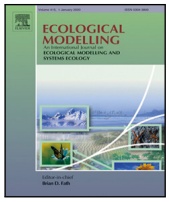
Ecological Modelling 489 (2024) 110630

Contents lists available at ScienceDirect

Ecological Modelling

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



Estimating the reproduction number, R_0, from individual-based models of tree disease spread

Laura E. Wadkin [a,*], John Holden [b], Rammile Ettelaie [b], Melvin J. Holmes [b], James Smith [b], Andrew Golightly [c], Nick G. Parker [a], Andrew W. Baggaley [a]

[a] School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[b] Faculty of Environment, University of Leeds, Leeds, LS2 9JT, UK
[c] Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, UK

ARTICLE INFO

Dataset link: https://doi.org/10.25405/data.ncl.24787752

Keywords:
Tree disease
Reproduction number
Individual-based model
Epidemic model
SIR model
Bayesian inference
Agent-based model

ABSTRACT

Tree populations worldwide are facing an unprecedented threat from a variety of tree diseases and invasive pests. Their spread, exacerbated by increasing globalisation and climate change, has an enormous environmental, economic and social impact. Computational individual-based models are a popular tool for describing and forecasting the spread of tree diseases due to their flexibility and ability to reveal collective behaviours. In this paper we present a versatile individual-based model with a Gaussian infectivity kernel to describe the spread of a generic tree disease through a synthetic treescape. We then explore several methods of calculating the basic reproduction number R_0, a characteristic measurement of disease infectivity, defining the expected number of new infections resulting from one newly infected individual throughout their infectious period. It is a useful comparative summary parameter of a disease and can be used to explore the threshold dynamics of epidemics through mathematical models. We demonstrate several methods of estimating R_0 through the individual-based model, including contact tracing, inferring the Kermack–McKendrick SIR model parameters using the linear noise approximation, and an analytical approximation. As an illustrative example, we then use the model and each of the methods to calculate estimates of R_0 for the ash dieback epidemic in the UK.

1. Introduction

The loss of biodiversity due to the spread of tree diseases and invasive pests within forest ecosystems has an enormous environmental, economic, and social impact worldwide (Freer-Smith and Webber, 2017; Cuthbert et al., 2021). This threat has been escalating rapidly due to increasing globalisation resulting in a greater number of accidental imports and climate change creating a more favourable environment for many pests and pathogens.

Mathematical and computational models are powerful tools for deepening our understanding of the fundamental behaviours of different pests and pathogens, as well as describing and forecasting their future spread (Cornell et al., 2019; Gertsev and Gertseva, 2004; Wang and Song, 2008; Meentemeyer et al., 2011; Cunniffe et al., 2016). Individual-based models, where each individual tree (or group of trees) is represented, with properties governed by a set of probabilistic rules, have the advantage of providing information about the whole system (i.e., the macroscale and collective behaviour) from modelling the individual (microscale) behaviour and are ideal for exploring spatial patterning. Previous compartmental individual-based models of tree disease have considered a square lattice representing susceptible trees with pathogen spread stochastically through nearest neighbour interactions (Orozco-Fuentes et al., 2019). Although suitable for some fungal diseases (Orozco-Fuentes et al., 2019; Goleniewski and Newton, 1994), many tree diseases are spread through airborne pathogens on longer scales and are not suitable for nearest neighbour contact spread (Parnell et al., 2009; Grosdidier et al., 2018). Thus, here we adopt a spatially explicit individual-based model with a dispersal kernel representing the spatial dispersal of the pathogen or pest which results in the probability of infection, similar to those previously used to explore controls for human epidemics (Suprunenko et al., 2021).

The infectivity of a disease can be quantified by the basic reproduction number, R_0, defined as the total number of expected secondary infections arising from one newly infected individual introduced into a fully susceptible population. It helpfully defines the threshold between epidemic ($R_0 > 1$), and containment ($R_0 < 1$), and can be used as a comparative parameter for diseases with differing behaviours and spread mechanisms. In many mathematical models (including the Kermack–McKendrick SIR model considered in this text) R_0 can be used

* Corresponding author.
E-mail address: laura.wadkin@ncl.ac.uk (L.E. Wadkin).

https://doi.org/10.1016/j.ecolmodel.2024.110630
Received 30 June 2023; Received in revised form 15 December 2023; Accepted 16 January 2024
Available online 25 January 2024

to determine the steady-state conditions of the disease or pathogen density in a host population (Suprunenko et al., 2021; Kermack and McKendrick, 1927; Jeger and Van den Bosch, 1994; Segarra et al., 2001; Diekmann and Heesterbeek, 2000; Van den Bosch et al., 2008; Wang and Zhao, 2012; Van den Driessche and Watmough, 2002).

It is possible to calculate R_0 from knowledge of the pathogen's life-cycle and its interactions with the host plant (Van den Bosch et al., 2008), however this information is not always readily available. Previous work estimating R_0 in real-world ecological scenarios has included utilising a spatially-explicit population dynamic model for wheat stripe rust (Mikaberidze et al., 2016), the SIR model with a time-varying infestation rate for the oak processionary moth pest (Wadkin et al., 2022) and stochastic epidemiological landscape models for sudden oak death (Filipe et al., 2012).

For varying implementations of spatially explicit individual-based models, previous work has sought a corresponding analytic approximation to R_0, which can provide quick insights into epidemic properties (Suprunenko et al., 2021; Bolker, 1999; Keeling, 1999; Brown and Bolker, 2004; Filipe and Maule, 2003; North and Godfray, 2017). For an individual-based stochastic compartmental model, similar to the one we consider here, a recent approximation exploits localised invasions at the start of an epidemic to estimate R_0 analytically and to explore the impact of spatial control strategies (Suprunenko et al., 2021). We will consider this approximation, as well as our own derivation of an analytic expression, throughout the manuscript.

In this paper, we present a versatile individual-based model with a Gaussian infectivity kernel (referred to henceforth as the IBM) and explore several methods for estimating the basic reproduction number from this model, including contact tracing, parameter inference for a SIR model and an analytic approximation. We then take an illustrative case-study of the UK ash dieback epidemic to compare these methods. We outline the IBM, the compartmental SIR model with accompanying parameter inference methodology, and the analytic approximation of R_0 in Section 2, present the results in Section 3, and discuss the findings in Section 4.

2. Methods

In this section we introduce the basic reproduction number R_0 (Section 2.1), present the IBM for a generic tree disease (Section 2.2), summarise the stochastic SIR model (Section 2.3), outline the statistical methodology for parameter inference of the stochastic SIR model (Section 2.4), and derive an analytical approximation of R_0 for the IBM (Section 2.5).

2.1. The reproduction number R_0

The reproduction number is a key parameter quantifying the spread of a disease. The basic reproduction number, R_0, describes the expected number of secondary infections produced by one primary infected tree, over their infectious period. Some of the methods presented in Section 3 estimate the basic reproduction number at every time-step of the simulation, which we will refer to as the instantaneous basic reproduction number, $R_0(t)$. We will also consider the instantaneous effective reproduction number, calculated at a time t as $R(t) = R_0 S(t)/S(0)$, which takes into account that as a disease progresses the number of susceptibles, $S(t)$, is decreasing and thus limiting disease transmission.

2.2. The individual-based model

We consider individual trees as points, randomly distributed at a density ρ within a bounded square of $L \times L$. All trees are classified as being in one of three states: susceptible (S), infected (I) or removed ($R^\dagger$) (a compartmental SIR model as described in Section 2.3 Kermack and McKendrick, 1927). Susceptible trees have yet to be infected and are at risk, infected trees currently have the disease and are infective to surrounding susceptible trees during their infectious period T_I, and removed trees have previously been infected, but are no longer infectious. Trees transition through the states $S \to I \to R^\dagger$ with opportunities to transition in every iteration of an arbitrary discrete time-step which can be rescaled to handle different time courses.

We assume that susceptible trees at shorter distances from infected trees are more likely to become infected than those further away. Thus, infection spreads through the trees based on a spatial kernel, allowing a dependence on the distance between a susceptible and an infectious tree. The choice of this kernel is flexible, linking to the spatial dispersal of a pathogen and the probability of infection upon pathogen presence. A Gaussian kernel is commonly used, however, in some ecological cases other kernels may be more appropriate (Nathan et al., 2012; Grosdidier et al., 2018). For simplicity, here we apply a Gaussian kernel to capture the decay of infection probability with increasing distance between trees, but this is easily transferable to any other kernel function. In this case, the probability of a susceptible tree S_i becoming infected due to an infectious tree I_j, separated by a distance r, in an arbitrary time-step of the IBM Δt, is described by the function

$$Pr(S_i \to I_i | I_j) = B \exp\left(\frac{-r^2}{2l^2}\right) \Delta t + o(\Delta t),$$

where B is an infectivity parameter, l is the length scale of the pathogen dispersal and $o(\Delta t)/\Delta t \to 0$ as $\Delta t \to 0$. To make epidemics comparable as the transmission length scale varies, we scale the infectivity parameter, leading to the kernel

$$Pr(S_i \to I_i | I_j) = \frac{b}{2\pi l^2} \exp\left(\frac{-r^2}{2l^2}\right) \Delta t + o(\Delta t), \tag{1}$$

where $\Delta t = 1$ and therefore $b/(2\pi l^2) = B$ takes a value between 0 and 1. Two example kernels of the above form with contrasting length scales are shown in Fig. 1(a).

We also assume that a susceptible tree is more likely to become infected if it is surrounded by a greater number of infectious trees, thus we require the probability that tree S_i becomes infected through any of the infectious trees I_j (with $j = 1, \dots, N_I$) and their respective probability of infecting S_i, i.e., p_{ij}. To avoid a lengthy union calculation, we calculate the probability of S_i remaining uninfected, through $\prod(1-p_{ij})$. The infectious period is set by a fixed parameter T_I. Once a tree has been infectious for T_I time-steps, it transitions into the removed ($R^\dagger$) category.

The simulation begins with an initial number of infected trees, I_0, chosen stochastically from trees closest to the centre of the domain. In one time-step all the possible infections are assessed and the corresponding compartmental transitions are applied. Example snapshots from the model are shown for the two illustrative infectivity kernels with contrasting length scale parameters in Fig. 1(b) and (c).

2.2.1. Contact tracing through the IBM

The most direct way to calculate R_0 in a computational setting is through the contact tracing of secondary infections throughout the simulation. This is straightforward if the disease dynamics are one-to-one, with a tree becoming infected due to a 'contact' with a single infectious tree. However, in the IBM described above, we assume infections can occur due to pressure from multiple sources through the infectious kernels surrounding each infected tree, as shown in Fig. 2.

In this case, after assessing the transition probabilities for all pairwise combinations of susceptible and infected trees, we can assign a resulting number of newly caused infections to each infectious tree that is proportional to the number of trees contributing towards the transition, as illustrated in Fig. 2. For example, if tree S_1 is successfully infected with transition probabilities positively assessed from tree I_1 and I_2, then both tree I_1 and I_2 are deemed to have caused 0.5 secondary infections for the transition of S_1 into the infected category. These secondary infections can then be summed for each I_i after all S_j have been considered. This method conserves the total number

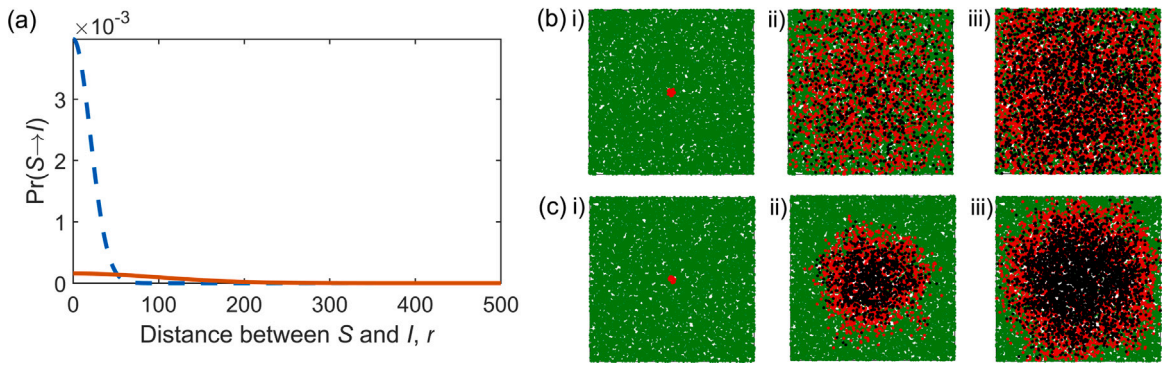


Fig. 1. (a) Two example Gaussian infectivity kernels from (1), with $b = 10$, $l = 100$ (orange, long-range spread) and $b = 10$, $l = 10$ (blue dashed, short-range spread). The resulting infection dynamics are shown in (b) and (c) for the long and short-range spreads respectively, showing snapshots at (i) $t = 0$, (ii) 20% tree mortality and (iii) 50% tree mortality. Both cases begin with $I_0 = 20$ infected trees with an infectious period of $T_I = 10$ time-steps, amongst a population of susceptible trees with density $\rho = 0.05$ in a bounded box of size $L = 500$.

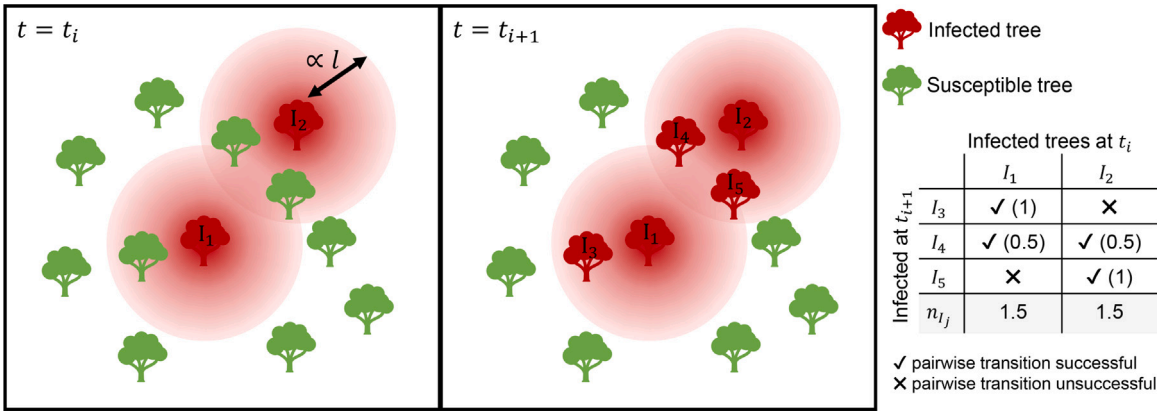


Fig. 2. Schematic illustrating the number of secondary infections, n_{I_j}, caused by two infected trees, I_1 and I_2. The circles around each of I_1 and I_2 at time-step $t = t_i$ illustrate the surrounding infectivity kernel, with a length scale described by parameter l, and infectivity by b. For each tree within the infectious zone, the probability of transition is assessed as described in Section 2.2. For newly infected trees (i.e., I_3, I_4 and I_5), the contributing infection pressures from previous infected trees (i.e., I_1 and I_2) are assessed independently, with proportional responsibility allocated to each previous infectious tree, as shown in the table.

of secondary infections caused, whilst allowing an insight into the individual tree contributions.

We can then estimate $R(t)$ and R_0 through this proxy-contact tracing method. The number of new secondary infections caused by each infected tree in each time-step can be averaged, giving the mean number of new infections in a time-step, $\overline{n_{I_j}}$. The instantaneous effective reproduction number is then $R(t) = \overline{n_{I_j}} T_I$ where T_I is the infectious period. For example, if in a single time-step there were 20 initially infected trees leading to 10 new infections, with a mean infection duration of 10 timesteps then $\overline{n_{I_j}} = 0.5$ and $R(t = 1) = 0.5 \times 10 = 5$, corresponding to the fact that at this stage of the dynamics there are five new infections (on average) from one initial infection over its infectious period. We present examples of this method as an estimate for R_0 in Section 3.1.

2.3. The compartmental SIR model

We consider estimating the parameters for a standard SIR model to describe the IBM output as a method of calculating R_0 (see Section 2.4 for inference details). The established formulation of a compartmental SIR model (Andersson and Britton, 2000; Kermack and McKendrick, 1927) describes the rates of change of the populations in each compartment by

$$\frac{dS}{dt} = -\beta IS, \qquad \frac{dI}{dt} = \beta IS - \gamma I, \qquad \frac{dR^{\dagger}}{dt} = \gamma I,$$

where β describes the rate of infection at which contact of one infected with one susceptible will result in infection due to pathogen dispersal (sometimes referred to as the effective contact rate), and γ describes the rate of removal due to a limited infectious period.

Since the IBM described above is inherently probabilistic, we will consider the more flexible stochastic SIR system (as opposed to the deterministic system) described by the Itô stochastic differential equation (SDE)

$$dX_t = a(X_t, \theta)dt + \sqrt{d(X_t, \theta)}dW_t, \tag{2}$$

where $X_t = (S_t, I_t)'$ is the state of the system at time t, $\theta = (\beta, \gamma)'$ is a vector of parameter values, and $dW_t = (W_{1,t}, W_{2,t})'$ denotes a vector of uncorrelated standard Brownian motion processes. The SDE drift function $a(X_t, \theta)$ and diffusion coefficient $d(X_t, \theta)$ are given by

$$a(X_t, \theta) = \begin{pmatrix} -\beta S_t I_t \\ \beta S_t I_t - \gamma I_t \end{pmatrix} \quad \text{and}$$

$$d(X_t, \theta) = \begin{pmatrix} \beta S_t I_t & -\beta S_t I_t \\ -\beta S_t I_t & \beta S_t I_t + \gamma I_t \end{pmatrix}. \tag{3}$$

A derivation of the above can be found in Fuchs (2013). The drift and diffusion functions are the infinitesimal mean and variance that match the most natural Markov jump process representation of the SIR model (Gillespie, 2000). The conditions under which this leads to a reasonable approximation are also discussed in Gillespie (2000). In this model, the basic reproduction number is given by $R_0 = \beta N/\gamma$. In the next section we outline an inference scheme for estimating plausible values of β and γ (and thus R_0) from the time-series output of the IBM.

Note that other variations of this model exist, allowing flexibility to capture the dynamics of different diseases, such as the SEIRS

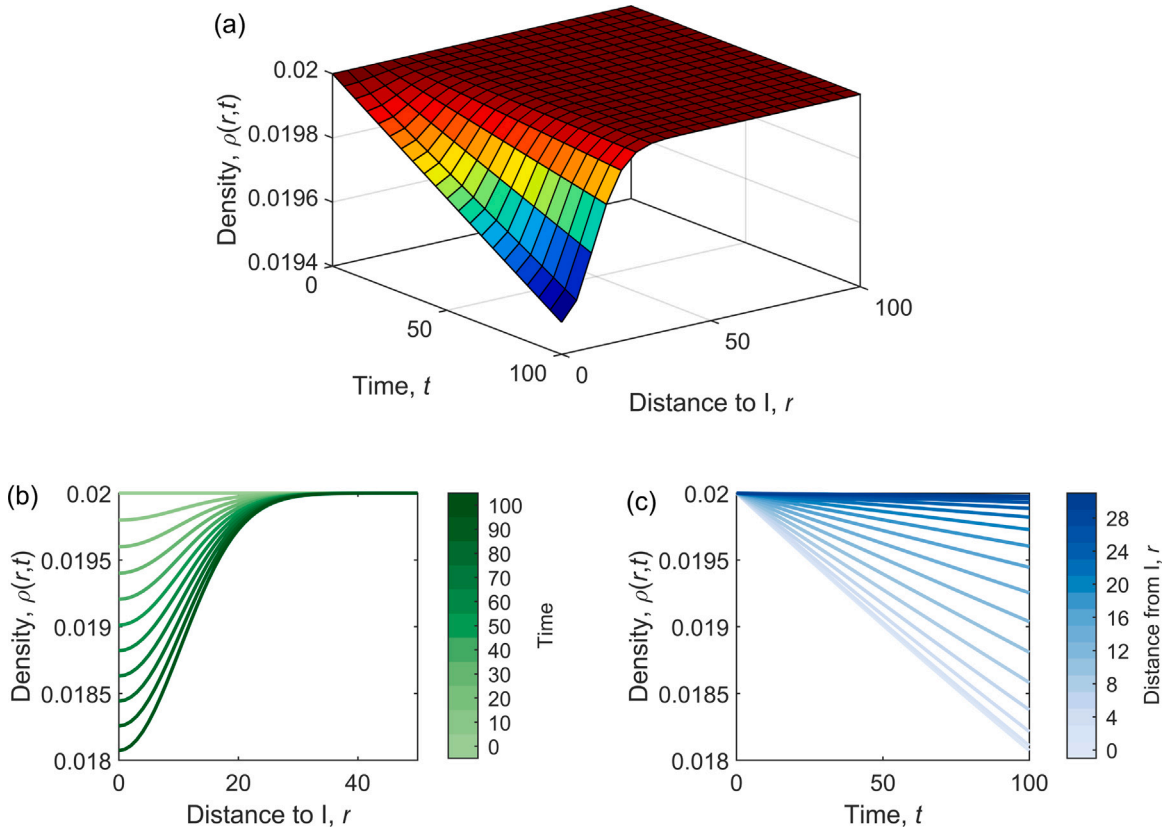


Fig. 3. The analytic density function $\rho(r,t)$ given by (7) for the idealised scenario of one infectious individual surrounded by susceptibles at an initial density of $\rho_0 = 0.02$ and infection parameters $b = 10$ and $l = 10$ showing (a) the 3D surface $\rho(r,t)$, (b) the changing density with distance from the initial infected and (c) the changing density with time.

model (Bjørnstad et al., 2020) which includes additional exposed category, E, and feeds removed individuals back into the susceptible (S) category after a period of immunity. Similarly, it would be straightforward to adapt the corresponding categories in the IBM.

2.4. Bayesian inference

To perform Bayesian inference using the stochastic SIR model, (2) and (3), we apply the linear noise approximation (LNA) to obtain a tractable likelihood function, assume a Normal prior, and use a Markov chain Monte Carlo algorithm (see e.g., Gamerman and Lopes (2006)) to sample values from the posterior distribution of the parameter β. We fix γ based on the set removal rate in the IBM, i.e., $\gamma = 1/T_I$. Further details of this process are given in Appendix A.

2.5. An analytic approximation of R_0

We consider an idealised, spatially explicit analytic expression of R_0 for a single infectious individual. A full derivation is given in Appendix B. Firstly, consider a single infectious individual with infectious period T_I, surrounded by susceptibles at varying distance r at a density ρ. We can estimate the number of new infections caused by the infected individual to be equal to the number of susceptibles at distance r, $S(t) = 2\pi r\rho$, multiplied by the probability of infection at r, in this case described by (1). The expected number of new secondary infections from the initial infected individual at time step t (for $1 < t < T_I$) for all possible values of r is therefore

$$n_I(t) \equiv I(t) - I(t-1) = \int_0^\infty 2\pi r\rho \frac{b}{2\pi l^2} \exp\left(\frac{-r^2}{2l^2}\right) dr, \\ = b\rho.$$

The cumulative sum of new secondary infections from the single infectious individual during a set time period $0 \le t \le T$, denoted $N_I(T)$, could then be approximated as

$$N_I(T) = \sum_{t=1}^{t=T} n_I(t) = b\rho T,$$

leading to a first approximation of R_0 as the sum of secondary infections at the end of the infectious period T_I:

$$R_0 \equiv N_I(T_I) = b\rho T_I. \tag{4}$$

We refer to (4) as R_0 approximation 1. However, this assumes that the density of the susceptibles remains fixed at each time step as the infection process progresses.

If we consider a large but finite domain of size L, we can introduce a time variant density taking into account the decreasing number of susceptibles due to the infection process, $\rho(t)$, leading to $n_I(t) = b\rho(t)$ with

$$\rho(t) = \rho_0 \exp\left(-\frac{bt}{L^2}\right), \tag{5}$$

where ρ_0 is the density at time $t = 0$. The cumulative sum of new secondary infections, during a set time period $0 \le t \le T$, is then given by

$$N_I(T) = \int_0^T b\rho_0 \exp\left(-\frac{bt}{L^2}\right) dt, \\ = \rho_0 L^2 \left[1 - \exp\left(-\frac{bT}{L^2}\right)\right].$$

This is equivalent to noting that $N_I(T)$ is the overall change in the density of susceptibles, multiplied by the area, i.e., $N_I(T) = L^2[\rho(t = 0) - \rho(t = T)]$. The value of R_0 is the sum of the new secondary infections across the whole infectious period,

$$R_0 \equiv N_I(T_I) = L^2\rho_0 \left[1 - \exp\left(-\frac{bT_I}{L^2}\right)\right]. \tag{6}$$

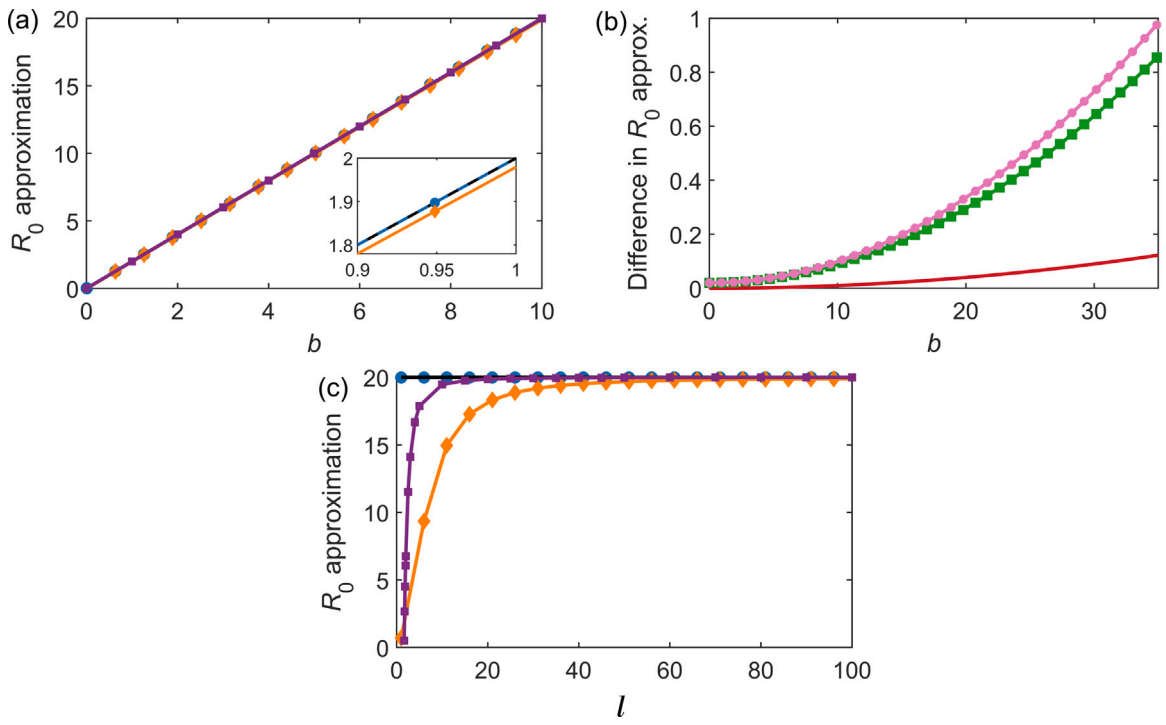


Fig. 4. A comparison of the three analytic approximations to R_0 presented in Section 2.5, (with $\rho = 0.02$, $L = 1000$ and $T_I = 100$), plus the analytic approximation presented in Suprunenko et al. (2021). (a) Approximation 1, (4), in black, approximation 2, (6), in blue dashed with circle markers and approximation 3, (8), in orange with diamond markers, for increasing values of infectious parameter b, and fixed length scale $l = 100$, with the approximation from Suprunenko et al. (2021) in purple with square markers. The inset shows a smaller range of b to highlight the difference between the overlapping lines for approximations 1–3. (b) The differences in the R_0 estimates between the three approximations, with the red solid line showing the difference between approximation 1 and 2, (4)–(6), the pink line with circle markers showing the difference between approximation 1 and approximation 3, (4)–(8), and the green line with square markers showing the difference between approximation 2 and approximation 3, (6)–(8). (c) Approximation 1, (4), in black, approximation 2, (6), in blue dashed with circle markers, approximation 3, (8), in orange with diamond markers, and the approximation from Suprunenko et al. (2021) in purple with square markers for increasing values of infectious length-scale parameter l.

This can be considered as a second approximation to R_0 where some reduction in the number of susceptibles has been taken into account. We refer to (6) as approximation 2. Note that this is an idealised scenario in which density reductions due to secondary and tertiary infections are neglected.

We can go a step further by considering the spatial influence of decreasing susceptibles. Due to the spatial infection kernel, we expect the density to vary both spatially (corresponding to the choice of infection kernel governing the spatial properties of the infection process) and temporally. We therefore consider a variant density of the form

$$\rho(r,t) = \rho_0 \exp\left(-\frac{b}{2\pi l^2} t g(r;l)\right), \tag{7}$$

where ρ_0 is the density at $t = 0$ and $g(r;l)$ is a kernel equivalent to the infection kernel in (1). This density function, (7), is shown in Fig. 3 for infection parameters $b = 10$ and $l = 10$ to illustrate its impact at short infection length scales. As with density (5) above, this function only considers the density reduction due to the secondary infections caused by the initial infected individual. The same approach could be taken for a different choice of infection kernel.

As above, the total number of new secondary infections from the primary infection during a set time period, $0 \leq t \leq T$, at a certain distance r, $N_I(r,T)$, can be calculated from the overall change in density

$$N_I(r,T) = L^2[\rho(r,t=0) - \rho(r,t=T)].$$

For all possible values of r this becomes

$$N_I(T) = \int_0^\infty 2\pi r(\rho_0 - \rho(r,T))dr = \int_0^\infty 2\pi r \rho_0 \times \left[1 - \exp\left(-\frac{b}{2\pi l^2} T g(r;l)\right)\right] dr.$$

Although the above is difficult to solve directly, it can be integrated by performing a series expansion on the exponential term and integrating on a term-by-term basis, leading to

$$N_I(T) = 2\pi\rho_0 l^2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}(\frac{b}{2\pi l^2}T)^n}{(n)(n!)}.$$

If $b/2\pi l^2$ and T are small, the first order term is sufficient to approximate R_0 as a linear function of T. This is confirmed by the numerical simulations in Section 3.3. Evaluating the summation gives

$$N_I(T) = 2\pi\rho_0 l^2 \left[E_1\left(\frac{b}{2\pi l^2}T\right) + \ln\left(\frac{b}{2\pi l^2}T\right) + \Gamma\right],$$

where the function $E_1(x)$ is the mathematically well studied exponential function $E_1(x) = \int_x^\infty t^{-1}\exp(-t)dt$ and Γ is the Euler–Mascheroni constant ≈ 0.57721. The estimated value of the basic reproduction number is therefore

$$R_0 \equiv N_I(T_I) = 2\pi\rho_0 l^2 \left[E_1\left(\frac{b}{2\pi l^2}T_I\right) + \ln\left(\frac{b}{2\pi l^2}T_I\right) + \Gamma\right], \tag{8}$$

referred to as approximation 3. A comparison of the three approximations presented here, (4), (6), and (8), plus the approximation from Suprunenko et al. (2021), is shown in Fig. 4. Note that for all the approximations derived here, reductions in susceptibles were only taken into account due to secondary infections caused from the singular primary infection, and not due to any further infections through the epidemic process, thus making all the approximations an over-estimate of R_0; we discuss this further in Section 3.3. In the parameter ranges considered, there is little difference between approximation 1 and approximation 2. Both overestimate the value of R_0 in comparison to approximation 3, (8), particularly at a short-range infectious length scale, i.e., for $l < 50$ for $b = 10$. For this reason, for the remainder of this manuscript we consider (8) as an analytic approximation to R_0

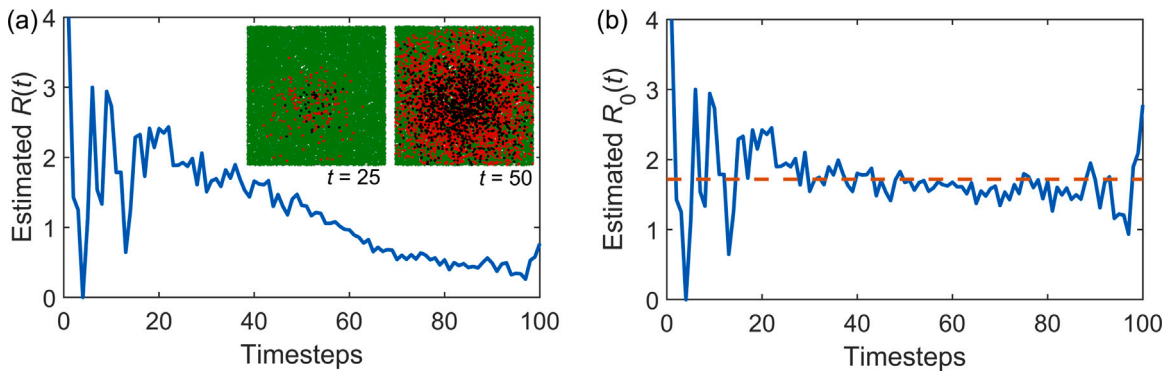


Fig. 5. Proxy contact tracing through the IBM (with parameters $b = 10$, $l = 100$, $I_0 = 5$, $L = 1000$, $\rho = 0.02$ and $T_I = 10$) to estimate (a) the instantaneous effective reproduction number, $R(t)$, with snapshots of the disease spread with susceptible (green), infected (red) and removed (black) trees at $t = 25$ and $t = 50$ in the inset, in a bounded box of $L = 1000$ and (b) the instantaneous basic reproduction number $R_0(t)$ (blue solid line) and basic reproduction number R_0 (orange dashed line, calculated as the mean of $R_0(t)$).

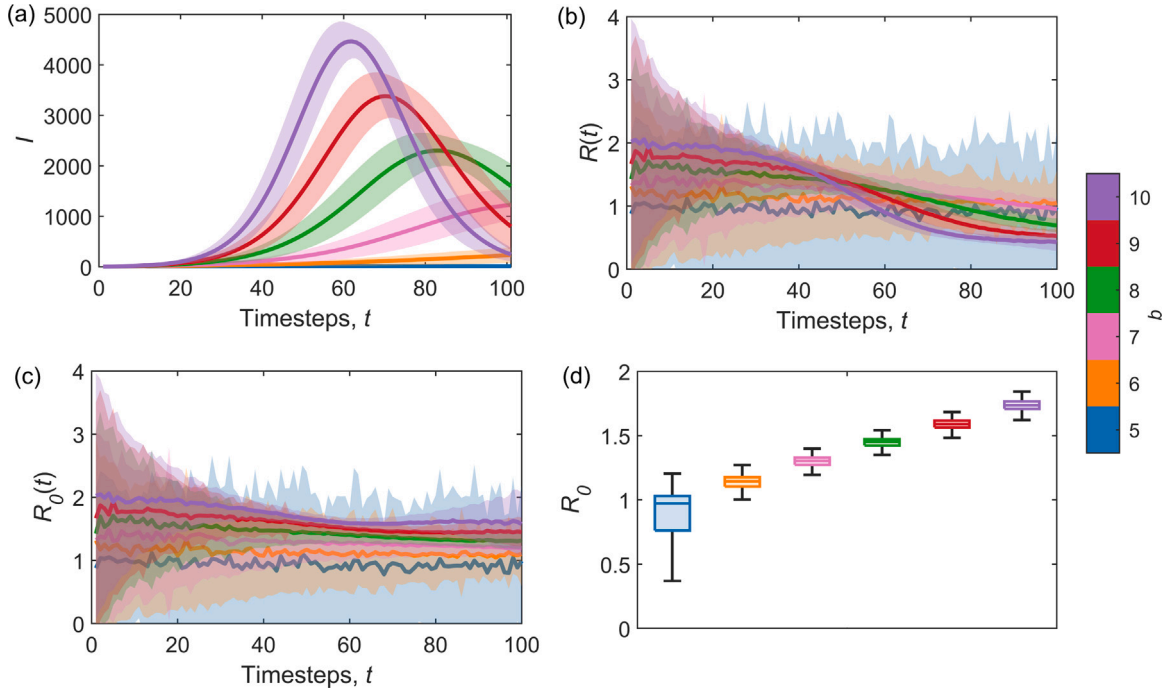


Fig. 6. Proxy contact tracing through the IBM (with parameters: $l = 100$, $I_0 = 5$, $L = 1000$, $\rho = 0.02$ and $T_I = 10$ at increasing values of infectivity b) to estimate R_0 for varying b. The mean (from 500 simulations) (a) infected population with time, along with (b) the instantaneous effective reproduction number $R(t)$ and (c) the instantaneous basic reproduction number $R_0(t)$. Error bars show the standard deviation. (d) Box plots of the estimated R_0 values from each of the 500 runs (calculated as the mean of each of the 500 corresponding $R_0(t)$ time series).

and further investigate its suitability to describe the IBM in Section 3.3. We also consider the analytic approximation from Suprunenko et al. (2021), calculating estimates of R_0 for our model parameters using the open source code provided.

3. Results

In this section we investigate the three methods of estimating the basic reproduction number from a typical individual-based model (IBM), described in Section 2. These methods include contact tracing (Section 3.1 using the methods described in Section 2.2.1), fitting to the standard SIR equations (Section 3.2 using the methods described in Section 2.4), and an analytical approximation (Section 3.3, using the methods described in Section 2.5). We then compare these methods by applying them to a simulation of the ash dieback pathogen in the UK (Section 3.4).

3.1. Contact tracing through the IBM

In the IBM described in Section 2.2, we assume infections can occur due to pressure from multiple sources through the infectious kernels surrounding each infected tree, as shown in Fig. 2. This results in infection dynamics that are not one-to-one. We thus use the method for proxy-contact tracing described in Section 2.2.1, resulting in a time-step estimate for the (instantaneous) effective reproduction number, $R(t)$. An example estimate of $R(t)$, the corresponding instantaneous basic reproduction number calculated as $R_0(t) = R(t)S(0)/S(t)$, and a point estimate of R_0 (calculated as the mean of $R_0(t)$) for a single illustrative simulation are shown in Fig. 5.

We can consider how varying the two IBM infection parameters, the infectivity b and the length scale l, will impact the reproduction number, summarised in Figs. 6 and 7. The mean infected population time series (averaged over 500 IBM simulations for each parameter

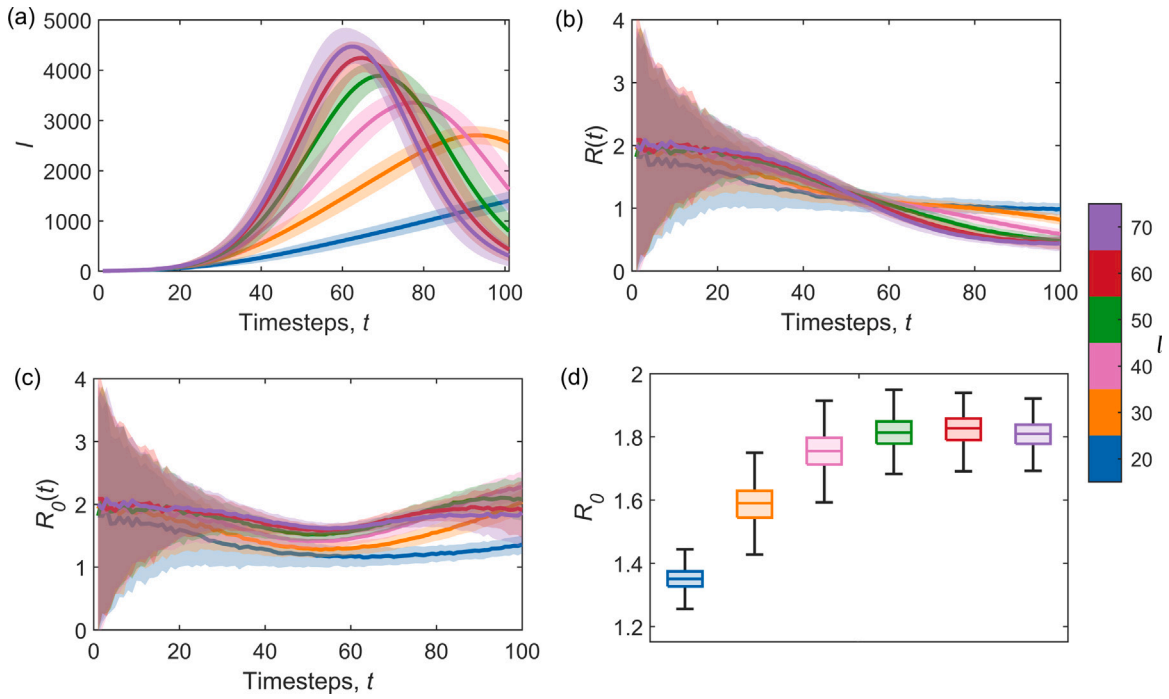


Fig. 7. Proxy contact tracing through the IBM (with parameters: $b = 10$, $I_0 = 5$, $L = 1000$, $\rho = 0.02$ and $T_I = 10$ at increasing values of infectious length scale l) to estimate R_0 for varying l. The mean (from 500 simulations) (a) infected population with time, along with (b) the instantaneous effective reproduction number $R(t)$ and (c) the instantaneous basic reproduction number $R_0(t)$. Error bars show the standard deviation. (d) Box plots of the estimated R_0 values from each of the 500 runs (calculated as the mean of each of the 500 corresponding $R_0(t)$ time series).

set) show the epidemic dynamics for increasing values of b and l, shown in Figs. 6(a) and 7(a), respectively. The mean instantaneous effective reproduction number $R(t)$ for increasing b and l are shown in Figs. 6(b) and 7(b), respectively. As the 'strength' of the epidemic increases (through an increasing b, or to a lesser extent, an increasing l) the early values of $R(t)$ increase. This results in a faster decreasing pool of susceptibles, and thus a more significant decrease in $R(t)$ with time. The instantaneous basic reproduction number, $R_0(t)$, takes into account this decrease in susceptibles, and is shown for increasing b and l in Figs. 6(c) and 7(c), respectively, showing the expected increase in $R_0(t)$ with both increasing b and l. A single value of the basic reproduction number, R_0, can be calculated as the mean of the instantaneous basic reproduction number for each simulation, shown in the box plots in Figs. 6(d) and 7(c) for increasing b and l, respectively. For increasing infectivity b, R_0 increases linearly. For increasing length-scale l, R_0 increases but saturates due to the presence of the bounding box (here at $L = 1000$), artificially decreasing the number of infections.

3.2. Fitting the stochastic SIR equations

It may be necessary to avoid the computational expense of tracking infections through the IBM. In this case, we can compare the simulations to the SIR model (Kermack and McKendrick, 1927) described with introduced stochasticity by (2) and (3). In this formulation, R_0 is defined as $\beta N/\gamma$. Here, we employ the inference scheme detailed in Section 2.4 and Appendix A to estimate the SIR model parameter β. Since γ is known from the IBM input parameters ($\gamma = 1/T_I$), we can then calculate R_0. Example fittings of the stochastic SIR model for three characteristic spread dynamics with increasing dispersal length scale l are shown in Fig. 8. At shorter length scales, we can see the slower increase in the infectious population, due to the spatial structure and a loss of the homogeneous mixing assumption when the interactions are short-range, particularly visible at the start of the epidemic due to a restricted number of susceptibles falling within the infection kernel. An advantage of this technique is the ability to obtain a posterior distribution of plausible R_0 values (through the posterior distribution for β and our estimate of $\gamma = 1/T_I$ from the IBM), shown in Fig. 8 for each of the example infection spread dynamics.

3.3. An analytical approximation

An idealised, spatially explicit analytical approximation for R_0 is derived in Section 2.5. In this section, we compare the analytic predictions of R_0 resulting from (8) to the numerical results from proxy contact tracing within the IBM (as described in Section 2.2.1 and Section 3.1).

The cumulative number of expected secondary infections due to a primary infection over an infectious period of $T_I = 100$, $N_I(t)$, is shown in Fig. 9(a) for a fixed dispersal parameter $l = 100$ and increasing b. The linear relation between time and $N_I(t)$ was predicted by (8) in Section 2.5. For lower values of b, the analytic approximation captures the estimation from the proxy contact tracing. For $b = 8$, the analytic estimate from (8) overestimates $N_I(t)$ from around $t = 50$, as it only takes into account a reduction in susceptibles from infections caused by the primary infectious individual, and not other infections involved in the epidemic process. The contact-traced estimate shows the actual saturation of $N_I(t)$ that occurs as the number of available susceptibles limits the infection spread. Similarly, Fig. 9(b) shows $N_I(t)$ for fixed $b = 5$ and increasing dispersal parameter l. In this case, the overestimation from the analytic $N_I(t)$ occurs at shorter length scales (e.g., for $l = 10$ in this parameter regime), where the pool of susceptibles has been reduced in a small area around the primary infection, limiting the epidemic process, as shown by contact tracing. A larger dispersal kernel encompasses a larger local neighbourhood of infectivity around the primary susceptible, resulting in less sensitivity to reductions in susceptibles through subsequent infections. The approximation of the basic reproduction number R_0 is given by the cumulative number of secondary infections resulting from the primary infection over its infectious period, i.e., $R_0 = N_I(T_I)$. In general, (8) overestimates R_0 due to the neglect of any subsequent infections caused by the secondary infections. This is exacerbated when the infectivity is high, or the length scale of dispersal is low, as seen in the plateau of infections in Fig. 9(a) and (b). Thus, (8) describes constant transition

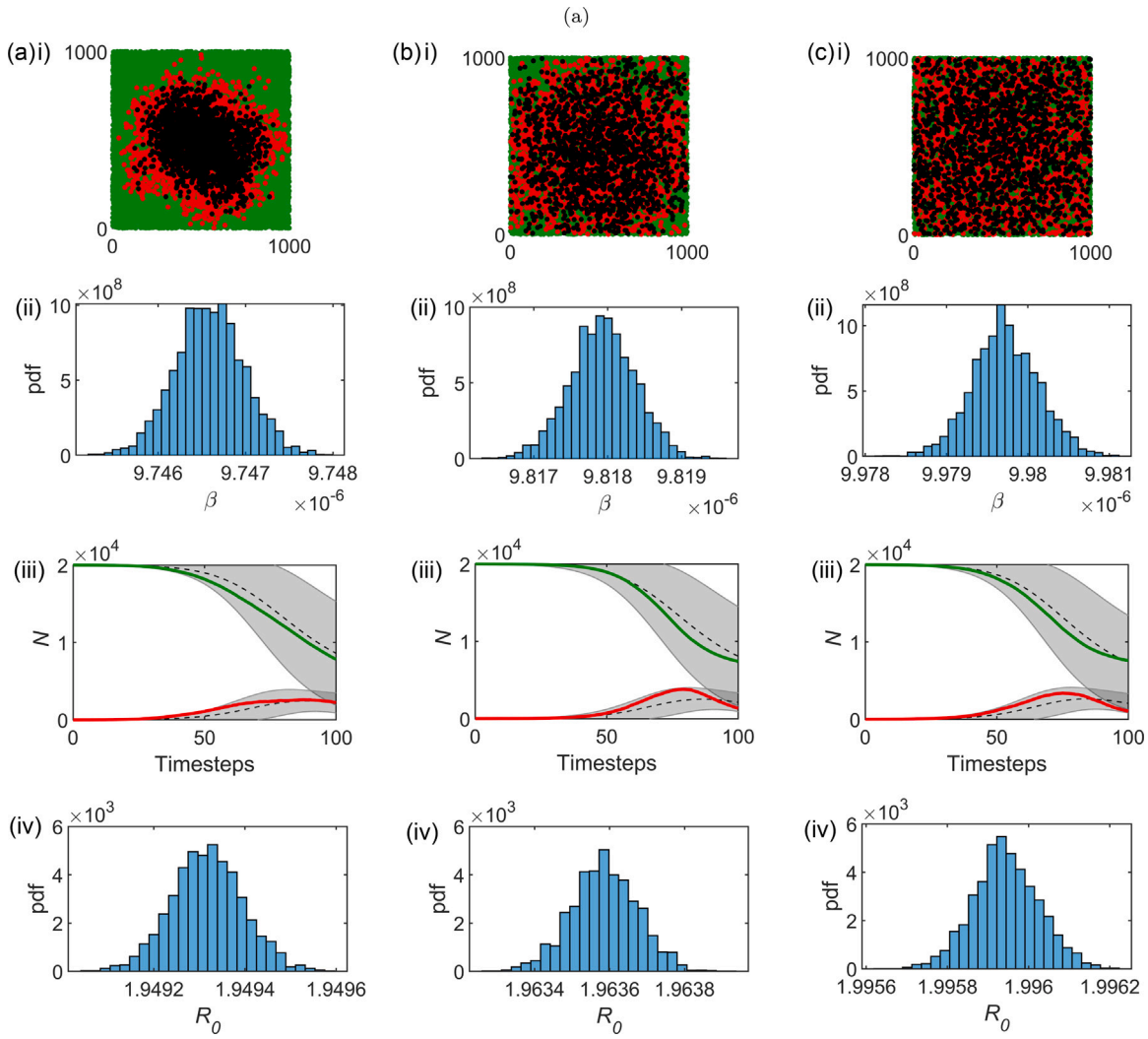


Fig. 8. Example epidemic regimes (with $\rho = 0.02$, $T_I = 10$ and $L = 1000$) taking place over similar time scales with an increasing radius of infection l: (a) $b = 10$ and $l = 40$, (b) $b = 10$ and $l = 100$, and (c) $b = 15$ and $l = 400$, showing (i) a snapshot at 20% mortality with susceptible (green), infectious (red) and removed (black) trees, (ii) posterior distributions of the inferred SIR parameter β, (iii) the number of infectious (I, red) and susceptible (S, green) trees from the IBM, with the corresponding stochastic SIR fitting using the median inferred β, and $\gamma = 1/T_I = 0.1$, shown as the mean of 500 runs in black dashed with standard deviation error bars in grey and (iv) posterior distribution of estimated $R_0(t)$, using all posterior estimates of β, and $\gamma = 1/T_I = 0.1$.

rates accurately but deviates from model simulations when subsequent infections cause a significant reduction in local susceptible density. The values of R_0 estimated through both the analytic and numeric contact tracing approximations for an arbitrary fixed infectivity of $b = 5$ and an increasing dispersal parameter l are shown in Fig. 9(c). The analytic approximation captures the general trend of increasing R_0, followed by saturation due to a limited susceptible population in the infectious area, as in Fig. 7(d). The analytic approximation previously defined in Suprunenko et al. (2021) (also shown in Fig. 9(c)) shows a similar trend and is in agreement with the predictions from (8) at greater length scales.

A useful property of the basic reproduction number is its quantification of the transmission threshold, defined as $R_0 = 1$, predicting the separation of states between confinement and epidemic. In Fig. 9(d), the threshold predicted by (8) is shown with a two-dimensional phase plot of all estimated R_0 over tree density and infection parameter b, allowing a categorisation of disease confinement or epidemic from the model parameters.

We can also assess how the total tree mortality relates to the threshold $R_0 > 1$ predicted by (8). The total proportion of host trees in the Removed ($R^\dagger$) compartment (removal prevalence) for increasing predicted analytical R_0 from (8) is shown for an illustrative epidemic regime ($L = 500$, $\rho = 0.01$, $T_I = 100$, $r = 100$, $0 < b < 4$) in Fig. 9(e). When b and ρ result in an analytical prediction of R_0 less than one, correspondingly, tree mortality is low. When the infectivity is increased such that analytical $R_0 > 1$, tree mortality rises considerably. The approximation described by (8) therefore demonstrates the threshold-like behaviour defined by $R_0 = 1$. It is worth noting that despite being above the threshold, the numerical simulations from the IBM can still fail to produce an epidemic, due to the influence of early stochastic forces increasing the probability of epidemic extinction (Heffernan et al., 2005; Tildesley and Keeling, 2009), a disadvantage of the general concept of R_0 (Li et al., 2011). The threshold-like behaviour witnessed in Fig. 9(e) demonstrates that (8) provides a simple predictive framework for the IBM considered here.

3.4. Application to ash dieback

In this section we employ the idealised IBM, as described in Section 2.2, to describe the spread of ash dieback (*Hymenoscyphus fraxineus*), a highly destructive fungal tree disease, and use each of the above methods to estimate the basic reproduction number R_0.

The IBM is initialised with a density of $\rho = 0.0444$ trees/m^2, given the estimate of 444 ash trees per hectare (10,000 m^2) within small

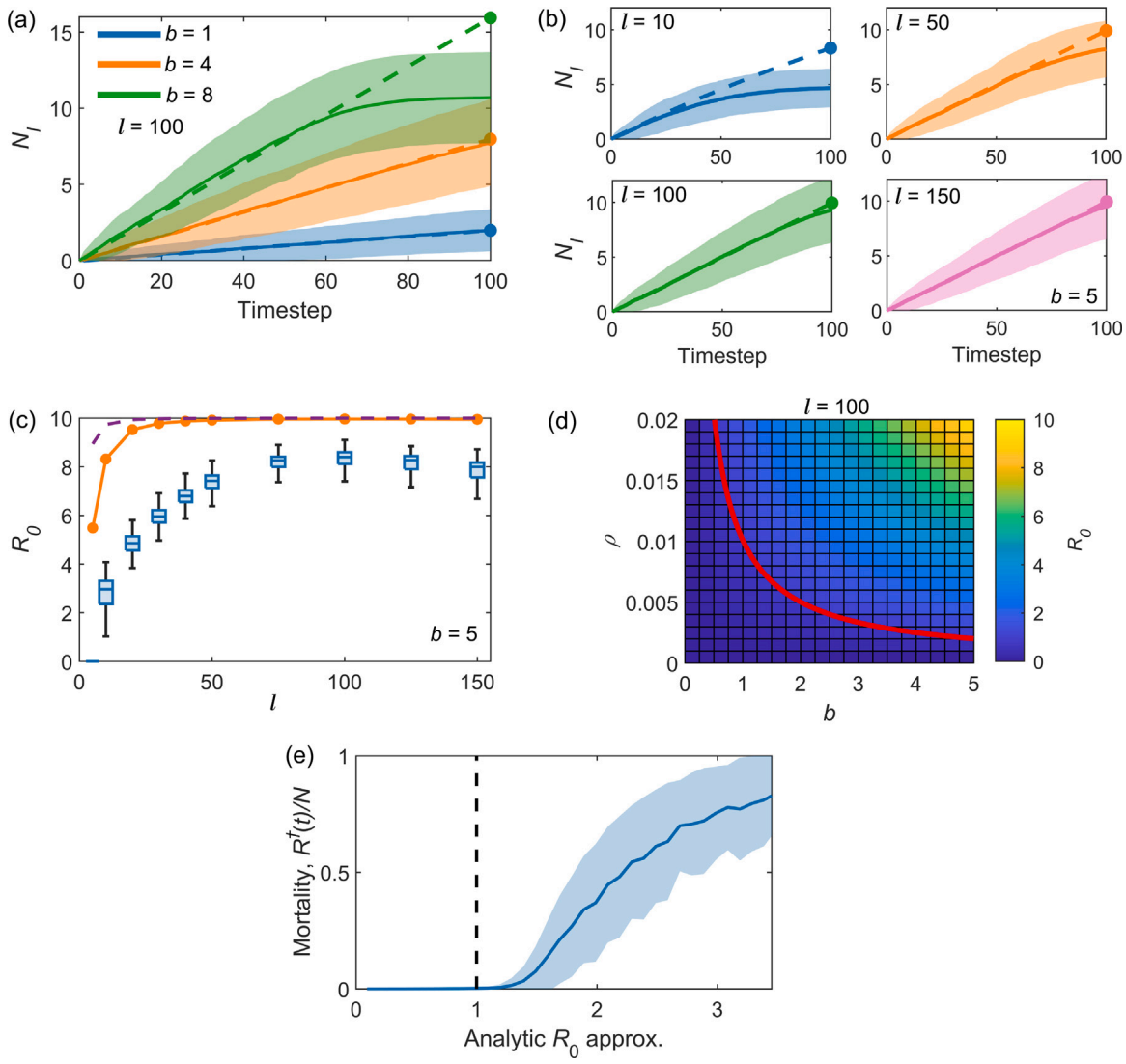


Fig. 9. Comparison between the analytical expression for $N_I(t)$ (the cumulative number of secondary infections resulting from the primary infection), and hence $R_0 = N_I(T_I)$ from (8), and the value of R_0 estimated through proxy contact tracing through the IBM (as in Section 3.1). In all cases $L = 1000$, $\rho = 0.02$, $I_0 = 1$ and $T_I = 100$. (a) N_I for increasing infectivity b (with fixed infectious length scale $l = 100$). The analytic estimate of $N_I(t)$ (from (8)) is shown as a dashed line, with a circle at the end of the infectious period indicating R_0. The solid lines show the median contact traced estimates of R_0 (over 500 runs of the IBM) with standard deviation error bars. (b) N_I for increasing infectious length scale l (with fixed infectivity $b = 5$). The analytic estimate of N_I is shown as a dashed line, with a circle at the end of the infectious period indicating R_0. The solid lines show the median contact traced estimates of R_0 (over 500 runs of the IBM) with standard deviation error bars. (c) The analytic estimates of R_0 from (8) for increasing length scale l (and fixed infectivity $b = 5$) is shown as the solid orange line with circle markers. The analytic approximation from Suprunenko et al. (2021) is shown as the dashed purple line. Box plots show the corresponding estimates of the contact traced R_0 values from 500 runs of the IBM. (d) The analytic R_0 phase plane predicted by (8) for increasing density ρ and infectivity b. The threshold, given by $R_0 = 1$, is plotted in red and illustrates the separation between confinement and epidemic. (e) The relationship between the total tree mortality (removal prevalence) over 500 runs of the IBM and the R_0 value predicted by (8), demonstrating the threshold-like behaviour at $R_0 = 1$.

woodlands in the UK (The Tree Council, 2014). We set the dispersal parameter l in (1) to be 138 m, based upon estimates of the local dispersal kernels for the ash dieback pathogen from spore-trapping data (Grosdidier et al., 2018). The pathogen can remain active for around 4 years after infection (Wylder et al., 2018) and so the infectious period is chosen to be $T_I = 4$ years. This leaves the unconstrained infectivity parameter, b, and the time-step over which new infections are assessed, dt. Here we arbitrarily take $dt = 1$ week, and consider a range of b (with units $m^2 week^{-1}$) to focus on regimes resulting in $0 < R_0 < 7$.

Assuming there will be some variability in the estimate of ρ in different areas, we show the phase plane of the analytic estimation of R_0 for a range of densities and infectivity parameters, along with the threshold value of $R_0 = 1$, in Fig. 10(a). This illustrates the values of ρ and b that would result in the epidemic regime. The estimates of R_0 for 500 simulations with fixed $\rho = 0.0444$ trees/m^2 and $l = 138$ m through each of the methods described above, plus the analytic approximation from Suprunenko et al. (2021), are shown in Fig. 10(b). The overestimate of R_0 through both analytical expressions is clear, as discussed in Section 3.3, with similar estimates of R_0 resulting through both contact tracing and inference of the SIR parameters. Given an estimate of the infectious parameter b (discussed further in Section 4), this would allow the prediction of R_0 through several comparative methods without requiring physical contact tracing of the infection.

4. Discussion

Stochastic IBMs are a popular choice for describing the spread of tree diseases and pests through woodland areas, due to their flexibility and ability to describe large-scale collective behaviours from the programming of individual behaviours. Here we use a typical IBM for the spread of an arbitrary disease or pest through a synthetic forest, and explore different methods of calculating the basic reproduction number R_0, a key parameter for disease characterisation and forecasting.

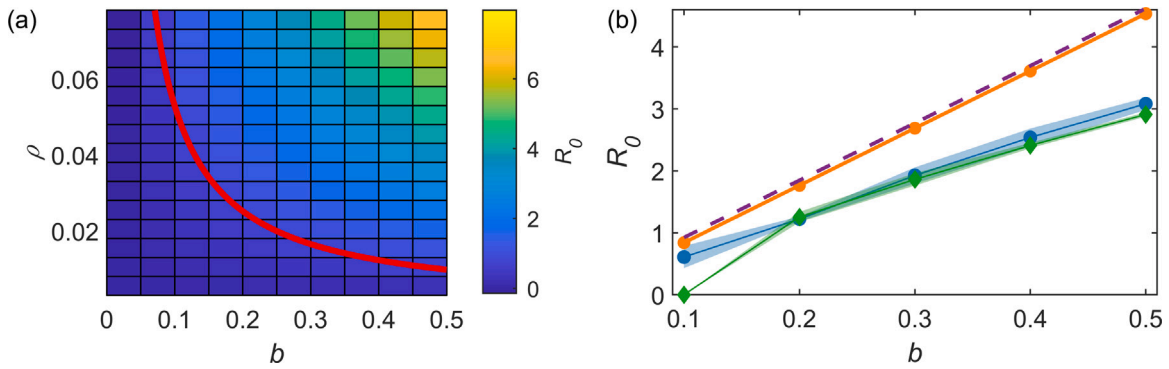


Fig. 10. (a) The analytic R_0 phase plane predicted by (8) for increasing density ρ (including the estimated tree density relevant for ash dieback, $\rho = 0.0444$) and infectivity b. The threshold, given by $R_0 = 1$, is plotted in red and illustrates the separation between confinement and epidemic. (b) The estimates of R_0 with the ash dieback parameters $l = 138$ m, $\rho = 0.0444$ trees/m^2 and $T_I = 4$ years through the analytic approximation (8) (solid line with orange circles), previous analytic approximation from Suprunenko et al. (2021) (purple dashed line), contact tracing where R_0 is the mean of $R_0(t)$ (blue circles), and inference of the SIR parameters (green diamonds). Markers indicate the mean of 500 simulations, with standard deviation error bars.

We have expanded upon similar compartmental IBMs of tree disease (Orozco-Fuentes et al., 2019), introducing a flexible spatial component to the infection spread through the inclusion of an infectious kernel, as used in Parnell et al. (2009) to describe the spread of Asiatic citrus canker disease. The infectious dispersal kernel characterises the spatial spread of the disease and thus has a large impact on the epidemic outcome (Fabre et al., 2021). In this case we use an IBM with a Gaussian probability of infection due to its simplicity and prevalence in similar models (Nathan et al., 2012; Mikaberidze et al., 2016; Fabre et al., 2021; Prussin II et al., 2015), but this is entirely flexible and the most appropriate choice will depend on the nature of the disease or pest considered. For example, for the ash dieback epidemic, both Gaussian and power law kernels have been estimated for the dispersal at different scales (Grosdidier et al., 2018). Future work should explore the impact of different kernels on the IBM behaviour and the corresponding estimations of R_0.

The choice of infectious kernel is linked to dispersal (Nathan et al., 2012; Bullock et al., 2017) describing the probability of pathogen presence over distance. Dispersal kernels can be estimated through either mechanistic models (Nathan et al., 2011; Thompson et al., 2014), or through fitting statistical functions to dispersal data (Bullock et al., 2017; Grosdidier et al., 2018), but can be challenging to ascertain. A further complication is the mapping of a dispersal kernel to the corresponding infectious kernel, i.e., how does probability of pathogen presence correspond to the probability of successful infection? In the IBM kernel this is captured in the infectivity parameter $B = b/(2\pi l^2)$, and we consider a range of arbitrarily chosen values to illustrate the methodology. For the ash dieback simulation, we take the length scale from the dispersal kernel estimated from ecological data describing the density of pathogen spores at varying distance from infected ash trees (Grosdidier et al., 2018). Since B (and thus b) is not known, we again consider a range of parameter values leading to plausible R_0 estimations.

Inference for B under the IBM may be possible by leveraging recently proposed approximate Bayesian computation techniques (ABC, see e.g. Minter and Retkute (2019)) although such approaches typically require several millions of model simulations. The practical applicability of ABC for our proposed spatio-temporal model remains the subject of ongoing work. It may also be possible to constrain B through estimations of R_0 (through empirical means or from data of infections) in previous outbreaks of the disease or pest under similar conditions. Climate change is driving the geographical expanse of many ecological epidemics and invasions, and so ecological data may be available from multiple locations.

The most straightforward way of calculating R_0 within the IBM is through the contact tracing of infections. This can be done directly, by storing the number of secondary infections resulting from a singular tree, but requires a model in which trees are considered individually in the infection time-step, thus requiring more computational expense. In the case of a vectorised IBM, where individual secondary infections are not tree specific, we can calculate an estimation of $R(t)$ through the number of average new infections, and sharing 'the blame' for these between the currently infected population. Although noisy, particularly at the beginning of an epidemic regime (if the length scale of infectivity is low and initial infected trees are clustered, resulting in a limited number of susceptible trees in the local area) and end of epidemic simulation (as the number of susceptible trees available approaches zero), the median of the $R_0(t)$ series provides an estimation consistent with the other methods considered here. True contact-tracing methods, where the source of individual infections is tracked, have the advantage of providing a direct measurement of R_0, even with increasing model complexity, but are more computationally expensive. Vectorised implementations of infection spread will require approximations to tracking the average number of infections caused by individuals and thus result in more noisy estimations.

The temporal output from the IBM can be compared to the classic SIR equations, allowing a straight-forward estimation of R_0 through the model parameters β and γ. The estimation of the model parameters through the inferential scheme results in a distribution of plausible R_0 values, which has the advantage of capturing the parameter uncertainty and is useful for forecasting best and worst-case scenarios. However, this method relies on the assumption that the standard SIR equations will effectively capture the time-series resulting from the IBM. Unless the length scale of dispersal is very large, we will not fulfil the homogeneous mixing assumption of the SIR model. Despite this, we still find the parameter estimates to be descriptive enough to provide a measure of R_0 consistent with other methods. Future work could explore the relationship between the dispersal kernel parameters and the SIR model parameters.

We also derive an analytical expression for approximating R_0 which predicts the epidemic threshold and the contact-traced reproduction number computed through the IBM simulations, with the caveat that it overestimates R_0, particularly when the epidemic severity is high. Our analytic approximation shows good consistency with the previous expression developed in Suprunenko et al. (2021) with a reduced over-estimate at shorter length scales, and is perhaps more simple to implement computationally. Future work could consider a thorough quantitative comparison of existing analytic R_0 approximations under varying infectious dynamics and models. The overestimation of R_0 can be compared to well-known results (Tildesley and Keeling, 2009; Keeling and Eames, 2005) showing that the first-generation basic reproduction number for farms infected with foot-and-mouth overestimates the growth rate of infection. The analytic estimate has the advantage of

Table 1
Advantages and disadvantages of the methods presented to estimate R_0 from an individual-based model.

Method	Advantages	Disadvantages
Contact tracing	Intuitive, accurate representation of model dynamics, straight-forward to implement, robust to other individual based model implementations	Computationally expensive, requires an approximation using the average number of infections if individual infection causes are not tracked
Analytic approx.	Requires IBM parameters only, provides a quick upper bound (worst-case) scenario estimate	Over-estimation at short length-scales, non-trivial to adapt to other IBM implementations, kernels and epidemic regimes, assumption of spatial homogeneity of individuals
Inference of SIR parameters	Require model output only (easily transferable to varying model implementations and to real-world applications), quantifies uncertainty	Computationally expensive, prior knowledge of parameters required, relies on the assumption that the standard SIR equations capture the model dynamics (homogeneous mixing)

calculation through the IBM model parameters only, however comparable analytical solutions are challenging to determine for more elaborate life-cycles, dynamics and aggregated host distributions.

Considering an illustrative simulation of ash dieback in the UK using the IBM allows a comparison of the different estimation methods of R_0 in context. If the infectivity parameter b were to be estimated (through estimation of B and l), as discussed above, the IBM model parameters could be used to efficiently calculate an upper-bound for a plausible estimation of R_0. Numerical methods such as contact tracing and inference of the SIR model parameters provide a more accurate estimation of R_0, albeit at more computational expense. Future work could expand the IBM to larger areas, considering varying densities of woodland across different geographical areas to generate landscape level predictive R_0 maps.

Exploring different underlying tree distributions to capture more realistic landscapes would also be an interesting avenue for future work. This would be straightforward to implement in the IBM and the approximations of R_0 through both proxy-contact tracing and inference of the SIR parameters would be estimated in the same way, however, an adjustment in the analytic R_0 estimation would be required as this derivation depends on an assumption of an homogeneous underlying host tree distribution. Similarly, the work here could be extended to consider other formulations of compartmental models, e.g., SIRS where removed individuals can transition back to the susceptible class (Golightly et al., 2023), or SEIR as considered in Suprunenko et al. (2021).

This work provides an additional framework to existing methods (Suprunenko et al., 2021) to estimate epidemic severity through the parameter R_0, using a flexible IBM for tree disease spread. Given knowledge of the dispersal kernel for a particular pathogen, the IBM can be used to estimate R_0 through several methods, the advantages and disadvantages of which are summarised in Table 1, and thus provides an extensible tool which can be further developed for ecological epidemic forecasting.

CRediT authorship contribution statement

Laura E. Wadkin: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **John Holden:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – review & editing. **Rammile Ettelaie:** Conceptualization, Supervision, Writing – review & editing, Methodology. **Melvin J. Holmes:** Conceptualization, Supervision, Writing – review & editing, Methodology. **James Smith:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing, Methodology. **Andrew Golightly:** Conceptualization, Methodology, Software, Supervision, Writing – review & editing, Funding acquisition. **Nick G. Parker:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing, Methodology. **Andrew W. Baggaley:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing, Methodology.

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

All codes relating to this manuscript are freely available at: https://doi.org/10.25405/data.ncl.24787752.

Acknowledgements

This research was supported by: EPSRC New Horizons, UK Grant EP/V048511/1 (AB, AG, NGP, and LW), NERC Knowledge Exchange Fellows, UK Grant NE/X000478/1 (LW) and DEFRA UK (JH, JS, and NGP). We also thank Dr Srivatsa Badariprasad for helpful discussions.

Appendix A. Bayesian inference details

For inference on the stochastic SIR model, (2) and (3), we apply the linear noise approximation (LNA) to obtain a tractable approximation to the stochastic differential equation. Formal details of the LNA can be found in Kurtz (1972), van Kampen (2001) and Komorowski et al. (2009), with an outline derivation below.

Consider a partition of the system X_t as

$$X_t = \eta_t + Z_t, \tag{A.1}$$

where $\{\eta_t, t \geq 0\}$ is a deterministic process satisfying the ordinary differential equation (ODE)

$$\frac{d\eta_t}{dt} = a(\eta_t, \theta), \qquad \eta_0 = x_0, \tag{A.2}$$

and $\{Z_t, t \geq 0\}$ is a residual stochastic process. The residual process Z_t satisfies

$$dZ_t = \{a(X_t, \theta) - a(\eta_t, \theta)\}\, dt + \sqrt{d(X_t, \theta)}\, dW_t,$$

which will typically be intractable. The assumption that $\|X_t - \eta_t\|$ is "small" motivates a Taylor series expansion of $a(x_t, \theta)$ and $d(x_t, \theta)$ about η_t, with retention of the first two terms in the expansion of a and the first term in the expansion of b. This gives an approximate residual process $\{\hat{Z}_t, t \geq 0\}$ satisfying

$$d\hat{Z}_t = H_t \hat{z}_t\, dt + \sqrt{d(\eta_t, \theta)}\, dW_t,$$

where H_t is the Jacobian matrix with (i, j)th element

$$(H_t)_{i,j} = \frac{\partial a_i(\eta_t, \theta)}{\partial \eta_{j,t}}.$$

For the SIR model in (2) and (3) we therefore have

$$H_t = \begin{pmatrix} -\beta I_t & -\beta S_t \\ \beta I_t & \beta S_t - \gamma \end{pmatrix}.$$

Given an initial condition $\hat{Z}_0 \sim \mathrm{N}(\hat{z}_0, \hat{V}_0)$, it can be shown that $\hat{Z}_t$ is a Gaussian random variable (Fearnhead et al., 2014). Consequently, the partition in (A.1) with Z_t replaced by $\hat{Z}_t$, and the initial conditions $\eta_0 = x_0$ and $\hat{Z}_0 = 0$ give

$$X_t \sim \mathrm{N}\left(\eta_t, V_t\right), \tag{A.3}$$

where η_t satisfies (A.2) and V_t satisfies

$$\frac{dV_t}{dt} = V_t H_t' + d(\eta_t, \theta) + H_t V_t, \qquad V_0 = 0. \tag{A.4}$$

Further details on the derivation of (A.4) are given in Wadkin et al. (2022). Hence, the linear noise approximation is characterised by the Gaussian distribution in (A.3), with mean and variance found by solving the ODE system given by (A.2) and (A.4), which can be solved numerically.

Given the observational process X_t, and parameter vector θ, the likelihood is then

$$L(\theta|X) = \prod_{i=0}^{n-1} \mathrm{N}_2(X_{t_{i+1}}; X_{t_i} + a(X_{t_i}, \theta)dt, V_{t_i} + (V_{t_i} H_{t_i} + d(\eta_{t_i}, \theta) + H_{t_i} V_{t_i})dt) \tag{A.5}$$

where $\mathrm{N}_2(\cdot, m, \upsilon)$ denotes the multivariate Gaussian density with mean vector m and variance matrix υ. In this case we fix γ according to the removal rate in the IBM and so $\theta = (\beta)$ only. We set a prior specification of $\beta \sim \log\mathrm{N}(0, 1)$. Since $\beta > 0$ we work on a unrestricted parameter space by letting $\lambda = \log\beta$. The posterior is given by

$$\pi(\lambda|X) \propto \pi(\lambda) L(e^{\lambda}|X), \tag{A.6}$$

where $\pi(\lambda) = \mathrm{N}(\lambda; a_\beta, c_\beta^2)$. We can then use an MCMC scheme (Algorithm 1) to generate draws of $\lambda|X$ and exponentiate to give draws of $\theta|X$. This results in a posterior distribution of plausible values of β. In cases where γ is not known, it is straightforward to expand the parameter search to target $\theta = (\beta, \gamma)'$.

Algorithm 1 Random walk Metropolis algorithm

1. Initialise at $\theta^{(0)}$ in the support of $\pi(\theta|X)$. Set the iteration counter $i = 1$.
2. Propose $\theta^* = \theta^{(i-1)} + \epsilon_i$ where $\epsilon_i \sim \mathrm{N}(0, \Omega)$
3. With probability

   $$\alpha(\theta^*|\theta^{(i-1)}) = \min\left\{1, \frac{\pi(\theta^*)\pi(X|\theta^*)}{\pi(\theta^{(i-1)})\pi(X|\theta^{(i-1)})}\right\}$$

   put $\theta^{(i)} = \theta^*$ otherwise put $\theta^{(i)} = \theta^{(i-1)}$.
4. If $i = M$ stop otherwise put $i := i + 1$ and go to step 2.

Appendix B. Analytic R_0 derivation

In this section, an idealised, spatially explicit expression of $R_0(t)$ is derived for the IBM. Firstly, consider a single infectious individual with infectious period T_I, surrounded by susceptibles at varying distance r at a density ρ. We can estimate the number of new infections caused by the infected individual to be equal to the number of susceptibles at distance r, $S(t) = 2\pi r\rho$, multiplied by the probability of infection at r, in this case described by (1). The expected number of new secondary infections from the initial infected individual at time step t (for $1 < t < T_I$) for all possible values of r is therefore

$$\begin{aligned} n_I(t) \equiv I(t) - I(t-1) &= \int_0^\infty 2\pi r\rho \frac{b}{2\pi l^2} \exp\left(\frac{-r^2}{2l^2}\right) dr, \\ &= b\rho. \end{aligned}$$

The cumulative sum of new secondary infections from the single infectious individual during a set time period $0 \le t \le T$, denoted $N_I(T)$, could then be approximated as

$$N_I(T) = \sum_{t=1}^{t=T} n_I(t) = b\rho T,$$

leading to a first approximation of R_0 as the sum of secondary infections at the end of the infectious period T_I:

$$R_0 \equiv N_I(T_I) = b\rho T_I. \tag{B.1}$$

However, this assumes that the density of the susceptibles remains fixed at each time step as the infection process progresses.

If host growth is neglected (likely appropriate for many tree based scenarios), fewer trees will be available to infect at time-step $t + 1$. The susceptible tree density can therefore be seen as a monotonically decreasing function of time, $\rho(t)$, leading to

$$n_I(t) = b\rho(t).$$

In a large but finite domain of size L, tree density approximately follows

$$\frac{d\rho}{dt} = -\frac{n_I(t)}{L^2} = -\frac{b\rho(t)}{L^2}.$$

Solving the above, with the initial condition ρ_0 at $t = 0$, leads to

$$\rho(t) = \rho_0 \exp\left(-\frac{b}{L^2} t\right).$$

The cumulative sum of new secondary infections, during a set time period $0 \le t \le T$, is then given by

$$\begin{aligned} N_I(T) = \sum_{t=1}^{t=T} n_I(t) &= \int_0^T b\rho_0 \exp\left(-\frac{bt}{L^2}\right) dt, \\ &= \rho_0 L^2 \left[1 - \exp\left(-\frac{bT}{L^2}\right)\right]. \end{aligned}$$

This is equivalent to noting that $N_I(T)$ is the overall change in the density of susceptibles, multiplied by the area, i.e., $N_I(T) = L^2[\rho(t = 0) - \rho(t = T)]$. The value of R_0 is the sum of the new secondary infections across the whole infectious period,

$$R_0 \equiv N_I(T_I) = L^2\rho_0 \left[1 - \exp\left(-\frac{bT_I}{L^2}\right)\right].$$

This can be considered as a second approximation to R_0 where some reduction in the number of susceptibles has been taken into account. Note that this is an idealised scenario in which density reductions due to secondary and tertiary infections are neglected.

However, the uniform density reductions in the above assume that secondary infections are equally likely at all spatial locations about the primarily infected tree, which is not the case for an infectivity kernel as in (1). On average, neglecting this spatial variation within the changing density results in an overestimation of the number of secondary infections induced by the tails of the dispersal kernel, thus giving rise to a greater R_0 value. Taking this into account, we can instead consider a spatial variant density of the form

$$\rho(r, T) = \rho_0 \exp\left(-\frac{b}{2\pi l^2} T g(r; l)\right), \tag{B.2}$$

where $g(r; l)$ is a Gaussian kernel as in (1). As above, the total number of new secondary infections from the primary infection during a set time period, $0 \le t \le T$, at a certain distance r, $N_I(r, T)$, can be calculated from the overall change in density

$$N_I(r, T) = L^2[\rho(r, t = 0) - \rho(r, t = T)].$$

For all possible values of r this becomes

$$\begin{aligned} N_I(T) = \int_0^\infty 2\pi r(\rho_0 - \rho(r, T))dr &= \int_0^\infty 2\pi r\rho_0 \\ &\times \left[1 - \exp\left(-\frac{b}{2\pi l^2} T g(r; l)\right)\right] dr. \end{aligned}$$

Here the finite lattice square of area L^2 has been replaced with integration in polar coordinates over dr. Although the above is difficult to solve directly, it can be integrated by performing a series expansion on the exponential term and integrating on a term-by-term basis, as follows:

$$
\begin{aligned}
N_I(T) &= \int_0^\infty 2\pi r\rho_0 \left[1-\exp\left(-\frac{b}{2\pi l^2}Tg(r;l)\right)\right]dr,\\
&= 2\pi\rho_0\int_0^\infty r\left[1-\sum_{n=0}^{\infty}\frac{(-\frac{b}{2\pi l^2}T)^n}{n!}\exp\left(-\frac{r^2}{2l^2}\right)\right]dr,\\
&= 2\pi\rho_0\int_0^\infty r\left[1-\sum_{n=0}^{\infty}\frac{(-1)^{n+1}(\frac{b}{2\pi l^2}T)^n}{n!}\exp\left(-\frac{nr^2}{2l^2}\right)\right]dr,\\
&= 2\pi\rho_0\sum_{n=0}^{\infty}\frac{(-1)^{n+1}(\frac{b}{2\pi l^2}T)^n}{n!}\int_0^\infty r\exp\left(-\frac{nr^2}{2l^2}\right)dr,\\
&= 2\pi\rho_0 l^2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}(\frac{b}{2\pi l^2}T)^n}{(n)(n!)}.
\end{aligned}
$$

If $b/2\pi l^2$ and T are small, the first order term in the above equation is sufficient to approximate $N_I(t)$ as a linear function of T. Finally, the above can be summed to give

$$
\begin{aligned}
N_I(T) &= 2\pi\rho_0 l^2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}\left(\frac{b}{2\pi l^2}T\right)^n}{(n)(n!)},\\
&= 2\pi\rho_0 l^2\left[E_1\left(\frac{b}{2\pi l^2}T\right)+\ln\left(\frac{b}{2\pi l^2}T\right)+\Gamma\right],
\end{aligned}
\tag{B.3}
$$

where the function $E_1(x)$ is the mathematically well studied exponential function $E_1(x) = \int_x^\infty t^{-1}\exp(-t)dt$ and Γ is the Euler–Mascheroni constant ≈ 0.57721. The estimated value of the basic reproduction number is therefore

$$
R_0 \equiv N_I(T_I) = 2\pi\rho_0 l^2\left[E_1\left(\frac{b}{2\pi l^2}T_I\right)+\ln\left(\frac{b}{2\pi l^2}T_I\right)+\Gamma\right].
$$

References

Andersson, H.K., Britton, T., 2000. Stochastic Epidemic Models and their Statistical Analysis. In: Lect. Notes Stat., Springer-Verlag, New York, p. x+137.

Bjørnstad, O.N., Shea, K., Krzywinski, M., Altman, N., 2020. The SEIRS model for infectious disease dynamics. Nature Methods 17 (6), 557–559.

Bolker, B.M., 1999. Analytic models for the patchy spread of plant disease. Bull. Math. Biol. 61, 849–874.

Van den Bosch, F., McRoberts, N., Van den Berg, F., Madden, L.V., 2008. The basic reproduction number of plant pathogens: matrix approaches to complex dynamics. Phytopathology 98 (2), 239–249.

Brown, D.H., Bolker, B.M., 2004. The effects of disease dispersal and host clustering on the epidemic threshold in plants. Bull. Math. Biol. 66, 341–371.

Bullock, J.M., Mallada González, L., Tamme, R., Götzenberger, L., White, S.M., Pärtel, M., Hooftman, D.A.P., 2017. A synthesis of empirical plant dispersal kernels. J. Ecol. 105 (1), 6–19.

Cornell, S.J., Suprunenko, Y.F., Finkelshtein, D., Somervuo, P., Ovaskainen, O., 2019. A unified framework for analysis of individual-based models in ecology and beyond. Nature Commun. 10 (1), 4716.

Cunniffe, N.J., Cobb, R.C., Meentemeyer, R.K., Rizzo, D.M., Gilligan, C.A., 2016. Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in California. Proc. Natl. Acad. Sci. 113 (20), 5640–5645.

Cuthbert, R.N., Bartlett, A.C., Turbelin, A.J., Haubrock, P.J., Diagne, C., Pattison, Z., Courchamp, F., Catford, J.A., 2021. Economic costs of biological invasions in the United Kingdom. NeoBiota 67, 299–328.

Diekmann, O., Heesterbeek, J.A.P., 2000. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. In: Wiley Series in Mathematical and Computational Biology, John Wiley and Sons, United States.

Van den Driessche, P., Watmough, J., 2002. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. 180 (1–2), 29–48.

Fabre, F., Coville, J., Cunniffe, N.J., 2021. Optimising reactive disease management using spatially explicit models at the landscape scale. In: Plant Diseases and Food Security in the 21st Century. Springer, pp. 47–72.

Fearnhead, P., Giagos, V., Sherlock, C., 2014. Inference for reaction networks using the Linear Noise Approximation. Biometrics 70, 457–466.

Filipe, J.A.N., Cobb, R.C., Meentemeyer, R.K., Lee, C.A., Valachovic, Y.S., Cook, A.R., Rizzo, D.M., Gilligan, C.A., 2012. Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: Sudden oak death in Northern Californian Forests. PLoS Comput. Biol. 8 (1), 1–13.

Filipe, J., Maule, M., 2003. Analytical methods for predicting the behaviour of population models with general spatial interactions. Math. Biosci. 183 (1), 15–35.

Freer-Smith, P.H., Webber, J.F., 2017. Tree pests and diseases: the threat to biodiversity and the delivery of ecosystem services. Biodivers. Conserv. 26 (13), 3167–3181.

Fuchs, C., 2013. Inference for Diffusion Processes: With Applications in Life Sciences. Springer Science & Business Media.

Gamerman, D., Lopes, H.F., 2006. Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference. CRC Press.

Gertsev, V.I., Gertseva, V.V., 2004. Classification of mathematical models in ecology. Ecol. Model. 178 (3–4), 329–334.

Gillespie, D.T., 2000. The chemical Langevin equation. J. Chem. Phys. 113, 297–306.

Goleniewski, G., Newton, A.C., 1994. Modelling the spread of fungal diseases using a nearest neighbour approach: effect of geometrical arrangement. Plant Pathol. 43 (4), 631–643.

Golightly, A., Wadkin, L.E., Whitaker, S.A., Baggaley, A.W., Parker, N.G., Kypraios, T., 2023. Accelerating Bayesian inference for stochastic epidemic models using incidence data. Stat. Comput. 33 (6), 1–18.

Grosdidier, M., Ioos, R., Husson, C., Cael, O., Scordia, T., Marçais, B., 2018. Tracking the invasion: dispersal of hymenoscyphus fraxineus airborne inoculum at different scales. FEMS Microbiol. Ecol. 94 (5).

Heffernan, J.M., Smith, R.J., Wahl, L.M., 2005. Perspectives on the basic reproductive ratio. J. R. Soc. Interface 2 (4), 281–293.

Jeger, M.J., Van den Bosch, F., 1994. Threshold criteria for model plant disease epidemics. I. Asymptotic results. Phytopathology 84 (1), 24–27.

van Kampen, N.G., 2001. Stochastic Processes in Physics and Chemistry. North-Holland.

Keeling, M.J., 1999. The effects of local spatial structure on epidemiological invasions. Proc. R. Soc. B 266 (1421), 859–867.

Keeling, M.J., Eames, K.T.D., 2005. Networks and epidemic models. J. R. Soc. Interface 2 (4), 295–307.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. Math. Phys. 115 (772), 700–721.

Komorowski, M., Finkenstadt, B., Harper, C., Rand, D., 2009. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. BMC Bioinform. 10 (1), 343.

Kurtz, T.G., 1972. The relationship between stochastic and deterministic models for chemical reactions. J. Chem. Phys. 57, 2976–2978.

Li, J., Blakeley, D., Smith, R.J., 2011. The failure of R0. Comput. Math. Methods Med. 2011.

Meentemeyer, R.K., Cunniffe, N.J., Cook, A.R., Filipe, J.A.N., Hunter, R.D., Rizzo, D.M., Gilligan, C.A., 2011. Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in California (1990–2030). Ecosphere 2 (2), art17.

Mikaberidze, A., Mundt, C.C., Bonhoeffer, S., 2016. Invasiveness of plant pathogens depends on the spatial scale of host distribution. Ecol. Appl. 26 (4), 1238–1248.

Minter, A., Retkute, R., 2019. Approximate Bayesian computation for infectious disease modelling. Epidemics 29, 100368.

Nathan, R., Katul, G.G., Bohrer, G., Kuparinen, A., Soons, M.B., Thompson, S.E., Trakhtenbrot, A., Horn, H.S., 2011. Mechanistic models of seed dispersal by wind. Theor. Ecol. 4, 113–132.

Nathan, R., Klein, E., Robledo-Arnuncio, J.J., Revilla, E., 2012. Dispersal Kernels, vol. 15, Oxford University Press, Oxford, UK.

North, A.R., Godfray, H.C.J., 2017. The dynamics of disease in a metapopulation: The role of dispersal range. J. Theoret. Biol. 418, 57–65.

Orozco-Fuentes, S., Griffiths, G., Holmes, M.J., Ettelaie, R., Smith, J., Baggaley, A.W., Parker, N.G., 2019. Early warning signals in plant disease outbreaks. Ecol. Model. 393, 12–19.

Parnell, S., Gottwald, T.R., van den Bosch, F., Gilligan, C.A., 2009. Optimal strategies for the eradication of Asiatic Citrus Canker in heterogeneous host landscapes. Phytopathology 99 (12), 1370–1376.

Prussin II, A.J., Marr, L.C., Schmale III, D.G., Stoll, R., Ross, S.D., 2015. Experimental validation of a long-distance transport model for plant pathogens: Application to fusarium graminearum. Agricult. Forest Meteorol. 203, 118–130.

Segarra, J., Jeger, M.J., Van den Bosch, F., 2001. Epidemic dynamics and patterns of plant diseases. Phytopathology 91 (10), 1001–1010.

Suprunenko, Y.F., Cornell, S.J., Gilligan, C.A., 2021. Analytical approximation for invasion and endemic thresholds, and the optimal control of epidemics in spatially explicit individual-based models. J. R. Soc. Interface 18 (176), 20200966.

The Tree Council, 2014. Chalara in non-woodland situations: Findings from a 2014 study.

Thompson, S.E., Assouline, S., Chen, L., Trahktenbrot, A., Svoray, T., Katul, G.G., 2014. Secondary dispersal driven by overland flow in drylands: Review and mechanistic model development. Mov. Ecol. 2 (1), 1–13.

Tildesley, M.J., Keeling, M.J., 2009. Is R0 a good predictor of final epidemic size: Foot-and-mouth disease in the UK. J. Theoret. Biol. 258 (4), 623–629.

Wadkin, L.E., Branson, J., Hoppit, A., Parker, N.G., Golightly, A., Baggaley, A.W., 2022. Inference for epidemic models with time-varying infection rates: Tracking the dynamics of oak processionary moth in the UK. Ecol. Evol. 12 (5), e8871.

Wang, X., Song, X., 2008. Mathematical models for the control of a pest population by infected pest. Comput. Math. Appl. 56 (1), 266–278.

Wang, W., Zhao, X.-Q., 2012. Basic reproduction numbers for reaction-diffusion epidemic models. SIAM J. Appl. Dyn. Syst. 11 (4), 1652–1673.

Wylder, B., Biddle, M., King, K., Baden, R., Webber, J., 2018. Evidence from mortality dating of fraxinus excelsior indicates ash dieback (Hymenoscyphus fraxineus) was active in England in 2004–2005. For.: Int. J. For. Res. 91 (4), 434–443.