = -3$ was iterated for 10^4 random and independent repetitions, each one drawing a different set of $N = 20$ sites. At each repetition, the linear gradient $\lambda(d)$ direction was maintained from East (0°) to West (180°). A pre-allocation matrix was built to store the labels corresponding to each directional octant (column-wise), and the values obtained from the Pearson's linear correlation coefficients R between λ and the ranked distances at each repetition (row-wise). At the end of the Monte Carlo simulation the above matrix was pivoted by averaging the R values column-wise and calculating the associated 95% Bias Corrected and accelerated confidence intervals ($_{95\%BCa}$CI) (DiCiccio and Efron 1996), whose bounds were obtained from 10^4 bootstrap resamplings as described in Lione et al. (2021). The correctness of outcome resulting from the Monte Carlo simulation of the DirGrad algorithm was appraised by checking that the barplot of the R averages displayed the lowest negative bar for the East octant column, and the symmetric highest and positive bar for the opposite direction (i.e. West octant column). Identical Monte Carlo simulations were replicated with the same number of iterations, by reparametrizing the equation of $\lambda(d)$ by setting $m = -5 \cdot 10^{-1}$, $m = -5 \cdot 10^{-2}$, $m = -5 \cdot 10^{-3}$, and $m = -5 \cdot 10^{-14}$.

In field validation on case studies

The DirGrad algorithm was validated on two case studies documenting the spread of invasive plant pathogens and the spatial progression of their invasion front over the years.

The first case study deals with the colonization of a 3030 ha forest stand in Central Italy (Sabaudia Forest, Circeo National Park) by the alien fungal pathogen *Heterobasidion irregulare*, well established in that site mostly in association with its main host *Pinus pinea* L. (Gonthier et al. 2012). Since its introduction in 1944 in the Castelporziano area (Rome), *H. irregulare* has spread by means of spores southward invading pine stands along the Tyrrhenian coast and has reached the Sabaudia forest, located approximately 70 km far from Castelporziano (Garbelotto et al. 2013). The presence and abundance of *H. irregulare* were assessed by monitoring the airborne inoculum of the pathogen through a spore trapping assay ten years apart (i.e. in 2006 and 2016) (Garbelotto et al. 2022). A total of 33 sampling points were established across the Sabaudia forest, and the spore deposition rate (DR, spores·m^{-2}·h^{-1}) of *H. irregulare* was quantified for each of such points in both time periods. The DirGrad model was validated by testing whether the algorithm could detect the most likely direction of the gradient resulting from the movement of the front of invasion of *H. irregulare* in the Sabaudia forest. The DirGrad algorithm was run on data published in Garbelotto et al. (2022), by using as input variables the sampling point coordinates and their associated DR of *H. irregulare* assessed in 2016 (i.e. λ).

The second case study deals with the progression of the invasion front of *Ophiostoma novo-ulmi* affecting *Ulmus minor* Mill. across the Swedish island of Gotland, whose area covers approximately 3000 km^2 (Menkis et al. 2015). The first detection of *O. novo-ulmi* in Gotland dates back in 2005 (Menkis et al. 2015). From 2005 to 2013, monitoring was conducted on a yearly basis and the presence/absence of *O. novo-ulmi* was assessed and mapped across the whole island (Menkis et al. 2015). Since DirGrad was not designed to cope with binary variables (e.g. presence/absence) and georeferred data about the incidence of *O. novo-ulmi* were not available, the latter were appraised as follows. The distribution map of the presence/absence of *O. novo-ulmi* in 2013 published in Menkis et al. (2015) was acquired and load into WebPlotDigitizer version 4.7 to extract point coordinates, whose

unit was set at raster-image pixel-level (Rohatgi 2024). Fifty random points were selected and a buffer analysis was conducted by counting the number of positive and negative records of *O. novo-ulmi* within a circular radius of 10 units from each point. The incidence of the pathogen (i.e. λ) for each point was calculated as the ratio (%) between the positive records and the total number of records included within each buffer. The DirGrad algorithm was run by using as input variables the sampling point coordinates and their associated incidence of *O. novo-ulmi*.

Results

Script of the DirGrad algorithm

The general algorithm of DirGrad resulted in the following user-friendly script in R language. Each line of the algorithm (in italics) can be copy/pasted as such to the R console, by following the instructions provided for data preparation (point 1) and the explanatory comments added to clarify the code (points 2–15). The algorithm section to assess the performance of the DirGrad fitting is reported and commented from point 16 to 20.

1. Prepare the relevant data storing them in a .csv spreadsheet file with 4 columns titled "id_site", "x", "y", "phytovar". The first column represents the sites labels, the second and the third the sites coordinates, the fourth the value of the variable of phytopathological interest scored at the site level (e.g. disease incidence, severity, and so on). The file should be coded in MS-DOS format (i.e. column separator set as ";" and decimal place separator at ".") and look as shown in Online Resource 1.
2. Load and attach the data into the R console:
   data<-read.table(file.choose(),header=T, sep= ";",dec= ".")
   s_mat<-as.matrix(data)
3. Calculate the current value of :
   *xE<-max(sqrt(s_mat[,"x"]^2+ s_mat[,"y"]^2))+0.5*sd(s_mat[,"x"])*
4. Calculate the current value of the distance and embed it within the data matrix:
   dist_E<-abs(s_mat[,"x"]-xE)
   s_mat<-cbind(s_mat, dist_E)
   s_mat<-s_mat[order(s_mat[,"dist_E"], decreasing=FALSE),]
5. Vector listing the compass directions analyzed:
   dir<-c("E","NE","N","NW","W","SW","S","SE")
6. Vector listing the radians corresponding to each compass direction:
   *alpha<-c(0,pi/4,pi/2,(3/4)*pi, pi, (5/4)*pi, (3/2)*pi, (7/4)*pi)*
7. Vector listing the sexagesimal degrees corresponding to each compass direction:
   degrees<-c("0°","45°","90°","135°","180°","225°","270°","315°")
8. Pre-allocation vector of the Pearson's linear correlation coefficients (R):
   R_Pearson<-numeric(8)
9. Starting the loop assessing the 8 compass octants:
   for (i in 1:8) {
10. Rotation of the reference system and re-calculation of site coordinates and distances from :
    *x_rot<-s_mat[,'x']*cos(alpha[i]) + s_mat[,'y']*sin(alpha[i])*
    *y_rot<--s_mat[,'x']*sin(alpha[i]) + s_mat[,'y']*cos(alpha[i])*
    dist_E_rot<-abs(x_rot-xE)
11. Retrieval of data excluded from the reference system rotation (i.e. the site label and the variable of phytopathological interest)
    id_site_original<-s_mat[,'id_site']
    phytovar_original<-s_mat[,'phytovar']
12. Building of the new data matrix following the coordinate system rotation:
    s_mat_rot<-cbind(id_site_original, x_rot, y_rot, phytovar_original, dist_E_rot)
    s_mat_rot<-s_mat_rot[order(s_mat_rot[,'dist_E_rot'], decreasing = FALSE),]
    s_mat_rot
13. Pearson's linear correlation coefficient R calculated after each rotation of the reference system:
    R_Pearson[i]<-cor.test(s_mat_rot[,'phytovar_original'], rank(s_mat_rot[,"dist_E_rot"]))$estimate

14. Ending the loop started at point 9:
    }
15. Resulting output listing for each octant the corresponding R values:
    data.frame(dir, alpha, degrees, R_Pearson)
16. This section of the algorithm starts the process to conduct the bootstrap tests assessing the performance of the DirGrad fitting. The user can specify the number of bootstrap samples to draw from the column-vector present in the original dataset. The number of samples is set by default to 10^4. A pre-allocation matrix is built to store the bootstrap distribution of the R^* values for each octant direction:
    NBOOT<-10,000
    Rsim = matrix(nrow = 8,ncol = NBOOT)
    rownames(Rsim) = c("E","NE","N","NW","W","SW","S","SE")
17. The loop to iterate the bootstrap resampling of the column-vector is initiated:
    VARIABILE = data$phytovar
    for (k in 1:NBOOT)
    {
    phytovar = sample(VARIABILE, length(VARIABILE), replace = T)
18. The DirGrad algorithm described in points 2–15 is run iteratively on each bootstrap sample:
    id_site = data$id_site
    x = data$x
    y = data$y
    data<-data.frame(id_site, x,y, phytovar)
    s_mat<-as.matrix(data)
    *xE<-max(sqrt(s_mat[,"x"]^2 + s_mat[,"y"]^2)) + 0.5*sd(s_mat[,"x"])*
    dist_E<-abs(s_mat[,"x"]-xE)
    s_mat<-cbind(s_mat, dist_E)
    s_mat<-s_mat[order(s_mat[,"dist_E"], decreasing = FALSE),]
    dir<-c("E","NE","N","NW","W","SW","S","SE")
    *alpha<-c(0,pi/4,pi/2,(3/4)*pi, pi, (5/4)*pi, (3/2)*pi, (7/4)*pi)*
    degrees<-c("0°","45°","90°","135°","180°","225°","270°","315°")
    R_Pearson<-numeric(8)
    for (i in 1:8) {.
    *x_rot<-s_mat[,'x']*cos(alpha[i]) + s_mat[,'y']*sin(alpha[i])*
    *y_rot<--s_mat[,'x']*sin(alpha[i]) + s_mat[,'y']*cos(alpha[i])*
    dist_E_rot<-abs(x_rot-xE)
    id_site_original<-s_mat[,'id_site']
    phytovar_original<-s_mat[,'phytovar']
    s_mat_rot<-cbind(id_site_original, x_rot, y_rot, phytovar_original, dist_E_rot)
    s_mat_rot<-s_mat_rot[order(s_mat_rot[,'dist_E_rot'], decreasing = FALSE),]
    s_mat_rot
    R_Pearson[i]<-cor.test(s_mat_rot[,'phytovar_original'], rank(s_mat_rot[,"dist_E_rot"]))$estimate
    }
19. The bootstrapped series of R^* values are stored in the pre-allocation matrix and the loop is ended:
    Rsim[, k] = R_Pearson
    }
20. The critical thresholds $R_{0.05}$ and $R_{0.95}$ are calculated for each octant direction:
    inf = function(vect) {quantile(vect, prob = 0.05)}
    sup = function(vect) {quantile(vect, prob = 0.95)}
    apply(Rsim,1,inf)
    apply(Rsim,1,sup)

A script without comments ready to be copied/pasted in block into the R console is provided as supplementary material (Online Resource 2).

In silico application of the DirGrad algorithm

For the first run application of the DirGrad algorithm, the construction of the decreasing gradient from East (0°) to West (180°) of the variable of phytopathological interest $\lambda(d)$ resulted in a linear function whose trend decreased with increasing values of the distance d, consistently with the parametrization $m = -3$ (Fig. 1).

The 20 sites randomly located in a Cartesian plane showed coordinates x ranging from 4.082 to 12.401, and y from 1.437 to 10.488. The value of x_E attained 17.442, and was external to the cloud of site locations $(x_E > \max(x_i))$. The distance d separating sites from x_E achieved a minimum of 5.041 and a maximum of 13.360. The assignment of the λ values to sites was consistent with the gradient applied, attaining the highest value (24.957) in site 1, which was the closest to x_E, while the lowest value (0) was reached in site 20, the furthest to x_E. At increasing distances from x_E, the sites displayed decreasing values of λ from East to West (Fig. 2).

The visual assessment of the ordinary spatial kriging interpolation indicated that the underlying gradient was mainly oriented from East to West (corresponding to octant E and rotation $\alpha = 0°$). Conversely, all other rotations

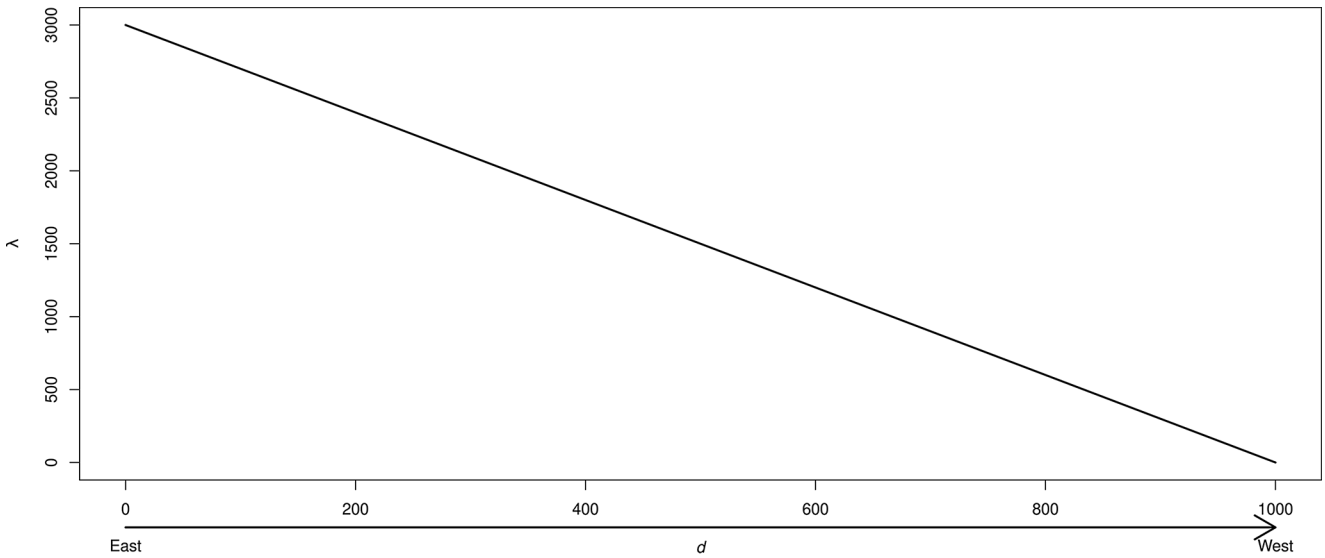

Fig. 1 Linear decreasing gradient from East to West (see arrow) of the variable of phytopathological interest $\lambda(d)$ parametrized with $m = -3$ as a function of the distance d (adimensional units)

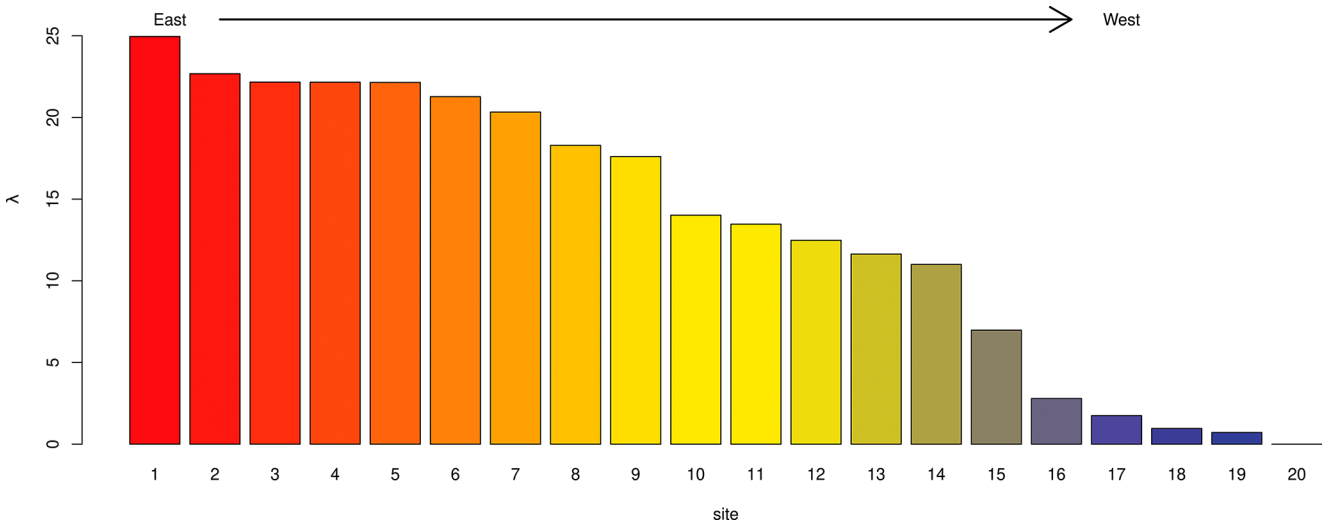

Fig. 2 Values of the variable of phytopathological interest λ displayed by sites simulated to run the DirGrad algorithm. In the barchart, sites are labelled from 1 to 20 based on their location from East to West, as indicated by the arrow representing the gradient direction. Warmer colors are used to score higher levels of λ

($\alpha = 45°, 90°, 135°, 180°, 225°, 270°, 315°$) did not produce raster images whose heatmaps could display a right-to-left color ramp consistent with the simulated gradient direction set by construction (Fig. 3).

The outcome of the DirGrad algorithm was consistent with the results obtained from the visual assessment of the raster heatmaps produced through the kriging spatial interpolation. DirGrad was able to successfully reconstruct the direction of the gradient set by construction by simulating a front of invasion coming from East and moving towards West. Indeed, the value of α^* minimizing Pearson's linear correlation coefficient R between the observed values of λ and $rank(d)$ was 0°, leading to the minimum and negative $R = -0.979$, while the corresponding $\alpha^* + \pi$ (180°) resulted in the maximum and positive R value of 0.979 (i.e. front of invasion movement E→W, ∓ 0.979). Such R values fulfilled the conditions $R_{\alpha^*} < R_{0.05}$ and $R_{\alpha^*+\pi} > R_{0.95}$, with $R_{0.05} = -0.382$ and $R_{0.95} = 0.382$. All R values produced by the other rotation angles α were either higher than −0.979, if negative, or lower than 0.979, if positive, thus indicating that other underlying gradient directions were less likely than the East-West one (Table 1).

For the second run application of the DirGrad algorithm, the construction of a circular spatial gradient violating the assumptions of linearity and with a pattern not monotonically increasing/decreasing along any given direction resulted in a raster with highest probability density values close to the origin of the Cartesian system, circularly decreasing towards the borders of the squared window, as shown by the corresponding kriging interpolation (Fig. 4). The 50 sites randomly located in a Cartesian plane displayed coordinates x ranging from −2.869 to 2.973, and y from −2.724 to 2.886, with probability density assigned to λ included between $3.867 \cdot 10^{-5}$ and $1.500 \cdot 10^{-1}$. The second run of the DirGrad algrorithm excluded the presence of a linear front of invasion when the front shape was circular. Indeed, all octants displayed Pearson's linear correlation coefficients R close to 0, with values between $\pm 3.42 \cdot 10^{-2}$ and $\pm 7.88 \cdot 10^{-2}$, and none of the R values fulfilled the conditions $R_{\alpha^*} < R_{0.05}$ and $R_{\alpha^*+\pi} > R_{0.95}$ since $R_{0.05} < -0.230$ and $R_{0.95} > 0.230$.

The third run application of the DirGrad algorithm resulted in a pattern circularly decreasing towards the borders of the first quadrant of the Cartesian plane, as shown by the corresponding kriging interpolation (Fig. 5). The outcome of the algorithm showed the presence of three significant linear monotonic decreasing gradients SW→NE (∓ 0.896), W→E (∓ 0.727), and S→N (∓ 0.640) that fulfilled the conditions $R_{\alpha^*} < R_{0.05}$ and $R_{\alpha^*+\pi} > R_{0.95}$, with $R_{0.05}$ scoring at −0.468, −0.474, −0.472 and $R_{0.95}$ at −0.468, −0.474, −0.472, respectively. Instead, NW→SE (∓ 0.151) and SE→ NW (± 0.151) were not scored as significant with critical values having absolute values over 0.45.

In silico validation of the DirGrad algorithm

The DirGrad algorithm was successfully validated through the in silico experiment conducted by using Monte Carlo simulations. The iterations carried out with the equation of $\lambda(d)$ parametrized by setting $m = -3$ led to results similar to those outlined in the single run described in the previous section. The barplot of the Monte Carlo averages obtained for the Pearson's linear correlation coefficient displayed the lowest negative bar for the East octant column ($R = -0.981$), and the symmetric highest and positive bar ($R = -0.981$) for the opposite direction (i.e. West octant column) (Fig. 6). Comparable results were obtained from the Monte Carlo simulations carried out when reparametrizing the equation of $\lambda(d)$ by setting $m = -5 \cdot 10^{-1}$, $m = -5 \cdot 10^{-2}$, $m = -5 \cdot 10^{-3}$, and $m = -5 \cdot 10^{-14}$. Hence, the in silico validation confirmed that DirGrad could successfully reconstruct the direction of the gradient set by construction by simulating a front of invasion coming from East and moving towards West.

In field validation on case studies

The application of the DirGrad algorithm to the case study reported in Garbelotto et al. (2022) showed that the front of invasion of *H. irregulare* at the stand scale moved more likely from NE to SW (NE→SW, ∓ 0.430). The conditions $R_{\alpha^*} < R_{0.05}$ and $R_{\alpha^*+\pi} > R_{0.95}$ were fulfilled, with $R_{0.05} = -0.289$ and $R_{0.95} = 0.289$. Although with milder R values, also the direction from N to S (N→S, ∓ 0.371) was significant, fulfilling the above conditions with $R_{0.05} = -0.295$ and $R_{0.95} = 0.295$. Conversely, all other octants did not display significant R values (Table 2).

The run of the DirGrad algorithm to the case study described in Menkis et al. (2015) was applied on λ values ranging from 0 to 95%, with an average of 54%. The outcome of DirGrad showed that *O. novo-ulmi* progressed across the island of Gotland through a front of invasion likely moving from N to S (N→S, ∓ 0.628). The conditions $R_{\alpha^*} < R_{0.05}$ and $R_{\alpha^*+\pi} > R_{0.95}$ were fulfilled, with $R_{0.05} = -0.235$ and $R_{0.95} = 0.235$. Other significant directions displaying lower R values were NE→SW (∓ 0.563) and NW→SE (∓ 0.598), while the remaining directions did not fulfill the conditions $R_{\alpha^*} < R_{0.05}$ and $R_{\alpha^*+\pi} > R_{0.95}$ (Table 3).

Discussion

Reliable statistical and geostatistical tools are required to model the spatial and temporal dynamics of diseases affecting crops, forests and ornamentals (Cunniffe et al. 2015). In particular, explicit spatial modelling is of paramount importance not only

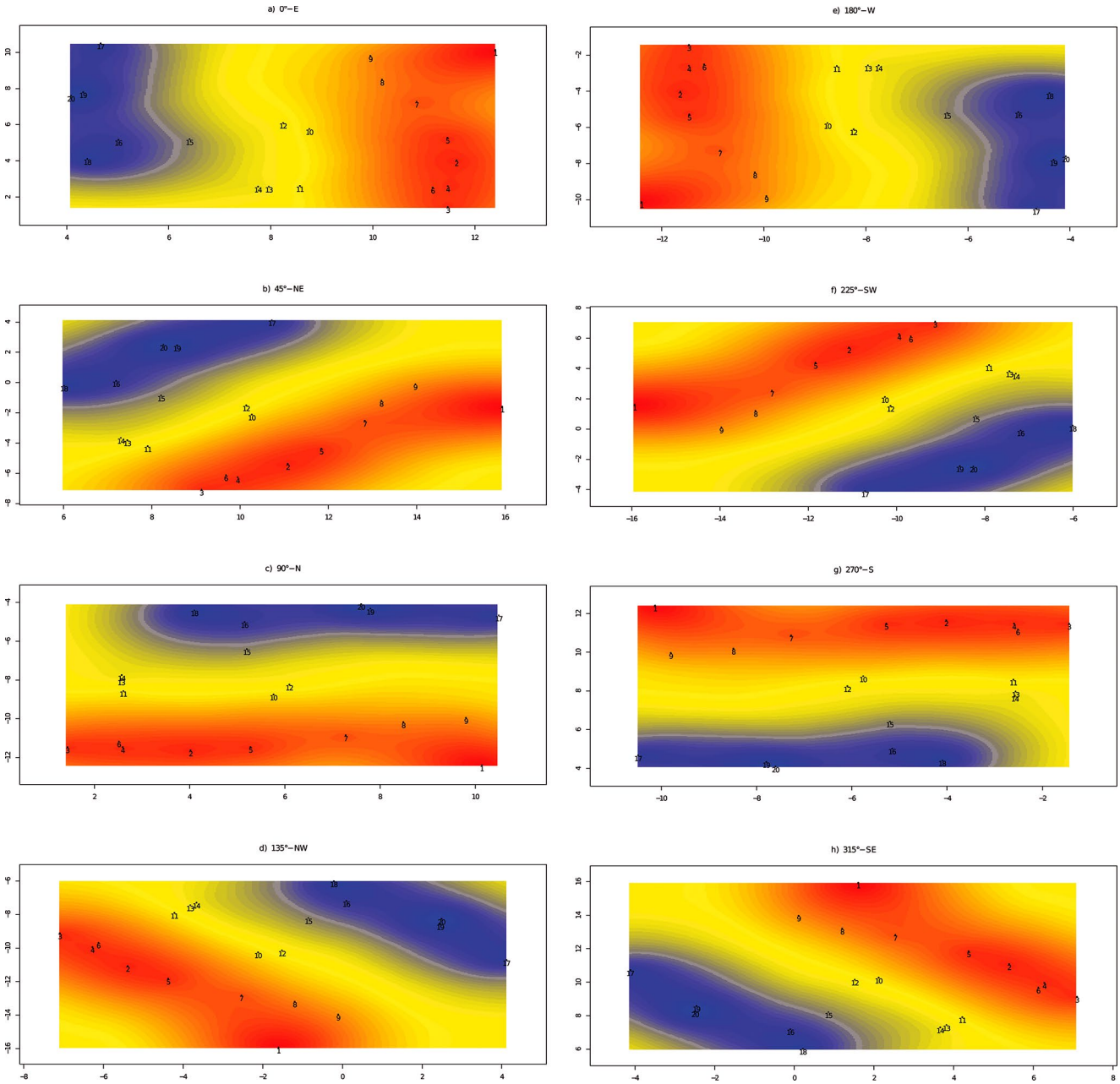

Fig. 3 Maps of the sites whose location was simulated for the first run application of the DirGrad algorithm and corresponding heatmaps of the ordinary spatial kriging interpolation of the variable of phytopathological interest λ . Sites are labelled from 1 to 20 based on the original simulated gradient decreasing linearly from East to West by construction (corresponding to octant E and rotation $\alpha = 0°$) (panel a). The panels from b) to h) correspond to the rotations of the reference system based on angles α other than 0°, which are indicated on the top of each panel in sexagesimal degrees, along with the corresponding compass directions expressed as octants. Warmer colors of the heatmap are associated with higher values of λ . The color ramp from warm to cold colors indicates a decreasing gradient of λ . Spatial coordinates are in adimensional units

to investigate the spread of invasive alien plant pathogens, but also to better understand the resurgence patterns of native ones (Anderson et al. 2004; Raza and Bebber 2022). Models based on geostatistical approaches are constantly implemented in the framework of computational biology, numerical ecology and quantitative risk assessment applied to plant pathology, with the ultimate goal of supporting more practical activities promoting plant health (Lione and Gonthier 2016; Gilioli et al. 2017; Garrett et al. 2018). The latter may include the optimization of strategies for phytosanitary monitoring, the implementation of more efficient decision-making processes in plant health policy, and the design and application of more timely and effective approaches for disease control and management (Benninga et al. 2012; Lione et al. 2017; Rossi et al. 2010).

Table 1 Outcome of the first run of the DirGrad algorithm applied to a simulated linearly decreasing spatial gradient

Compass octant	α (sexagesimal degrees)	α (radians)	R	Gradient direction	Critical R
E	0°	0	−0.979	α^*	−0.382
NE	45°	0.785	−0.632	-	−0.377
N	90°	1.571	0.212	-	0.388
NW	135°	2.356	0.780	-	0.383
W	180°	3.142	0.979	$\alpha^* + \pi$	0.382
SW	225°	3.927	0.632	-	0.377
S	270°	4.712	−0.213	-	−0.388
SE	315°	5.498	−0.780	-	−0.383

For each compass octant and related angle α expressed in sexagesimal degrees (°) and radians, the corresponding Pearson's linear correlation coefficient R is reported. The direction of the gradient reconstructed by the algorithm (i.e. the direction from which the front of invasion is more likely to come from) is indicated as α^*, while the opposite direction (the direction towards which the front of invasion is more likely to move) is labelled as $\alpha^* + \pi$. The first angle results in the lowest and negative R value, the second to the highest and positive one. The critical values $R_{0.05}$ and $R_{0.95}$ are reported for negative and positive R coefficients, respectively

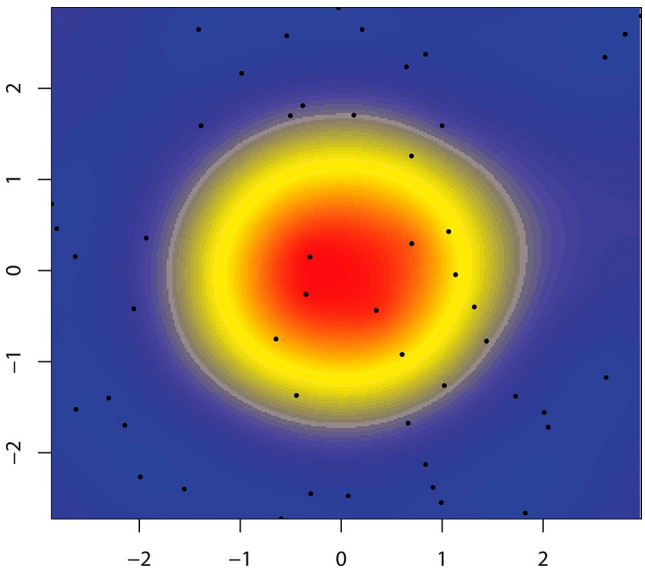

Fig. 4 Maps of the sites whose location was simulated for the second run application of the DirGrad algorithm and corresponding heatmap of the ordinary spatial kriging interpolation of the variable of phytopathological interest λ. The values of λ are drawn from a bivariate normal distribution centered in the origin of the Cartesian window to simulate a circular spatial gradient. Warmer colors of the heatmap are associated with higher values of λ. The color ramp from warm to cold colors indicates decreasing values of λ. Spatial coordinates are in adimensional units

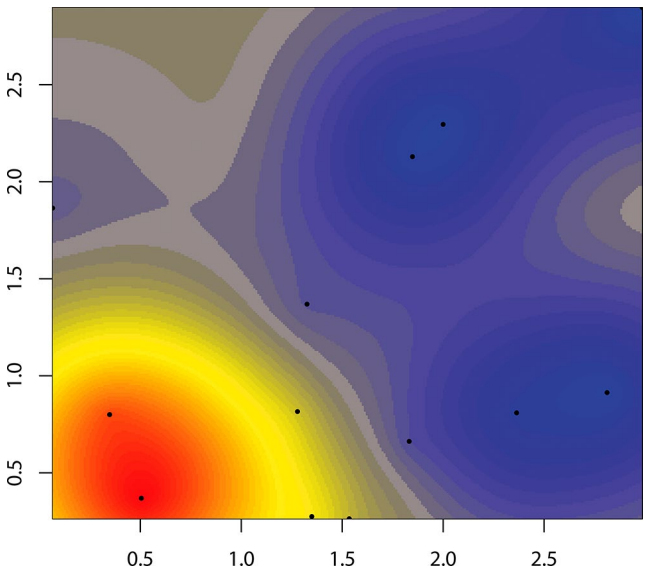

Fig. 5 Maps of the sites whose location was simulated for the third run application of the DirGrad algorithm and corresponding heatmap of the ordinary spatial kriging interpolation of the variable of phytopathological interest λ. The values of λ are drawn from a bivariate normal distribution centered in the origin of the Cartesian window to simulate a circular sector spatial gradient cropped within the first quadrant of the Cartesian plane. Warmer colors of the heatmap are associated with higher values of λ. The color ramp from warm to cold colors indicates decreasing values of λ. Spatial coordinates are in adimensional units

Methods to model plant disease dynamics and impact can take advantage from several deterministic or stochastic techniques that can be used to appraise spatial and temporal patterns of plant diseases for different host-pathogen-environment combinations (for a comprehensive overview see, for instance, Van Maanen and Xu 2003; Garrett et al. 2004; Hothorn et al. 2006; Kéry 2010; Carsey and Harden 2013; Crawley 2013; Lahcene 2013; Cunniffe et al. 2015; Dobrow 2016; Lione and Gonthier 2016; Lantz 2019; Thompson and Brooks-Pollock 2019). Nonetheless, the application of modern modelling techniques may require specific competencies and skills in advanced mathematics, statistics, computational biology, GIS science and programming (Mitchell 2005, 2013; Crawley 2013; Oliver and Webster 2014; Awange and Paláncz 2016; Lione and Gonthier 2016). Hence, the availability of cutting-edge, yet user-friendly computational tools could help plant pathologists to test relevant hypotheses about the biology, ecology and epidemiology of pathogens, based on results gathered from geostatistics (Lione and Gonthier 2016).

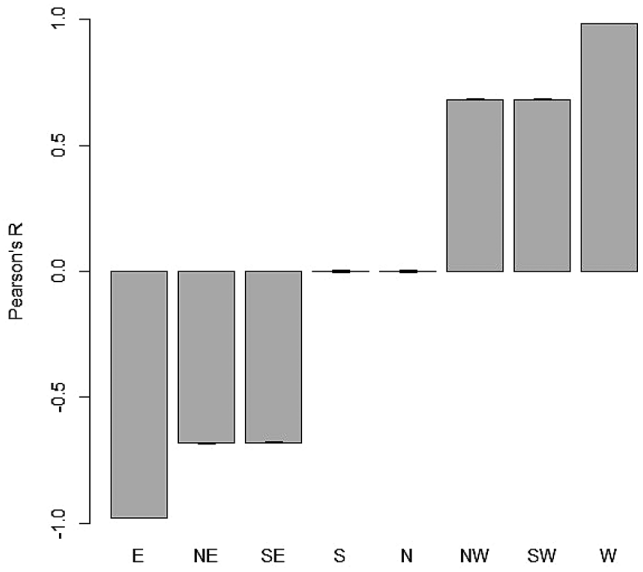

Fig. 6 Barplot of the Pearson's linear correlation coefficients R averages obtained from Monte Carlo simulations conducted over 10^4 iterations. Whiskers display the bounds of the associated 95% Bias Corrected and accelerated confidence intervals ($_{95\%BCa}$CI). R averages are shown for each octant, with the lowest negative bar for the East octant, and the symmetric highest positive bar for the West one

Modelling the front of invasion of alien pathogens, or the front of colonization of native ones infesting new areas within their current distribution range (in this study both referred to as "front of invasion") is a key challenge in epidemiology (Kottelenberg et al. 2021). Three main aspects need to be elucidated when modelling the front of invasion: its shape, movement rate and direction (Kottelenberg et al. 2021). Through extensive phytosanitary monitoring, real-time mapping of the pathogen's spread can provide huge amounts of data to build accurate models unravelling all the above aspects (Menkis et al. 2015; Kottelenberg et al. 2021), yet several practical constraints may hamper the feasibility of this approach, including the lack of adequate economic resources (Coulston et al. 2008). In the absence of a real-time mapping, the need of *ex post* reconstruction of the invasion front dynamics may arise, hence the opportunity of developing user-friendly geostatistical tools supporting plant pathologists in this task.

In this study we designed and tested DirGrad (*Dir*ection of *Grad*ient), an algorithm we specifically constructed to assess the overall spatial direction of the gradient determined by a variable of phytopathological interest associated with the impact of a plant disease. Such quantitative variable might be the incidence of the target pathogen, the severity of the disease on its hosts, an estimate of the economic losses determined by the onset of an epidemic, and so on.

DirGrad was designed to analyze data resulting from the assessment of a variable of phytopathological interest in different sites, for which the spatial coordinates are known. If an underlying increasing or decreasing gradient affects the spatial pattern of the variable along a given direction, DirGrad is able to identify such direction by providing a series of quantitative likelihood metrics (i.e. values of the Pearson's linear correlation coefficient R) and a criterion for their selection (i.e. minimum/maximum R). More importantly, since spatial gradients may result from the passage of the front of invasion of plant pathogens (Gregory 1968), a key application of the DirGrad algorithm is to reconstruct the prevalent direction of the invasion front. By construction, DirGrad produces as output an estimate of the likelihood that the front may originate from a given compass octant and move towards its opposite (e.g. from North to South N→S, from Northeast to Southwest NE→SW, and so on). One of the core advantages of DirGrad relies on its user-friendly appeal. Indeed, the algorithm can be run by simply copy/pasting the code on the R console (R Core Team 2023), without any additional setting operation needed. The only necessary input is a spreadsheet reporting the names of the sites, their spatial coordinates, and the corresponding value of the variable of interest (e.g. disease incidence,

Table 2 Outcome of the run of the DirGrad algorithm applied to the invasion front of *Heterobasidion irregulare* at Sabaudia Forest

Compass octant	α (sexagesimal degrees)	α (radians)	R	Gradient direction	Critical R
E	0°	0	−0.006	-	−0.285
NE	45°	0.785	−0.431	α^*	−0.289
N	90°	1.571	−0.371	-	−0.295
NW	135°	2.356	−0.245	-	−0.297
W	180°	3.142	0.006	-	0.285
SW	225°	3.927	0.430	$\alpha^* + \pi$	0.289
S	270°	4.712	0.371	-	0.295
SE	315°	5.498	0.245	-	0.297

For each compass octant and related angle α expressed in sexagesimal degrees (°) and radians, the corresponding Pearson's linear correlation coefficient R is reported. The direction of the gradient reconstructed by the algorithm (i.e. the direction from which the front of invasion is more likely to come from) is indicated as α^*, while the opposite direction (the direction towards which the front of invasion is more likely to move) is labelled as $\alpha^* + \pi$. The first angle results in the lowest and negative R value, the second to the highest and positive one. The critical values $R_{0.05}$ and $R_{0.95}$ are reported for negative and positive R coefficients, respectively

Table 3 Outcome of the run of the DirGrad algorithm applied to the invasion front of *Ophiostoma novo-ulmi* in the Swedish island of Gotland

Compass octant	α (sexagesimal degrees)	α (radians)	R	Gradient direction	Critical R
E	0°	0	−0.087	-	−0.232
NE	45°	0.785	−0.563	-	−0.235
N	90°	1.571	−0.628	α^*	−0.235
NW	135°	2.356	−0.598	-	−0.234
W	180°	3.142	0.087	-	0.232
SW	225°	3.927	0.563	-	0.235
S	270°	4.712	0.628	$\alpha^* + \pi$	0.235
SE	315°	5.498	0.598	-	0.234

For each compass octant and related angle α expressed in sexagesimal degrees (°) and radians, the corresponding Pearson's linear correlation coefficient R is reported. The direction of the gradient reconstructed by the algorithm (i.e. the direction from which the front of invasion is more likely to come from) is indicated as α^*, while the opposite direction (the direction towards which the front of invasion is more likely to move) is labelled as $\alpha^* + \pi$. The first angle results in the lowest and negative R value, the second to the highest and positive one. The critical values $R_{0.05}$ and $R_{0.95}$ are reported for negative and positive R coefficients, respectively

severity, impact, etc.). Not secondarily, R is one of the most widespread software for statistics worldwide and it is freely available as open source tool (Crawley 2013). The DirGrad algorithm itself is released as an open-source script, thus it could be prospectively redistributed, modified, and improved, or embedded in other existing statistics or GIS software used by plant pathologists or by a more general audience. Another relevant advantage of DirGrad over other more complex methods (e.g. kriging spatial interpolation, kernel density estimates) is the lack of ambiguity in the outcome (Awange and Paláncz 2016; Lione et al. 2017; Olmedo 2022). Indeed, spatial interpolation resulting from kriging or kernels may vary depending on different settings of the algorithm, including the raster resolution and the arbitrary choice of the interpolating procedure (Awange and Paláncz 2016; Lione et al. 2017; Olmedo 2022). Thus, the resulting raster either needs a visual interpretation, or some further advanced computational analyses (Mitchell 2005, 2013; Crawley 2013; Awange and Paláncz 2016; Lione et al. 2017). Conversely, the DirGrad outcome is rather basic, providing a single list of octants associated with a vector of Pearson's R values, among which the choice is predetermined by a unique selection criterion. The criterion is rather easy to implement because the most likely provenance direction of the decreasing gradient, putatively resulting from the movement of the front of invasion, is indicated by the minimum negative value of R. Since by construction R is a correlation index whose strength is higher for values approaching -1 (for $R<0$) and lower for values close to 0 (Crawley 2013), the magnitude of R is by itself a proxy of the magnitude of the gradient directionality. Nonetheless, to assess the goodness of fit of the model, a one-tailed bootstrap test (Tibshirani and Efron 1993; Carsey and Harden 2013; Crawley 2013) was included in the DirGrad algorithm based on the geostatistical approach proposed by Lione and Gonthier (2016). Such approach allows to test the strength of a spatial pattern by constructing the empirical distribution of the variable of phytopathological interest under the null hypothesis implying the absence of an explicit spatial structure of the variable. Since the DirGrad output is symmetric, meaning that opposite octants result in R value identical in module but opposite in sign, a one-tailed test is likely more appropriate than a two-tailed alternative. In case multiple R values are significant, this would mean that different directions of the gradient are possible, although their likelihood decreases with decreasing modules of the associated R values.

Notwithstanding the above mentioned advantages, as any model, DirGrad relies on theory and assumptions that should be carefully accounted for prior to its application and, with even more emphasis, for the biological interpretation of the outcomes. For these reasons, computational tools need to be validated prior to their release. In our study, DirGrad was successfully validated both in silico on simulated datasets, and on data gathered from real case studies investigating the invasion biology of plant pathogens at different spatial scales (from some thousand hectares to some thousand square kilometers) and involving different pathosystems (a basidiomycete causing root rot and wood decay on conifers and an ascomycete causing a vascular disease on broadleaves).

The most relevant assumptions refer to the mathematical structure of the gradient, assumed to be linear and with a monotonically decreasing spatial pattern. When such assumptions were met in silico, DirGrad proved to be reliable and correctly identified the overall gradient direction, as shown by the results obtained from the first run application and from the subsequent Monte Carlo simulations. Instead, as expected, DirGrad failed to detect the presence of a linear decreasing gradient during the second run application, where the gradient was non-linear and non-monotonically decreasing. Interestingly, in the third run application, where

linearity was violated but monotonicity was not, DirGrad correctly identified three equally likely directions of the simulated decreasing circular gradient. Along the only direction where the gradient was both non-linear and non-monotonic, DirGrad did not display a significant R value. Such results suggest that the model is reliable in the detection of linear and monotonically increasing or decreasing gradients, is robust when linearity is violated, while it cannot detect gradients whose spatial pattern does not meet the monotonicity assumption (i.e. gradients that alternate increasing/decreasing values along a given direction). It seems likely that the rank operator embedded in the DirGrad algorithm increases its robustness (Conover and Iman 1981), minimizing the influence of the non-linearity of the gradient.

The application of DirGrad for the modelling of the front of invasion is based not only on mathematical assumptions, but also on conditions that the front of invasion should fulfill: (I) the front shape is linear; (II) the front and its associated gradients are unique (i.e. there are no multiple fronts or overlapping gradients); and (III) once the front of invasion has crossed the site, the associated variable of phytopathological interest increases in magnitude over time. In the absence of historical records from phytosanitary surveys, it is unknown whether those conditions are met, or not. Applying DirGrad to two case studies documenting the past dynamics of invasion fronts allowed for the validation of the model's performance in the field. Indeed, although the pathosystems and the spatial scales of the two selected case studies were clearly different, the algorithm correctly identified the directions of displacement of the invasion fronts consistently with the scenarios described in the literature based on field evidence (Garbelotto et al. 2022; Menkis et al. 2015). At the regional scale, both *H. irregulare* and *O. novo-ulmi* spread across the invaded ecosystems from single points of introduction, hence a circular rather than a linear gradient would have been expected. However, the front of invasion of *H. irregulare* crossed the Sabaudia forest several km far from the site of first introduction (i.e. approximately 70 km) and at that spatial scale (i.e. a forest stand of some tenth km^2 over an invaded area of several hundred km^2) its shape may well approximate a straight line. In the case of *O. novo-ulmi*, although the available maps (Menkis et al. 2015) show the progression of the Dutch elm disease from a single point of introduction, the shape of the front of invasion is more linear than circular, possibly because of the host spatial distribution pattern. Northward the point of introduction in Gotland, a large gap is visible where no elm trees are present. Hence, the pathogen had likely moved asymmetrically towards the southern sectors of the island where hosts were much more abundant. The host distribution is likely a key factor influencing the spatial progression of the front of invasion. Not surprisingly, *H. irregulare* in the Sabaudia forest showed significant displacements of the front of invasion from NE to SW and from N to S, which are the directories across which large patches of *P. pinea* are present (Gonthier et al. 2012). It should be noted that the validation in silico and on different case studies in the field also suggests that the algorithm can be used to assess gradients at different spatial scales, including the host plant, crop, ecosystem, landscape, regional or intercontinental levels.

In addition to the host distribution, the properties of the front of invasion may depend upon several other factors, such as the landscape fragmentation, the pathogen's population density, its reproductive potential, the means through which the infectious inoculum can be dispersed (e.g. wind, water, insect vectors) and the distance it can reach, the effects exerted on the pathosystem by environmental variables (e.g. climate) and agricultural/silvicultural/land management practices, the carrying capacity of the invaded ecosystems, the presence of multiple points of introduction and physical or biological barriers (Plantegenest et al. 2007; Mundt et al. 2009; Azimzade et al. 2020). Nonetheless, based on in silico and on field case-studies validations, DirGrad is expected to perform adequately when: (I) in the region under investigation, the pathogen establishment and spread occurred from a single point; (II) the newly colonized area is rather homogeneous (i.e. lack of excessive fragmentation of the spatial distribution of suitable habitats, uniform management practices, and absence of additional biotic or abiotic random disturbances); and (III) the disease progress led to symptoms whose severity or impact increased over time. In particular, the latter point deserves a special attention. For instance, root rots and vascular diseases caused by pathogens spreading within a short range by means of root contacts or grafts between neighboring host plants are likely to fulfill this assumption. In relation to the case study referring to the invasion of *H. irregulare* in Central Italy, mortality centers and canopy gaps present in the invaded ecosystems have been documented to enlarge over time (Gonthier et al. 2014). At the field level, the impact of a foliar disease may be likely decreasing with increasing distance from the inoculum source, while augmenting its severity over time (Parker et al. 1997) due to the rate of colonization of the host tissues by the pathogen, or because of the presence of re-infection cycles. Conversely, if chemical control treatments are carried out to control the disease, the resulting spatial gradient of the severity at the field level might depend on the orientation of the treated plots and on the applied dose of the product, rather than on the movement of the front of invasion, as shown by Franke and Menz (2007) targeting powdery mildew (*Blumeria graminis* (DC.) Speer) and leaf rust (*Puccinia recondita* Roberge ex Desm.) pathogens on wheat.

It is worth noting that, as for any other model or statistical test, the application of DirGrad may be instrumental to

highlight that some of the underlying model assumptions are not met (Acutis et al. 2012; Crawley 2013), thus obtaining clues to shed light on some peculiar trait of the pathosystem under investigation. For instance, if the magnitudes of the Pearson's correlation coefficient R resulting from DirGrad are comparable among different octants, this might suggest that the pathogen was introduced from multiple points. Since gradients are vector-based mathematical objects, their direction obeys to the rules of vector algebra, hence a two-fold front of invasion from North and from East might lead to a higher R value corresponding to a single front coming from Northeast, just to cite an example. Also the lack of significant R values may suggest that some of the DirGrad underlying conditions are violated, indicating that: (I) no gradient is present; or (II) the gradient is not monotonic. In the last case, either the gradient is non-linear, or multiple introduction points of the pathogen are leading to alternating increasing/decreasing patterns of the gradient. Since the assumptions of the model cannot be tested a priori, the blind application of DirGrad (and of other similar geostatistical tools) is not recommended. Indeed, while the detection of a gradient and its direction may be successfully assessed by using an algorithm, the expert advice from a plant pathologist and the knowledge of the target pathosystem are pivotal to decide how to run the model and to interpret the results and their biological meaning. For instance, if a pathogen is suspected to spread unevenly under some circumstances leading to multiple disease gradients, the operator could run the algorithm on a series of spatial scales and interpret the outcomes accordingly. Indeed, while at local scale the gradient may look multiple due to site-specific conditions, at regional scale it might turn unique and display a single consistent pattern across different sites. Similarly, the algorithm outcome may be influenced by a high landscape fragmentation or by the presence of spatial gradients in the host distribution. In such a case, comparing the host abundance, disease impact and other environmental-related gradients through DirGrad may shed light on the underlying causes of the observed spatial patterns and unravel spatial-mediated correlations interrelating the host-pathogen-environment interactions. As an example, if the gradient of rainfall is directed concordantly with that of disease impact, it is likely that rainfall may be a relevant climatic variable explaining the pathogen's spread.

Although we did not test DirGrad on biological invasions of organisms other than plant pathogens, such as alien plants or insects pests, in principle no constraints should arise when the algorithm applications are extended to other model-systems. This holds true on condition that the resulting spatial gradients are deemed comparable to those resulting from invasive pathogens. It is worth noting that the case study on *O. novo-ulmi* used to successfully validate DirGrad refers to a fungal pathogen whose transmission can occur both through root grafts and insect vectors (in the case study, *Scolytus multistriatus* Marsham), suggesting that the geostatistical tool developed and presented in this paper could be flexible and suitable for other scientific fields such as entomology and botany.

In conclusion, in this study we designed, tested and successfully validated DirGrad, a new user-friendly geostatistical tool to analyze the directional properties of spatial gradients resulting from the impact of plant diseases. We released DirGrad as open-source algorithm, discussing its application for the modelling of the invasion front dynamics of plant pathogens.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s42161-024-01704-1.

Funding No funding was received for conducting this study.

Data availability All data relevant to this study are embedded within the manuscript or included as supplementary material (Online Resources).

Declarations

Employment Not applicable.

Research involving human participants and/or animals Not applicable.

Informed consent Not applicable.

Competing interests The Authors declare that no financial competing interests are directly or indirectly related to this work. Two out of the three Authors are Editorial Board Members and Editors of the Journal of Plant Pathology, but they did not handle the manuscript or influence in any way its processing.

References

Acutis M, Scaglia B, Confalonieri R (2012) Perfunctory analysis of variance in agronomy, and its consequences in experimental

results interpretation. Eur J Agron 43:129–135. https://doi.org/10.1016/j.eja.2012.06.006
Anderson PK, Cunningham AA, Patel NG, Morales FJ, Epstein PR, Daszak P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. Trends Ecol Evol 19:535–544. https://doi.org/10.1016/j.tree.2004.07.021
Awange JL, Paláncz B (2016) Geospatial algebraic computations. Theory and applications, 3rd edn. Springer, Berlin
Azimzade Y, Sasar M, Maleki I (2020) Invasion front dynamics in disordered environments. Sci Rep 10:18231. https://doi.org/10.1038/s41598-020-75366-1
Benninga J, Hennen WHGJ, Schans J (2012) Supply chain risk model for quantifying the cost-effectiveness of phytosanitary measures. Crop Prot 32:64–70. https://doi.org/10.1016/j.cropro.2011.11.001
Bradley BA, Blumenthal DM, Wilcove DS, Ziska LH (2010) Predicting plant invasions in an era of global change. Trends Ecol Evol 25:310–318. https://doi.org/10.1016/j.tree.2009.12.003
Campbell CL, Noe JP (1985) The spatial analysis of soilborne pathogens and root diseases. Annu Rev Phytopathol 23:129–148. https://doi.org/10.1146/annurev.py.23.090185.001021
Carsey TM, Harden JJ (2013) Monte Carlo simulation and resampling methods for social science. Sage, Los Angeles
Conover WJ, Iman RL (1981) Rank transformations as a bridge between parametric and nonparametric statistics. Am Stat 35:124–129. https://doi.org/10.2307/2683975
R Core Team (2023) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
Crawley MJ (2013) The R book, second edn. Wiley, Chichester
Cunniffe NJ, Koskella B, Metcalf CJE, Parnell S, Gottwald TR, Gilligan CA (2015) Thirteen challenges in modelling plant diseases. Epidemics 10:6–10. https://doi.org/10.1016/j.epidem.2014.06.002
DiCiccio TJ, Efron B (1996) Bootstrap confidence intervals. Stat Sci 11:189–212. https://doi.org/10.1214/ss/1032280214
Dobrow RP (2016) Introduction to stochastic processes with R. Wiley, New York
Franke J, Menz G (2007) Multi-temporal wheat disease detection by multi-spectral remote sensing. Precis Agric 8:161–172. https://doi.org/10.1007/s11119-007-9036-y
Garbelotto M, Linzer R, Nicolotti G, Gonthier P (2010) Comparing the influences of ecological and evolutionary factors on the successful invasion of a fungal forest pathogen. Biol Invasions 12:943–957. https://doi.org/10.1007/s10530-009-9514-4
Garbelotto M, Guglielmo F, Mascheretti S, Croucher PJP, Gonthier P (2013) Population genetic analyses provide insights on the introduction pathway and spread patterns of the north American forest pathogen *Heterobasidion irregulare* in Italy. Mol Ecol 22:4855–4869. https://doi.org/10.1111/mec.12452
Garbelotto M, Lione G, Martiniuc AV, Gonthier P (2022) The alien invasive forest pathogen *Heterobasidion irregulare* is replacing the native *Heterobasidion annosum*. Biol Invasions 24:2335–2349. https://doi.org/10.1007/s10530-022-02775-w
Garrett KA, Madden LV, Hughes G, Pfender WF (2004) New applications of statistical tools in plant pathology. Phytopathology 94:999–1003. https://doi.org/10.1094/PHYTO.2004.94.9.999
Garrett KA, Alcalá-Briseño RI, Andersen KF, Buddenhagen CE, Choudhury RA, Fulton JC, Hernandez Nopsa JF, Poudel R, Xing Y (2018) Network analysis: a systems framework to address grand challenges in plant pathology. Annu Rev Phytopathol 56:559–580. https://doi.org/10.1146/annurev-phyto-080516-035326
Gilioli G, Schrader G, Grégoire JC, MacLeod A, Mosbach-Schulz O, Rafoss T, Rossi V, Urek G, van der Werf W (2017) The EFSA quantitative approach to pest risk assessment – methodological aspects and case studies. EFSA J 8:1495. https://doi.org/10.1111/epp.12377
Gonthier P, Warner R, Nicolotti G, Mazzaglia A, Garbelotto M (2004) Pathogen introduction as a collateral effect of military activity. Mycol Res 108:468–470
Gonthier P, Nicolotti G, Linzer R, Guglielmo F, Garbelotto M (2007) Invasion of European pine stands by a north American forest pathogen and its hybridization with a native interfertile taxon. Mol Ecol 16:1389–1400. https://doi.org/10.1111/j.1365-294X.2007.03250.x
Gonthier P, Lione G, Giordano L, Garbelotto M (2012) The American forest pathogen *Heterobasidion irregulare* colonizes unexpected habitats after its introduction in Italy. Ecol Appl 22:2135–2143. https://doi.org/10.1094/phyto-95-0759
Gonthier P, Anselmi N, Capretti P, Bussotti F, Feducci M, Giordano L, Honorati T, Lione G, Luchi N, Michelozzi M, Paparatti B, Sillo F, Vettraino AM, Garbelotto M (2014) An integrated approach to control the introduced forest pathogen *Heterobasidion irregulare* in Europe. Forestry 87:471–481. https://doi.org/10.1093/forestry/cpu015
Gravatt F (1949) Chestnut blight in Asia and North America. Unasylva 3:2–7. https://doi.org/10.1017/s0953756204240369
Gregory PH (1968) Interpreting plant disease dispersal gradients. Annu Rev Phytopathol 6:189–212. https://doi.org/10.1146/annurev.py.06.090168.001201
Harwood TD, Tomlinson I, Potter CA, Knight JD (2011) Dutch elm disease revisited: past, present and future management in Great Britain. Plant Pathol 60:545–555. https://doi.org/10.1111/j.1365-3059.2010.02391.x
Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat 15:651–674. https://doi.org/10.1198/106186006X133933
IPPC Secretariat (2021) Scientific review of the impact of climate change on plant pests – a global challenge to prevent and mitigate plant pest risks in agriculture, forestry and ecosystems. Rome. FAO on behalf of the IPPC Secretariat. https://doi.org/10.4060/cb4769en
Kéry M (2010) Introduction to WinBUGS for ecologists: bayesian approach to regression, ANOVA, mixed models and related analyses. Academic, London
Kottelenberg D, Hemerik L, Saponari M, van der Werf W (2021) Shape and rate of movement of the invasion front of *Xylella fastidiosa* spp. Pauca Puglia Sci Rep 11:1061. https://doi.org/10.1038/s41598-020-79279-x
Krishnamoorthy K (2006) Handbook of statistical distributions with applications. Chapman and Hall/CRC
Labbé F, Fontaine MC, Robin C, Dutech C (2017) Genetic signatures of variation in population size in a native fungal pathogen after the recent massive plantation of its host tree. Heredity 119:402–410. https://doi.org/10.1038/hdy.2017.58
Lahcene B (2013) On Pearson families of distributions and its applications. Afr J Math Comput Sci Res 6:108–117. https://doi.org/10.5897/AJMCSR2013.0465
Lantz B (2019) Machine learning with R: expert techniques for predictive modeling. Packt publishing ltd., Birmingham
Liebhold AM, Brockerhoff EG, Kalisz S, Nuñez MA, Wardle DA, Wingfield MJ (2017) Biological invasions in forest ecosystems. Biol Invasions 19:3437–3458. https://doi.org/10.1007/s10530-017-1458-5
Lione G, Gonthier P (2016) A permutation-randomization approach to test the spatial distribution of plant diseases. Phytopathology 106:19–28. https://doi.org/10.1094/PHYTO-05-15-0112-R
Lione G, Giordano L, Ferracini C, Alma A, Gonthier P (2016) Testing ecological interactions between *Gnomoniopsis castaneae* and *Dryocosmus kuriphilus*. Acta Oecol 77:10–17. https://doi.org/10.1016/j.actao.2016.08.008
Lione G, Gonthier P, Garbelotto M (2017) Environmental factors driving the recovery of bay laurels from *Phytophthora Ramorum*

infections: an application of numerical ecology to citizen science. Forests 8:293. https://doi.org/10.3390/f8080293
Lione G, Giordano L, Sillo F, Brescia F, Gonthier P (2021) Temporal and spatial propagule deposition patterns of the emerging fungal pathogen of chestnut *Gnomoniopsis castaneae* in orchards of north-western Italy. Plant Pathol 70:2016–2033. https://doi.org/10.1111/ppa.13451
Menkis A, Östbrant IL, Wågström K, Vasaitis R (2015) Dutch elm disease on the island of Gotland: monitoring disease vector and combat measures. Scand J Forest Res 31:237–241. https://doi.org/10.1080/02827581.2015.1076888
Mitchell A (2005) The ESRI guide to GIS analysis – volume 2: spatial measurements and statistics. ESRI, Redlands
Mitchell A (2013) The ESRI guide to GIS analysis – volume 3: modeling suitability, movement, and interaction. ESRI, Redlands
Mundt CC, Sackett KE, Wallace LD, Cowger C, Dudley JP (2009) Long-distance dispersal and accelerating waves of disease: empirical relationships. Am Nat 173:456–466. https://doi.org/10.1086/597220
Numminen E, Laine AL (2020) The spread of a wild plant pathogen is driven by the road network. PLoS Comput Biol 16:e1007703. https://doi.org/10.1371/journal.pcbi.1007703
Oliver MA, Webster R (2014) A tutorial guide to geostatistics: computing and modelling variograms and kriging. CATENA 113:56–69. https://doi.org/10.1016/j.catena.2013.09.006
Olmedo OE (2022) kriging: Ordinary Kriging R package version 1.2. https://cran.r-project.org/web/packages/kriging/kriging.pdf
Parker SK, Nutter FW Jr, Gleaso ML (1997) Directional spread of *Septoria* leaf spot in tomato rows. Plant Dis 81:272–276. https://doi.org/10.1094/PDIS.1997.81.3.272
Plantegenest M, Le May C, Fabre F (2007) Landscape epidemiology of plant diseases. J R Soc Interface 4:963–972. https://doi.org/10.1098/rsif.2007.1114
Prospero S, Cleary M (2017) Effects of host variability on the spread of invasive forest diseases. Forests 8:80. https://doi.org/10.3390/f8030080
Queloz V, Hopf S, Schoebel CN, Rigling D, Gross A (2017) Ash dieback in Switzerland: history and scientific achievements. In: Vasaitis R, Enderle R (eds) Dieback of European ash (*Fraxinus* spp.) –consequences and guidelines for sustainable management. Swedish University of Agricultural Sciences, Uppsala, pp 68–78
Raza MM, Bebber DP (2022) Climate change and plant pathogens. Curr Opin Microbiol 70:102233. https://doi.org/10.1016/j.mib.2022.102233
Rohatgi A (2024) Web Plot Digitizer. https://apps.automeris.io/wpd/. Accessed 2 May 2024
Rossi V, Giosuè S, Caffi T (2010) Modelling plant diseases for decision making in crop protection. In: Oerke EC, Gerhards R, Menz G, Sikora R (eds) Precision crop protection - the challenge and use of heterogeneity. Springer, Dordrecht, pp 241–258. https://doi.org/10.1007/978-90-481-9277-9_15
Thompson RN, Brooks-Pollock E (2019) Detection, forecasting and control of infectious disease epidemics: modelling outbreaks in humans, animals and plants. Philos Trans R Soc B 374:20190038. https://doi.org/10.1098/rstb.2019.0038
Tibshirani RJ, Efron B (1993) An introduction to the bootstrap. CRC Press
Timmermann V, Børja I, Hietala AM, Kirisits T, Solheim H (2011) Ash dieback: pathogen spread and diurnal patterns of ascospore dispersal, with special emphasis on Norway. EPPO Bull 41:14–20. https://doi.org/10.1111/j.1365-2338.2010.02429.x
Van Maanen A, Xu XM (2003) Modelling plant disease epidemics. Eur J Plant Pathol 109:669–682. https://doi.org/10.1023/A:1026018005613