UNIVERSITY OF LEEDS

Modelling spatio-temporal tree disease epidemics in Great Britain

John W. Holden

Submitted for the degree of Doctor of Philosophy

University of Leeds

Faculty of Environment, University of Leeds

December 2021

Intellectual Property

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Signed: J. Holden

Acknowledgements

Firstly, I would like to thank my primary supervisor Dr James Smith for the support, guidance, great company (particularly in Brazil) and everything he taught me along the way. My gratitude also extends to my co-survivors in Newcastle, Dr Nick Parker and Dr Andrew Baggeley, for all their support and insight. In particular, I'm deeply grateful for being hosted in Newcastle during the first year. Similarly, I want to thank my co-supervisors in Leeds, Dr Melvin Holmes and Dr Rammile Ettelaie, for their help, support and encouragement. In addition, I thank Dr Sam Grant for introducing me to the role of policymaking in tree health and for helpful discussions. I'm also thankful for the generous funding made available from DEFRA and Newcastle University. Additionally, I thank Dr Siro Orozco-Fuentes, who gave this project a firm footing during the initial phases.

I'm forever indebted to my friends and family for giving reason to the madness and hardship. I sincerely thank Dr Marcin Kupilas for being a magnificent friend and an influential part of this PhD adventure; his curiosity and endless desire to learn inspired me throughout the project. Moreover, I'm deeply grateful to my good friend, Bradley Lister, for being there throughout the tough times. Brad's unconditional support, ongoing encouragement, and positive energy kept me sane during the lockdown. Likewise, I want to thank my wonderful cousin, Luke Holden. He was always there when times went beyond challenging and consistently offered support, perspective and a positive mindset.

Lastly, I wholeheartedly thank my parents, Sarah and Tony Holden, and little sister Sophie for their undying support and constant belief in me. Without them, my PhD journey would be impossible.

Abstract

Presently, tree populations worldwide face unprecedented threats from invasive pests and pathogens endangering biodiversity, timber production and human wellbeing. From first principles, this thesis incrementally extends a simple percolation model of forest-based epidemics into a more involved stochastic dispersal framework combined with tree canopy data. The approach developed here couples two spatially-explicit epidemic models at different scales. First, a non-local stochastic model of pathogen dispersal between trees is constructed. Second, the small-scale epidemic model is projected onto a large-scale distribution of host abundance, resulting in an R_0-map across Great Britain. Subsequently, a clustering algorithm is employed to identify high-risk regions in the R_0-map. Initial results indicate a global epidemic phase transition across the distribution, conditional on an infectivity parameter. The approach to 'spatially scale-up' an epidemic model over the entire landscape is computationally efficient, flexible and adaptable to many pests and pathogens. In addition, numerous studies have sought to understand and optimise epidemic control in botanical populations. The mainstream control paradigm generally seeks to optimise an 'eradication radius' about infected symptomatic trees over a relatively small spatial scale. However, large-scale epidemic control based solely on the spatial distribution of hosts has yet to be explored in depth. As such, this thesis will also examine how host heterogeneity, combined with targeted epidemic control, can give rise to natural 'pinch-points' that may slow the epidemic spread between regions. Ultimately, this investigation intends to help policymakers reach informed decisions about where to focus control in the landscape of Great Britain.

Contents

List of Figures

List of Tables

Abbreviations

ADB Ash dieback.
ALB Asian longhorn beetle.

BSBI Botanical Society of Britain and Ireland.

CCA Connected component analysis.
COM Center of infectious mass.
CS Countryside survey.

DED Dutch elm disease.
DEFRA Department for Environment Food and Rural Affairs.

EID Emergent infectious disease.
EWS Early warning signal.

FE *Fraxineus excelsior*.
FKPP Fisher-Kolmogorov-Petrovsky-Piskunov.

GB Great Britain.

HF *Hymenoscyphus fraxineus*.

LDD Long distance dispersal.

NFI National forest inventory.
NLM Non-local model.
NN Nearest-neighbour.

OPM Oak processionary moth.

PDE Partial differential equation.
PDF Probability density function.
PM Participatory modelling.
PR *Phytophthora ramorum*.

RMSE Root-mean-square error.

SDM Species distribution models.
SEIR Susceptible-exposed-infected-removed.
SIR Susceptible-infected–removed.
SLM Simple lattice model.
SOD Sudden oak death.

WSR Wheat stripe rust.

Chapter 1

Introduction

Currently, pests and pathogens threaten the survival of numerous tree species around the globe. An upsurge in the trade and transport of non-native plants and the widespread adoption of monoculture plantations have drastically increased the risk of large-scale outbreaks in native tree populations. Moreover, following the effects of climate change and increased temperatures, the threats posed by invasive non-indigenous pathogens are only set to increase. Accordingly, efforts to understand the precise mechanisms that underpin epidemics in tree populations are essential for human wellbeing and ecosystem stability.

When an epidemic takes hold, numerous challenges complicate an effective on-the-ground response. In addition, epidemic drivers are multifaceted, hard to quantify, and often unknown. Consequently, epidemic models are significantly challenging to parameterise. Further still, communicating research insight to policymakers and stakeholders poses a significant challenge for modellers even with accurate parameterisation.

Despite the complex challenges that threaten tree health, scientists, policymakers, and stakeholders can cooperate to prevent the spread of emergent infectious tree diseases within a country. Notably, the output from mathematical models can help advise practical management strategies to control epidemics through rural and urban tree populations. Accurate epidemic models can also help forest and woodland managers structure plantations to minimise epidemic impact. Aiding forest managers with mathematical models has become particularly vital given the UK government's recent measures to sequester carbon from the atmosphere with land-use change and afforestation, thereby increasing tree coverage throughout Great Britain.

This Chapter introduces the reasons, challenges, and value in modelling tree disease epidemics. First, the overarching rationale behind researching tree disease models is highlighted, followed closely by reviewing the most pressing epidemic drivers. Second, the challenge of implementing epidemic control is summarised, along with the principal socioeconomic relationships between modellers, policymakers and stakeholders. Finally, the Chapter concludes by outlining the topics covered in this thesis.

1.1 Invasive tree diseases

The modern world relies heavily on imports and exports characterised by global trade networks. Unfortunately, the trade and transport of foreign plant material can introduce invasive pests and pathogens into non-native landscapes. As such, crops, flowering plants and tree populations that lack evolutionary adaptations to these invasive (pathogenic) species face unprecedented perils [Wingfield et al., 2010]. Epidemics through botanical populations can be devastating. Classic historical examples include Irish potato blight, Dutch elm disease (DED) [Harwood et al., 2011] and North American chestnut blight [Worthen et al., 2010]. Two high-profile epidemics currently underway in the UK include ash dieback (ADB) affecting European ash *Fraxinus excelsior* (FE) [Hill et al., 2019], and *Phytophthora ramorum* (PR), a prevalent disease affecting over 150 plant species, including oak, larch, and sweet chestnut [Brasier and Webber, 2010].

Given the fundamental significance of tree populations in terrestrial ecosystems, ensuring tree health forms a critical challenge for society to address. In particular, policymakers can lead control initiatives to help impede the spread of disease [Gilligan, 2002]. However, implementing an uninformed and ineffective management strategy is costly and does little to stop the spread. Accordingly, research into optimal disease management underpins the central issue of botanical epidemiology, where research should ideally seek to minimise both epidemic and economic impact [Hill et al., 2019, Freer-Smith and Webber, 2017, Boyd et al., 2013a, Tyrväinen et al., 2005]. Frequently, control strategies include reducing tree densities [Pietzsch et al., 2021, Cobb et al., 2017] and planting genetically diverse host distributions [Sconyers et al., 2005, Zhu et al., 2000, Huang et al., 1980].

1.1.1 Epidemic drivers

The terms 'pest' and 'pathogen' describe a broad spectrum of taxonomically diverse organisms. Pests denote any organism that harms humans or human interests, such as crops or livestock [Buckle and Smith, 2015, Oerke, 2006, De Bach et al., 1964]. Overwhelmingly, insects constitute the main pest threats to tree species [Metcalf and Luckmann, 1994]. In Great Britain (GB), the Asian longhorn beetle (ALB) [Haack et al., 2010], and oak processionary moth (OPM) [Tomlinson et al., 2015] are two such pests that currently threaten tree health. In contrast, the term 'pathogen' describes any organism that induces disease. In the context of tree populations, diseases include fungi, bacteria, viruses, and oomycetes [Balloux and van Dorp, 2017, Boyd et al., 2013b]. Presently, the oomycete *P. ramorum* [Brasier et al., 2005], and ADB caused by the fungus *H. fraxineus* (HF) [Hill et al., 2019, Mitchell et al., 2014] are two pathogens that threaten tree-health in GB.

Tree pathogens have evolved diverse reproductive modes and epidemiologies, characterising a distinctive challenge when controlling tree-based pathogens. Two contrasting examples include chestnut blight caused by the bark pathogen fungus *Cryphonectria parasitica* and the xylem inhibiting bacterium *Xylella fastidiosa*[1]. Chestnut blight devastated American chestnut populations in the 1930s [Worthen et al., 2010], and more recently, infections have been spreading through European chestnuts, albeit to a lesser

---

[1] Several diseases are symptomatic expressions of *X. fastidiosa*, including phoney peach disease, quick olive decline, coffee leaf scorch, and most notably, Pierce's disease affecting Grapes and citrus variegated chlorosis [Hopkins and Purcell, 2002].

extent [Heiniger and Rigling, 1994]. The principal infection mechanism of *C. parasitica* occurs through windborne spore dispersal into wounds or openings in bark. In contrast, transmission of the bacterium *X. fastidiosa* is thought to occur exclusively through sap-sucking insect vectors, e.g leafhoppers of the genus *Homoptera* [Redak et al., 2004]. One remarkable facet of *X. fastidiosa* is its ability to infect numerous tree species. Current estimates indicate the host range of *X. fastidiosa* is upwards of 611 different plant species [(EFSA) et al., 2022], whereas *C. parasitica* predominantly infects chestnut species and occasionally nearby oak [Rigling and Prospero, 2018].

The trade and transport of non-native plant material are widely recognised to risk introducing invasive pests and pathogens into native landscapes [Potter and Urquhart, 2017, Lovett et al., 2016, Roy et al., 2014]. Ensuing epidemics can overwhelm botanical species lacking genetic resistance to the invasive species [Desprez-Loustau et al., 2016], better understood from an evolutionary perspective: in a natural environment unaltered by human transportation, tree species co-evolve with invasive pathogens in a 'gene-for-gene' arms race [Thrall and Burdon, 2003, Dangl and Jones, 2001, Flor, 1971]. The spread of Dutch elm disease in the UK [Harwood et al., 2011] and chestnut blight in North America [Worthen et al., 2010] are two classic examples that shook the world.

Trade regulations on plant imports are essential to prevent the initial introduction of invasive pathogens [Rodoni, 2009], exemplified by the Dutch elm disease epidemic in the UK. Had more stringent regulations been active in the 1960s, elm timber infected with scolytid bark beetles (carrying the fungus *Ophiostoma novo-ulmi*) might have prevented the devastating outbreak [Harwood et al., 2011]. Ordinarily, these preventative measures take the form of customs checks on imported/exported timber, crops or horticultural goods. Recently, the European Commission enacted plant passports to regulate how growers and traders can transport plant material between countries [Wulfert et al., 2010].

If checks and policy implementations fail, a pathogen might be introduced into the landscape and spread through natural dispersal pathways. Alternatively, a pathogen might be transported into foreign ecosystems through atmospheric long distance dispersal (LDD)

[Brown and Hovmøller, 2002]. In any case, biological control becomes necessary. The biological control of plant-based disease can be achieved in numerous ways. Commonly, this involves chemical agents such as pesticides, predatory insects or planting genetically resistant cultivars [Pal and Gardener, 2006, Baker et al., 1974].

In this work, we aim to construct epidemic models and investigate eradication strategies entailing the removal of trees through sanitation felling. In this pursuit, three questions are vital to consider: (1) How do we effectively identify an infected tree? (2) How can we optimise tree felling to slow the spread of disease given limited resources/budgets? (3) What is the risk that a large-scale epidemic will result from the initial observations of diseased trees?

1.2 Modelling and policy

The benefit of controlling an epidemic should outweigh the costs of letting an outbreak spread unchecked. Plant disease modellers can help infer well-designed control policies that maximally reduce epidemic impact and minimise the expenditure of resources—both natural and economic. However, achieving this in practice is problematic due to various unknowns [Cunniffe et al., 2015a], and history gives examples of insufficient control policies that fail to halt pathogen spread. Prominent examples include the mismanagement of Dutch elm disease in the late 1960s and early 1970s [Tomlinson and Potter, 2010], and more recently citrus canker in Florida [Schubert et al., 2001].

We can attempt to understand what dictates the optimal control of tree disease epidemics with mathematical models. Control strategies have been explored on both smaller [Hyatt-Twynam et al., 2017] and larger landscapes [Forster and Gilligan, 2007]. On all spatial scales, consensus agrees that the effort and resources required to control an epidemic depend significantly on the scale of the epidemic at hand. That is, *'aggressive pathogens should be met with aggressive control'*, as confirmed by modelling studies [Gilligan et al., 2007a, Cunniffe et al., 2015b]. In any scenario, an on-the-ground response must be carried out swiftly. Otherwise, the likelihood of successful management decreases rapidly, and the cost of inaction soars.

Conventional eradication strategies involve detecting symptomatic trees and culling neighbours within a radius, e.g. [Cunniffe et al., 2015b]. A naive strategy that indiscriminately culls hosts can be fine-tuned and optimised to reduce the epidemic impact maximally. For example, one strategy prioritises targets by ranking hosts according to the expected number of secondary infections [Hyatt-Twynam et al., 2017]. Over larger spatial scales, models of sudden oak death (SOD) in California indicate that epidemics are most effectively controlled by targeting infected trees either at or ahead of the infectious wave-front [Cunniffe et al., 2016]—see section 2.2.2 for a more in-depth discussion of the manuscript. Nevertheless, various factors complicate eradication regardless of the spatial scale, such as the cryptic nature of tree diseases and fluctuating government budgets [Bussell et al., 2019, Bussell and Cunniffe, 2020].

The management of citrus huanglongbing (HLB) disease epidemics in California provides an insightful example of how these mathematical models can help determine practical regulations to suppress the spread of disease, see [McRoberts et al., 2019]. Another relevant example comes from SOD management in California (mentioned above). In particular, modelling SOD epidemics in California indicates that culling infested trees on the disease front's leading-edge is more effective [Cunniffe et al., 2016]. Subsequently, this work helped inform targeting control in the work plan put forward by the California Oak Mortality Task Force[2] between 2017-2022.

Enhanced surveillance and monitoring strategies also seek to optimise resource allocations. Surveillance aims to detect infected individuals and disease incidence, generally requiring the collection and analysis of epidemic data [Parnell et al., 2017]. Ultimately, surveillance and monitoring comprise the last line of defence after preemptive border checks and inspections have failed to prevent the introduction of disease. Statistical approaches have been adopted to optimise the number of samples/surveys required to infer disease incidence accurately [Yamamura et al., 2016]. Moreover, optimal surveil-

---

[2] The reader can find further information on the California Oak Mortality Task Force here [Palmieri and Frankel, 2006], and download the work plan by visiting https://www.suddenoakdeath.org/

lance strategies have been examined with logistic [Parnell et al., 2012] and mechanistic [Cunniffe et al., 2015b] epidemic models—the next Chapter offers a more in-depth review of these modelling paradigms.

Whether collaboratively or individually, several governmental bodies have a stake in the UK's surveillance and tree health monitoring. For example, the National Forest Inventory (NFI) conduct general surveys to determine tree distributions and woodland composition—discussed more in section 4.2. Subsequently, output from NFI surveys is utilised by 'Forestry Research' (FR), the research agency of the Forestry Commission (FC). In addition, the FC plays a vital role in surveying landscapes for infectious tree diseases in the UK [Ryle et al., 1969, James et al., 1990], often responding to high-risk invasions in a more targeted capacity. Historically, the UK government has tasked the FC to undertake comprehensive large-scale surveys in response to Dutch elm disease in the 1960s-1970s [Potter et al., 2011] and ash dieback in 2012 [Tomlinson, 2016]. However, more recently, citizen science approaches have become increasingly utilised to monitor tree health [Brown et al., 2020], exemplified by the monitoring tool 'TreeAlert', a website owned by FR designed to collect information about tree health in woodlands and forests through user reports.

Even supposing an accurate and well-informed control strategy, a response on the ground only ensues when stakeholders implement control initiatives [Reed et al., 2018]. Here, the term 'stakeholder' is extensive, reflecting any interested individual, collective, or organisation with a stake in tree health that has the potential to influence or affect a policy direction or control decision [Brugha and Varvasovszky, 2000]. The UK's stakeholder landscape for tree health is equally broad, encompassing diverse governmental and private sector organisations and individuals. A conceptual framework for stakeholder categorisation was put forward by [Dandy et al., 2017], alongside a case study of who had a stake in ash dieback in the UK. In their analysis, [Dandy et al., 2017] listed various affected stakeholders:

- **governmental**: DEFRA, FERA, FC, FR and various local authorities

- **civil society/third sector**: National Trust, Wildlife Trust, Woodland Trust
- **private sector**: private land managers, outdoor recreationists, forest owners, citizen scientists, and the general public

Figure 1.1 presents a simplified view of the interactions which dictate research output, public awareness, policymaking and the eventual epidemic control in the UK[3]. Scientists in several disciplines, from molecular biology to mathematics, receive funding from and collaborate with policymakers and other research bodies, e.g. the Department for Environment Food & Rural Affairs (DEFRA). Policymaker-led control initiatives, resource allocation and recommendations then help to direct on-the-ground stakeholder responses. However, it is worthwhile to remark that stakeholder participation can be either voluntary or compelled by law in the UK. Typically, if the pest or pathogen possesses a significant risk, plant health authorities (e.g. FC and DEFRA) may serve a statutory plant Health notice (SPHN) to the landowner, thus mandating action.

Alternatively, scientists can engage stakeholders directly (discussed more below) or influence public opinion through outreach and scientific communication. In turn, the public can influence the decisions of policymakers by mounting sufficient political pressure [Fuller et al., 2016]. Unfortunately, several obstacles inhibit a well-informed, timely and effective response. In particular, poor accessibility to scientific research is widely-known to inhibit policy adoption, primarily because disseminating scientific information requires in-depth domain knowledge and technical skill [Jones and Kleczkowski, 2020]. In a bid to make their work more accessible to policymakers and stakeholders, modellers have endeavoured to construct user-friendly interfaces[4] [Cunniffe et al., 2015b]. Other strategies to leverage scientific output involve directly facilitating discourse between modellers and stakeholders, categorised as ‘participatory modelling’ (PM).

Recently, PM has become popular in ‘risk and natural disaster’ modelling research [Hämäläinen et al., 2020, Ravera et al., 2020, Hedelin et al., 2017]. Nonetheless, PM ap-

---

[3] In part, Figure 1.1 was informed by interviewing civil servants, researchers and policymakers at DEFRA and FERA as part of this thesis.

[4] The reader can find the user-friendly modelling interface constructed by [Cunniffe et al., 2015b] at `http://www.webidemics.com`

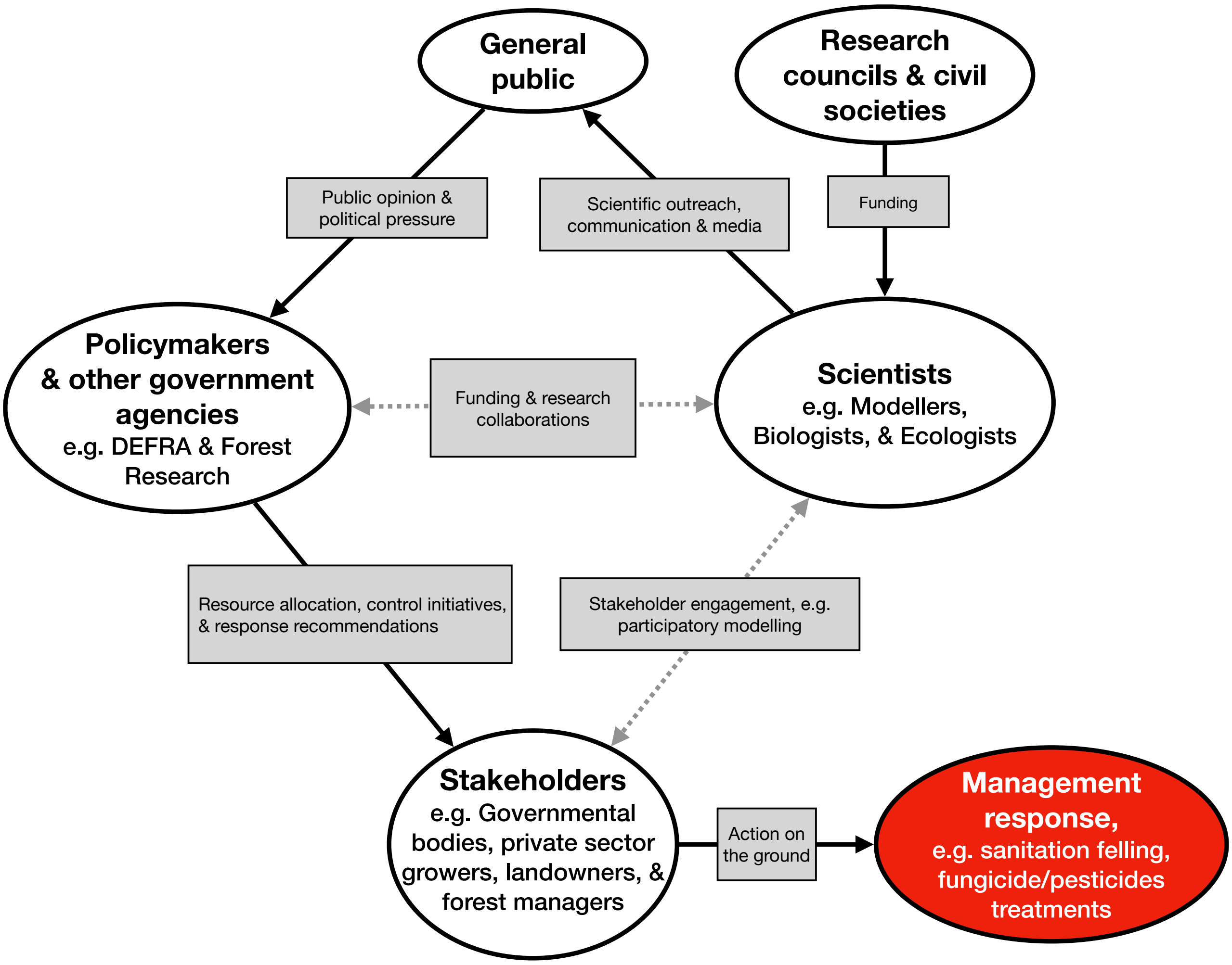


Figure 1.1: A simplified model representing the major socioeconomic interactions between the general public, scientists, policymakers and stakeholders in the UK. Scientists receive funding from and collaborate with Governmental bodies/policymakers. Policymakers make decisions and allocate resources to lead control initiatives to protect tree health. Affected stakeholders in the UK can choose to join voluntary control initiatives or be legally obliged to take action if served a statutory plant health notice.

proaches are rare in the context of plant disease, as reviewed by [Gaydos et al., 2019]. In addition to a literature review, [Gaydos et al., 2019] held an interactive workshop with stakeholders regarding the spread of *P. ramorum* in the United States. The workshop facilitated stakeholder engagement with an epidemic model [Tonini et al., 2017]—reviewed in Chapter 2. In particular, the authors reported that the stakeholders engaged well with the model and confirmed that the results were broadly consistent with observations in the field. However, and most interestingly, stakeholders with expert knowledge of the landscape remained sceptical of the host distributions' accuracy and resolution. In particular, stakeholders with domain knowledge on the ground thought that these inaccuracies could influence spatial dynamics in the simulation. Such insights are hard to deduce for modellers who generally remain less connected to the actual landscape. As such, [Tonini et al., 2017] demonstrated a positive motivation to facilitate the collaboration between plant-health modellers and stakeholders through PM.

An effective response generally relies on widespread adoption of voluntary policies among multiple stakeholders, which is thought to depend on several additional factors. As an example, [Milne et al., 2020] coupled an epidemic model of citrus huanglongbing disease (HLB) and stakeholder opinion dynamics. In the behaviour model, stakeholder opinions depended on research, other citrus growers, consultants, and the media. The perception of risks and trust in area-wide control led the stakeholders to join an area-wide control initiative. Subsequently, the analysis of [Milne et al., 2020] suggests that the efficiency of epidemic control led to more stakeholder-engagement than the perceived risks, and that frequent contact between stakeholders and advisors increases the probability of successful control.

1.3 Chapter summary

In this thesis, our motivation is to develop robust epidemic models of infectious tree disease epidemics to predict epidemic severity over GB and help guide policymakers. Firstly, we begin with a simple two-parameter percolation, from which more realistic and elaborate dispersal models follow. In particular, previous large-scale investigations have focused on specific pathosystems in a dynamic metapopulation-like setting [Cunniffe et al., 2016, Meentemeyer et al., 2011, Harwood et al., 2009]. We take an alternative approach and develop a general-purpose framework to spatially scale up a small-scale epidemic model (between individual trees) over large areas. The result is an R_0-map across GB with closer parallels to the emerging field of Infectious Disease Cartography in human and livestock diseases [Otieno et al., 2021, Kraemer et al., 2016, Messina et al., 2016].

Chapter 2 begins by outlining several requisite modelling themes. First, the review examines several seminal works that founded the field of quantitative botanical epidemiology. Following this, a suite of small, large and multi-scale spatial epidemic models are reviewed. Lastly, a case study of the emerging ash dieback epidemic is presented.

Chapter 3 sets the scene with a percolation-based simple lattice model (SLM) of tree disease spreading through a dense forest [Orozco-Fuentes et al., 2019]. The model is compartmentalised into a susceptible-infected-removed (SIR) framework and demonstrates a sharp transition threshold above which an epidemic will propagate. Above the threshold of transition, a travelling wave-like behaviour is demonstrated.

Chapter 4 builds on the percolation model constructed in Chapter 3. Firstly, we extend the work of [Orozco-Fuentes et al., 2019] and present an alternative method to detect an early warning signal in two-dimensional parameter space. Then, the Chapter moves toward a more applied setting by introducing several available host datasets in GB before proceeding with the maps of predicted abundance given by [Hill et al., 2017]. Thus, the percolation model is coupled with a map of predicted oak abundance over GB to outline a large-scale ‘toy’ model. Chapter 4 demonstrates that nearest-neighbour interactions are insufficient for describing epidemics through low tree densities across GB.

Chapter 5 introduces a generic Gaussian dispersal kernel into the epidemic model, denoted as the non-local model (NLM). Ensuing epidemics in the NLM are demonstrated to spread at lower tree densities, frequent across GB, thereby overcoming the inherent nearest-neighbour limitations witnessed in Chapters 3-4. Disease spread is then examined over a range of dispersal parameters and compared to the standard SIR model. Next, a spatially-explicit analytic expression for R_0 is derived and compared to a ‘*contact tracing*’ method of calculating R_0. Both methods of determining the reproductive ratio are shown to demonstrate a threshold at $R_0 = 1$.

Chapter 6 develops the dispersal model of Chapter 5 towards a mechanistic model reflecting the life cycle of ash dieback. The model involves susceptible-exposed-infected-removed (SEIR) compartments that repeat annually according to the sexual reproduction of ash dieback. Consequently, a method is presented to compose R_0-maps across GB using the map of predicted ash abundance given by [Hill et al., 2017]. Lastly, a connected-component-analysis (CCA) algorithm is used to visualise clustering and risk in the R_0-map. Examining the clustering as a function of infectivity reveals behaviour akin to a global epidemic phase transition across the map. That is, below a certain infectivity threshold, the pathogen would not be able to invade GB.

Chapter 7 proceeds from observations discussed in Chapter 6. Namely, Chapter 7 presents the first steps toward a novel landscape-level control strategy based on the large-scale host structure. More specifically, the epidemic control strategy targets natural pinch-points and fault lines in the spatial distribution of hosts to bottleneck the epidemic spread between at-risk regions. The Chapter ends by discussing the major assumptions in the control method and presents a series of research questions that need to be addressed before the control method is demonstrated sufficiently.

Finally, Chapter 8 presents a concrete view of the limitations and future developments of the work presented in this thesis.

Chapter 2

Interdisciplinary tree disease epidemics

Understanding modern-day tree disease epidemics requires a holistic, interdisciplinary approach, made possible only by the convergence of numerous scientific fields. Consequently, this Chapter reviews several key modelling themes. Models of tree disease aim to help design effective control policies and inform policymakers. Well-informed policymakers can help maintain tree health in rural, urban and commercial environments. However, myriad environmental, biological and anthropomorphic factors complicate the scientific understanding and thus effective disease control.

Here, the review begins by narrating some early historical developments in human and botanical epidemiology before presenting a variety of present-day modelling frameworks. After introducing the mainstream paradigm of plant-based epidemic models, an inspection of dispersal, thresholds, and epidemic control follow naturally. Finally, the Chapter ends with a case study of ash dieback, reflecting the multi-faceted difficulties posed by a recent emergent epidemic.

2.1 Historical perspectives

Historically, plant pathology and mathematical epidemiology fields existed in different spheres. Plant pathology research focused on understanding pathogen biology (e.g. [Walker, 1963]), not developing predictive mathematical theories. However, pioneering discoveries in mathematical epidemiology permitted a more quantitative treatment of botanical diseases. In particular, the seminal SIR model of [Kermack and McKendrick, 1927] provided a foundation to examine plant-based epidemics mathematically.

2.1.1 Standard SIR

The [Kermack and McKendrick, 1927] (K & M) model involves three compartmentalised fields, susceptible $S(t)$, infected $I(t)$ and removed $R(t)$. Each field models the evolution of a closed population of size N, where $N = S(t) + I(t) + R(t)$. A coupled system of ODEs then follow:

$$\frac{dS}{dt} = -\beta SI \tag{2.1}$$

$$\frac{dI}{dt} = \beta SI - \mu I \tag{2.2}$$

$$\frac{dR}{dt} = \mu I \tag{2.3}$$

where the term βSI dictates the flow of susceptible hosts into the infected compartment according to the rate β. Likewise, μI controls the transition of infected hosts into the removed compartment through a removal rate μ. Figure 2.1(a) illustrates the coupled SIR system for fixed μ and four values of β.

The coupled differential system of Equations (2.1-2.3) rely on several assumptions, including: (1) a closed population with no births or deaths (2) no exposed/incubation period (3) lifetime immunity following recovery (4) mass action population mixing, where contact mixing rates between individuals in S and I are proportional to the number of individuals in either field.

Today, countless articles have relaxed these assumptions and extended the standard SIR framework[1]. Nevertheless, a keystone result emerged from Equations (2.1-2.3), namely the existence of a critical epidemic threshold, captured through either:

$$R_0 = \frac{\beta}{\mu} \tag{2.4}$$

$$R_e = \frac{S(t)}{N}\frac{\beta}{\mu} \tag{2.5}$$

where R_0 and R_e are referred to as the basic and effective reproduction numbers, respectively. Following the introduction of one infected host at $t = 0$, $N \sim S(0)$ and we have $R_e = \big((N-1)/N\big)\beta/\mu$. Therefore, in the limit of a large population at $t = 0$, the term $(N-1)/N$ approximates unity and $R_e = R_0$, otherwise $R_e = \big(S(t)/N\big)R_0$.

Both quantities R_0 and R_e describe an epidemic threshold, though R_e captures the threshold in the face of a (finite) declining susceptible population by including the ratio $S(t)/N$. Equation 2.5 defines a critical threshold by the simple criterion[2]: (1) when $R_e > 1$, then $I(t)$ rises sharply, culminating in an epidemic before declining in the absence of newly infected hosts (2) if $R_e \leq 1$, then $I(t)$ quickly declines to zero as $t \to 0$ and the outbreak subsides.

2.1.2 Logistic growth

Numerous growth models have been used to describe plant disease, primarily based on the Gompertz [Berger, 1981], monomolecular [Neher and Campbell, 1992] and logistic [Kushalappa and Ludwig, 1982] growth equations. However, [Van der Plank, 1963] was the first to firmly establish ties between plant pathology and mathematical epidemiology using logistic growth. Fundamentally, Van der Plank modelled the growth of plant pathogens (or ‘inoculum’) by:

$$\frac{dI}{dt} = rI(1-I) \tag{2.6}$$

---

[1] Indeed, following the COVID-19 pandemic, SIR-type models remain an active field of research and dominate present-day epidemic literature [Atkeson et al., 2020].

[2] For a more comprehensive mathematical proof of SIR model thresholds, the reader is directed towards [Weiss, 2013]

where r describes the rate of pathogen growth and I reflects the proportion of infected tissue. In Equation 2.6, the amount of infected tissue I snowballs at first, in proportion to the amount of inoculum. Then, as time passes and I grows, $(1-I)$ approximates zero, and the system plateaus as all susceptible tissue becomes infected. The essential model behaviour is shown in Figure 2.1(b) over 10 growth rates parameters. From Equation 2.6, a simple method to determine the rate r follows:

$$r = \frac{1}{t_2 - t_1}\Big[\log_e(\frac{I_2}{1-I_2}) - \log_e(-\frac{I_1}{1-I_1})\Big] \tag{2.7}$$

where I_1 and $I_2 \in [0,1]$ are the proportions of infected tissue at times t_1 and t_2 respectively. Equation 2.7 is valid for all values of t, expect when $t_1 = t_2$ (see [Van der Plank, 1963] Chapter 3). Importantly, both I_1 and I_2, and by extension the infection rate r, are measurable in laboratory conditions. In Equation 2.6, newly infected tissue becomes infectious immediately following infection. Realistically, infectious tissue (and symptom expression) takes time to develop, described by an 'incubation period'. Accordingly, Van der Plank adapted Equation 2.6 to a delay differential equation (DDE):

$$\frac{dI_t}{dt} = RI_{t-p}(1-I_t) \tag{2.8}$$

where I_t and I_{t-p} describe the infectious tissue at times t and $t-p$ and p is the incubation period. Hence, infectious tissue grows in response to the factor RI_{t-p}, and saturates according to the logistic term $(1-I_t)$. The parameter R now describes pathogen growth at step $t-p$, as opposed to r that describes the spread of disease at step t, leading to the ratio:

$$\frac{R}{r} = \frac{x_t}{x_{t-p}} \tag{2.9}$$

where x_t represents the proportion of inoculum at time t, along with both r and R that denote the 'apparent' infection rate (measurable in laboratories from Equation 2.7) and the 'basic' infection rate, respectively. As Van der Plank explained, one usually seeks to determine R from the constant r. If r indeed stays constant over the epidemic, the basic

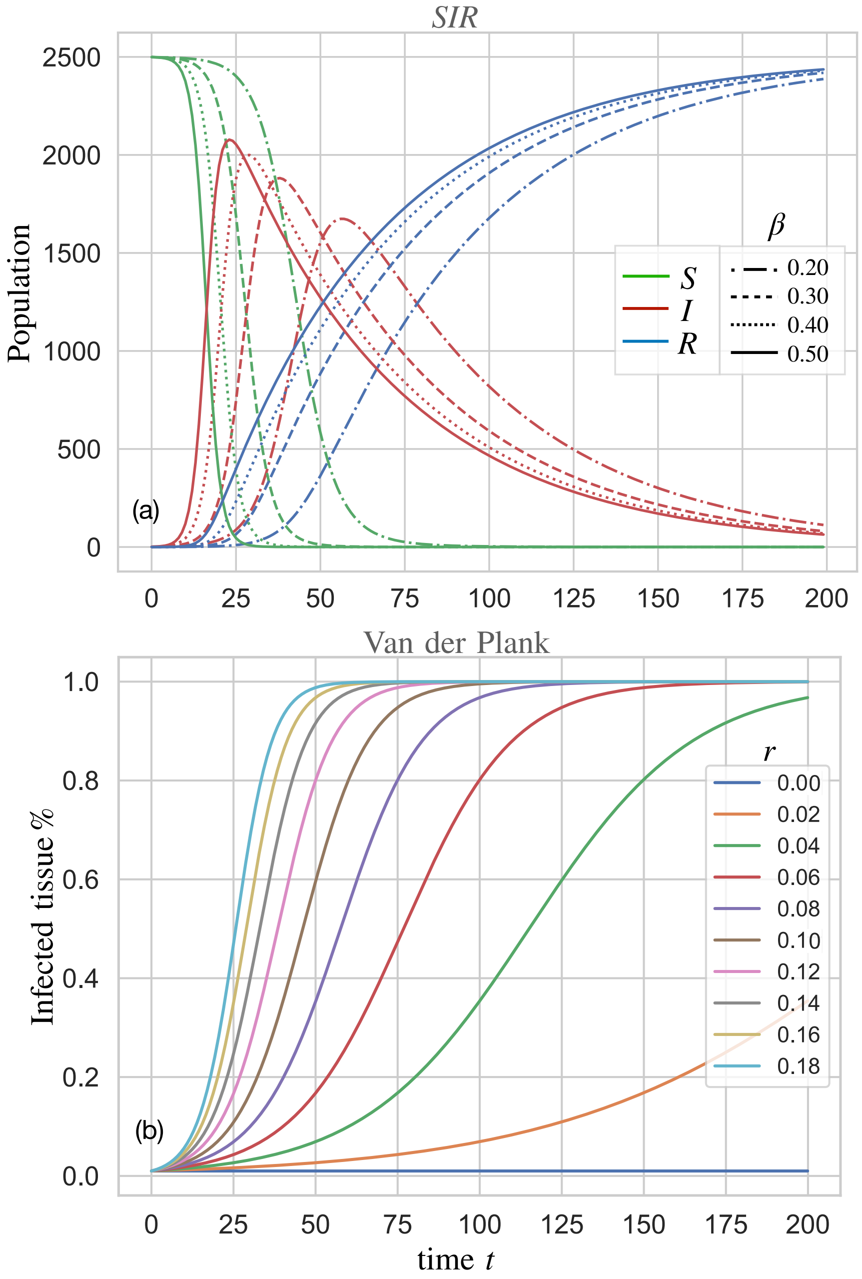


Figure 2.1: (a) The SIR model as presented by [Kermack and McKendrick, 1927] is shown for fixed removal rate $\mu = 0.20$ and variations of β. The coupled system of ODEs can be solved numerically with Euler's method. Here, simulations begin with 2500 susceptible and one infected individuals, and evolve for $t = 200$ steps. (b) The logistic growth model [Van der Plank, 1963] used to describe the growth of infected plant material. Simulations are numerically computed with a forward-time finite difference method for 10 different growth rates r.

infection rate R must decrease as x_t/x_{t-p} gets progressively smaller, i.e. as $x_{t-p} \rightarrow x_t$. At this point, the basic infection rate R begins to resemble the effective reproduction ratio R_e from Equation 2.5 because it takes into account a changing population.

Equation 2.8 describes a system where infections incubate for a period p before inducing the growth of more infectious plant tissue. However, it does not describe the infectious period (where infected tissue remains before becoming epidemiologically inert), which lead Van der Plank to extend Equation 2.8 to:

$$\frac{dI_t}{dt} = R_c(I_{t-p} - I_{t-i-p})(1 - I_t) \tag{2.10}$$

where R_c is the basic infection rate ‘corrected’ for removals and i is the infectious period. In this DDE, a unit of latently infectious tissue starts becoming infectious after p steps, and stops becoming infectious i steps. More formally, as $I_{t-i-p} \rightarrow I_{t-p}$, the rate of infectious tissue growth approaches zero, $dI_t/dt \rightarrow 0$.

2.1.3 Contrasting approaches

The SIR model does not include an incubation period, and therefore differs from the delayed differential formulation of Equation 2.10. Nonetheless, the compartmentalised SIR is easy to extend, leading to an SEIR system:

$$\frac{dS}{dt} = -\beta SI \tag{2.11}$$
$$\frac{dE}{dt} = \beta SI - \gamma E \tag{2.12}$$
$$\frac{dI}{dt} = \gamma E - \mu I \tag{2.13}$$
$$\frac{dR}{dt} = \mu I \tag{2.14}$$

where Equation 2.12 describes the population of latently (or exposed) infected hosts. Susceptible hosts transition into the exposed compartment at the same rate as before, namely βSI, but now have an exponentially distributed latency period of γ^{-1} before transitioning into I.

Both equations 2.11-2.14 and 2.10 outline similar behaviour, namely an incubatory period following infection. However, infected tissue in Equation 2.10 remains infectious for precisely i units, which contrasts with the exponentially distributed exposed lifetime implicit within Equation 2.12. Nevertheless, [Segarra et al., 2001] illustrated how Van der Plank's DDE and the SEIR system are both special cases of the SIR model proposed by [Kermack and McKendrick, 1927], when the SIR model incorporates a sporulation function, $\phi(\tau)$. The sporulation function appropriately models spore production in plant pathogens. In particular, if $\phi(\tau) = 0$ for small t, $\phi(\tau)$ can model incubation periods. Similarly, if $\phi(\tau) = 0$ for large t, we recover an infectious period. More recently, [Cunniffe et al., 2012] simplified the analysis of [Segarra et al., 2001], showing that both Van der Plank's DDE and the SEIR model can be recovered with an SE_nI_mR model. In an SE_nI_mR framework, both n and m represent an arbitrary number of distinct exposed and infectious compartments:

$$S \rightarrow E_1 \rightarrow E_2 \rightarrow \ldots \rightarrow E_n \rightarrow I_1 \rightarrow I_2 \rightarrow \ldots I_m$$

Here, the SEIR model is recovered when $n = 1$ and $m = 1$. However, when $n, m \rightarrow \infty$, [Cunniffe et al., 2012] demonstrated that we recover an expression equivalent to the DDE in Equation 2.10. Despite the equivalence of both DDE and SEIR formulations, the overarching theme of modern epidemiology is overwhelmingly compartmentalised, owing to the increased flexibility and easier analysis of compartmental models.

2.1.4 Progressive botanical epidemiology

After [Van der Plank, 1963] moved the field of botanical diseases into a more quantitative discipline, theoretical investigations (alongside the adoption of computer simulations) characterised the next few decades. Van der Plank's DDE was applied to numerous pathosystems, halo blight in beans [Taylor et al., 1979] and grape powdery mildew [Sall, 1980] to name a few. In particular, [Sall, 1980] adapted the logistic approach to include a time-varying infection rate $r(t)$. Moreover, diverse mathematical techniques, including multiple regression analysis, were incorporated into mainstream plant epidemiology [Butt and Royle, 1974]. Subsequently, [Zadoks et al., 1979] consolidated various early mathematical models of plant disease alongside [Jeger, 1984].

Developments in computing compounded advances in plant epidemiology through this period. Improved accessibility and computer architectures permitted faster calculations and more intensive models. The first epidemic simulator (EPIDEM) written in FORTRAN IV came by [Waggoner et al., 1969]. EPIDEM modelled the fungi *Alternaria solani* spreading through infected potato and tomato leaf tissue under different environmental conditions.

An interesting early simulator (EPIMUL76) developed by [Zadoks and Kampmeijer, 1977] adapted Van der Plank's logistic growth model into a spatio-temporal framework of two spatial dimensions. EPIMUL76 simulated the spread of disease on a two-dimensional domain, subdivided into 20×20 host units referred to as 'compartments', that took place inside a computer with 128K of memory. Arguably, subdividing the domain into separate compartments could be considered as an early agent-based model.

In their analysis, [Zadoks and Kampmeijer, 1977] alluded to the problem of scale in plant disease, as spatial scales were conceptualised as 'microscales' ($\leq$ 1m), 'mesoscales' (10^2m) and 'macroscales' (10^6m). In this picture, microscales ranged from plant leaves to individual plants, mesoscales reflected crop fields, and macroscales described large regional expanses over an entire country. Moreover, the probability of dispersal between infected hosts assumed a Gaussian distribution, in contrast to [Gregory, 1968].

In general, the ability to simulate more intricate models grew in proportion to the amount of computer memory available. For a review of early plant disease simulators, see [Teng, 1985].

2.1.5 Percolation: from forest fires to epidemics

Research on percolation occurred alongside the developing field of plant disease modelling. The development of percolation theory marked an early approach to modelling epidemic systems that are both spatially explicit and stochastic. The original formulation of percolation described the gelation of a fluid by modelling the bonds that form between molecules [Broadbent and Hammersley, 1957]. Moreover, these early formulations of percolation relied on a graph—illustrated in terms of vertices and edges. However, physicists studying material sciences put forward re-interpretations, naturally on a lattice [Essam, 1980].

An essential characteristic of percolation is its ability to describe a phase transition. More specifically, an abrupt change in the state of a system occurs when the system reaches a percolation threshold. Interestingly, percolation thresholds are treatable with scaling theory, used to study critical phenomena. In particular, when percolation is around the threshold, specific characteristics (e.g. cluster size and cluster numbers) were found to scale according to a 'critical exponent.' Amazingly, critical exponents remain fixed, regardless of the spatial dimension or the type of lattice geometry. Accordingly, early work continued study percolation around criticality in terms of critical exponents [Stauffer, 1979]. An expanded discussion of percolation theory continues in Chapter 3.

Different flavours of percolation models, such as site or bond percolation, were described to model different processes. Nevertheless, percolation proved a convenient theory and various phenomena including gelation, magnetism and telecommunications were described [Adler et al., 1983]. With only a short conceptual jump from time-dependent percolation used to study the growth of crystals [Family and Vicsek, 1985], forest fire models were subsequently related to percolation [MacKay and Jan, 1984].

Beginning with the SIR framework, mathematical epidemiology was already well-established around the time percolation theory was conceived [Baily, 1975]. A fire spreading through a population of trees is not too different to a disease spreading through a population. Hence, a general percolation-based epidemic-formalisation framework was put forward by [Cardy and Grassberger, 1985] who proposed that epidemics might be in the same universality class as percolation.

A fractal-like pattern of epidemics was observed by [Grassberger, 1986], shown in Figure

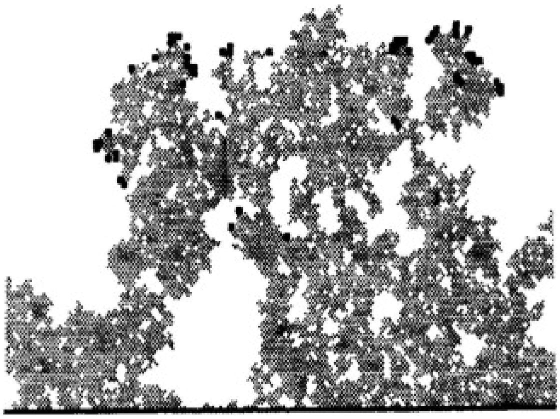


Figure 2.2: A space-time representation of an epidemic spreading at the critical threshold. The spatial (horizontal) and time (vertical) axis show self-similar propagation of diseased individuals in grey, produced by [Grassberger, 1986]

2.2. In Figure 2.2, lighter grey sites represent removed individuals, black sites indicate actively infected sites, and white sites indicate unaffected sites. All lattice sites in the bottom row were initially infected, and the infection can be seen to propagate from the bottom up. The lattice was initialised at the critical-density $p \sim p_c$ culminating in a fractal-like pattern. [Grassberger, 1986] did not attribute the hosts of this model to be trees, but instead a general host-population with low mobility. The authors noted that local interactions between hosts and infected were vast simplifications and proposed generalising the system with long-range interactions following a power law. Although, it must be remarked that including long-range interactions would cease to describe a percolation based system.

Percolation models of epidemics were examined in several contexts, including renormalisation-groups [Ohtsuki and Keyes, 1986] and Monte-Carlo methods [Janssen et al., 1988]. Studies generally focused on finding the systems critical exponents and categorising phase transition graph that characterised an epidemic (super-critical) or extinction (sub-critical) regimes [Grassberger, 1986]. The properties of both epidemics and forest fire percolation models were studied together in [Cox and Durrett, 1988], highlighting their similarity.

Early ecological applications of percolation were put forward by [O'Neill et al., 1992], who studied the effects of landscape distributions and percolation models (reflecting both forest fire and tree disease epidemics). In their study, [O'Neill et al., 1992] combined SIR-like mechanics inside a percolation-based distribution of hosts and subsequently found a narrow regime where disease epidemics would spread, thus mirroring the threshold-like behaviour witnessed in the SIR model. Although, here, the authors considered the possibility of host recovery.

A body of work applied percolation to model the economically damaging plant pathogen *Rhizoctonia Solani*. *R. Solani* affects numerous botanical hosts, including potatoes, soybean, and rice [Carling and Leiner, 1990, Chela Fenille et al., 2002, Kazempour, 2004, Tsror, 2010]. Primarily, the saprotrophic fungus spreads between infected and susceptible host roots through mycelial growth [Papavizas, 1970]. The first article put forward by [Bailey et al., 2000] relates the fungal growth dynamics of *R. Solani* to percolation. In their research, [Bailey et al., 2000] assembled a matrix of interspaced nutrient sources which together formed a triangular lattice. Experiments then sought to monitor how the spread of *R. Solani* depends on the distance between sites. Subsequently, using the conceptual percolation threshold, [Bailey et al., 2000] demonstrated the existence of a minimum distance between sites, below which no spread occurred. This result places constraints on the optimal proximity between hosts planted in a field.

Subsequent work on *R. Solani* continued to assess how random removals of nutrient sites below the percolation threshold could prevent the spread of disease [Otten et al., 2004]. In the following years, Markov chain Monte Carlo (MCMC) methods were developed by [Gibson et al., 2006] to estimate parameters for another percolation model applied to *R. Solani* [Filipe et al., 2004]. [Pérez-Reche et al., 2011] went on to put forward further generalisations of the percolation paradigm by incorporating synergistic interactions between hosts. Here, ‘synergy’ refers to the increased susceptibility of hosts surrounded by multiple infected neighbours—similarly, a reduced infection pressure results when multiple susceptibles surround an infected host. Additional work by [Ludlam et al., 2012] then formulated statistical approaches to find parameter estimates for the generalised percolation model incorporating synergistic interactions. In a later follow-up paper, [Poggi et al., 2013] proceeded to outline a risk-based percolation framework for soil-borne pathogens, assessing how the probability of invasion depends on agricultural host density.

2.2 Spatially-explicit epidemic models

Emergent infectious diseases (EIDs) in tree populations are multi-scale, spanning regional, country or continental spatial scales. Considerable variations in landscape composition, tree densities, population aggregation, and the climatic factors give rise to diverse spatio-temporal patterns of disease spread [He et al., 2019, Suzuki et al., 2003]. Incorporating spatial structure into tree disease models is therefore vital to accurately capture environmental influences, host-pathogen interactions and dispersal [Liu et al., 2007]. This section outlines several approaches to modelling the spread of tree disease over small, large and multiple spatial scales.

2.2.1 Small-scale: stochastic dispersal

Pathogen dispersal through wind, watercourse or human trade underpins a key feature of EDIs. Numerous functions have been used to model dispersal, but generally, all describe a continuous non-negative (real-valued) function that is normalisable: $\int_{-\infty}^{\infty} D(x)dx = 1$. Various review articles provide comprehensive functional examples [Bullock et al., 2017, Nathan et al., 2012, Howe and Smallwood, 1982]. However, the general class of dispersal kernels include thin-tailed Gaussian and exponential alongside fat-tailed inverse power law variants.

Early plant disease simulators (e.g. EPIDEM and EPIMUL76) included stochastic dispersal—discussed at length in section 2.1.4. A more recent article by [Parnell et al., 2010] modelled stochastic dispersal in Citrus canker to investigate the effects of landscape aggregation patterns and disease progression. [Parnell et al., 2010] studied landscape patterns of two kinds: 1) varying degrees of randomly distributed densities 2) varying degrees of host aggregation. The epidemic model was based on a series of prior Citrus canker works [Parnell et al., 2009a, Gilligan and van den Bosch, 2008, Cook et al., 2008]. The epidemic model is described by:

$$Pr(S_i \to I_i)_{\Delta t} = 1 - \exp\big[-\beta \sum_{j}^{N} \exp^{-\alpha d_{ij}} + \epsilon\big] \tag{2.15}$$

Here, a transition probability describes the i^{th} susceptible tree becoming infected during the step $t \to t + \Delta t$ on account of N infected trees. Equation 2.15 assumes that dispersal exponentially decreases with distance (i.e. $\exp^{-\alpha d_{ij}}$), where α denotes the dispersal scale parameter. Infection pressure from the j^{th} infected tree is multiplied by an infection rate, β. The last parameter to consider in Equation 2.15 is the primary infection rate ϵ. The primary infection rate reflects the chance of infection from sources outside the immediate system, i.e. at time $t = 0$, distant infected populations external to the closed host population under consideration.

The outside exponential term of Equation 2.15 depicts a cumulative exponential threshold above which trees become infected. Interestingly, this threshold is based on the 'Sellke construction' [Sellke, 1983]. More specifically, the Sellke construction assumes that susceptives need an arbitrary (cumulative) degree of infection exposure before becoming infectious. Using Equation 2.15, [Parnell et al., 2010] proceeded to define a control radius and found that both landscape aggregation and (randomly distributed) high host densities increase the optimal control radius.

A later paper by [Cunniffe et al., 2015b] proceeded to generalise the Citrus canker model. Primarily, the authors examined an $SECIR$ model[3], though several other model variants were included. [Cunniffe et al., 2015b] contrasted both Gaussian and Cauchy dispersal kernels. Cunniffe et al. included dispersal parameters provided by [Neri et al., 2014], who assessed Cauchy kernels—in addition to exponential kernels. The general model followed:

$$\phi_i(t) = w(t)\big[\beta \sum_{j} K(d_{ij}; \alpha) + \epsilon\big] \tag{2.16}$$

where $w(t)$ is a time-dependent infectivity function, ϵ is the primary infection (set to zero

---

[3] Here, compartments are (S)usceptible, (E)xposed, (C)ryptic and (R)emoved) where C denotes unobservable cryptic infections.

in the manuscript), β is the rate of secondary infection, and $K(d_{ij};\alpha)$ is the dispersal function. However, this time, $\phi_i(t)$ in Equation 2.16 presents a *rate* of transition for a single susceptible tree (from $S_i \to E_i$) under the influence of all infected neighbours (represented by the j^{th} index), as opposed to the probability of Equation 2.15 (hence its missing outside exponential term).

Cunniffe et al. examined the effects of a cull radius similar to [Parnell et al., 2010]. Yet this time, the results were aimed towards assessing control when 'there is epidemic uncertainty'. Consequently, [Cunniffe et al., 2015b] assessed different epidemic severity by varying β, alongside numerous control parameters, e.g. eradication response time, detection probabilities and revisit/survey intervals. All in all, [Cunniffe et al., 2015b] highlighted an intuitive result, namely, that the scale of control should reflect the "intrinsic epidemic scale". More succinctly, *aggressive pathogens should be met with an aggressive control strategy.* Moreover, Cunniffe et al. suggested that thick-tailed dispersal kernels (in this case, a Cauchy distribution) prove more challenging to control.

The Citrus canker models developed in Equations 2.15 and 2.16 contrast with non-spatial analytical systems. For example, the model of pine wilt disease (PWD) constructed by [Khan et al., 2020], who coupled a differential system of pine (H)osts and beetle (V)ectors following:

$$\frac{dS_H}{dt} = \lambda_H - \beta_1 \Psi S_H I_V - \beta_2 \Phi \alpha S_H I_V - \gamma_1 S_H \tag{2.17}$$
$$\frac{dE_H}{dt} = \beta_1 \Psi S_H I_V - \beta_2 \Phi \alpha S_H I_V - (\gamma_1 + m) E_H \tag{2.18}$$
$$\frac{dA_H}{dt} = m(1-\omega) E_H - \gamma_1 A_H \tag{2.19}$$
$$\frac{dI_H}{dt} = m\omega E_H - (\gamma_1 + \mu) I_H \tag{2.20}$$
$$\frac{dS_v}{dt} = \lambda_V - K S_v I_H - \gamma_2 S_V \tag{2.21}$$
$$\frac{dE_v}{dt} = K S_v I_H - (\gamma_2 + \eta) E_V \tag{2.22}$$
$$\frac{dI_v}{dt} = \eta E_V - \gamma_2 I_V \tag{2.23}$$

In this system, pine tree hosts interact with beetle vectors that carry pathogenic nematodes (*Bursaphelenchus xylophilus*). Stepping through the system, equations 2.17-2.20 describe the host population. Naturally occurring births and deaths in the host population occur at rates λ_H and γ_1, respectively. Pine trees become infected by two mechanisms: $\beta_1 \Psi$ that describes the incidence rate due to mature infected beetles, and $\beta_2 \Phi$ that describes the incidence rate due to the offspring of infected beetles. Once pine hosts become exposed, a fraction (ω) transition into the infectious symptomatic state I, while the remaining fraction $(1 - \omega)$ become asymptomatic—both pathways occur at rate m. Disease induced death happens at a rate μ. Equations 2.21-2.23 outline a (SEI) dynamic for beetle vectors. Natural births and deaths in the beetle population happen with rates λ_V and γ_2, respectively. Furthermore, susceptible beetles become exposed by feeding on infected pine trees at rate K and transition into the infected beetle class at rate η.

Equations 2.17-2.20 assume mass action population mixing of beetles without stochasticity. Presumably, this assumption led [Khan et al., 2020] to model PWD as a non-spatial system on account of the mobile population of Beetle vectors. In reality, a dispersal kernel is likely to describe beetle movements more accurately than the well-mixed system presented in Equations 2.17-2.20. Case in point, the spatio-temporal dynamics of Asian longhorned beetle were examined by [Smith et al., 2004] who inferred a median dispersal rate of 30m/day according to an exponential dispersal kernel (with only 2% of beetles exceeding 920m).

The particular method of analysis constitutes a major difference between spatially explicit stochastic models and their non-spatial analytic counterparts. Linear stability analysis was performed on the deterministic system of Equations 2.17-2.20. Whereas the stochastic spatio-temporal framework of Equations 2.15 and 2.16 were analysed by repeating simulations inside an ensemble.

2.2.2 Large-scale: landscape spread

Ultimately, microscopic (host-pathogen) interactions propagate the spread of disease. Although once disease-establishment has taken place, large-scale outbreaks can spread through vast areas, e.g. the spread of ash dieback through Europe [Alsop et al., 2015]. As a result, contemporary models of tree disease have examined the large-scale spread over entire landscapes. The previous section outlined some small-scale models (on the order of $1 \sim 10$km), however, in this section, we focus on large-scale epidemic models.

Over large scales, two primary disease drivers include long distance dispersal (LDD) through wind [Golan and Pringle, 2017, Gross et al., 2014a] and trade [Hill et al., 2019, Perrings, 2016, Harwood et al., 2009, Keeling and Eames, 2005]. However, linking human trade networks and dispersal in one large-scale model is challenging due to numerous complex parameters and epidemiological drivers.

A framework constructed by [Harwood et al., 2009] incorporated the growth and reproduction, dispersal and trade of infectious plant material into a single model. The study conducted by [Harwood et al., 2009] aimed to assess the risk of *Phytophthora ramorum* and *Phytophthora kernoviae* in the UK by employing a linked network approach. In the linked network, single grid cells of area 1km × 1km were coupled together by a (wind-borne) dispersal kernel and a trade network. Following earlier earlier work [Madden et al., 2007], the population inside each grid cell evolved according to an $SEIS$ model:

$$\frac{dS}{dt} = \mu E + \mu I - \beta SI \tag{2.24}$$

$$\frac{dE}{dt} = \beta SI - kE - \mu E \tag{2.25}$$

$$\frac{dI}{dt} = kE - \mu I \tag{2.26}$$

where host introductions were assumed to balance the total number of removals $\mu(S + E + I)$. Then, dispersal between 1km × 1km grids took place inside a domain of size 700km × 1300km covering the UK. Here, the host population was informed by the Country side survey data—discussed more below in section 4.2. An inverse square power law described wind-borne dispersal (with a scale constant of 2m), though parameterisation was qualitative and uniformed by experimental data.

In addition to wind-borne dispersal, a simulated trade network linked 1km × 1km grid cells. In the trade network, plant nurseries and retailers were connected by LDD trade and transport. In this manner, *Phytophthora ramorum* and *Phytophthora kernoviae* could jump between cells. The same linked-network approach was later used to reconstruct the highly popularised 1970s Dutch elm disease epidemic in Great Britain [Harwood et al., 2011, Potter et al., 2011].

A similar construction was put forward by [Meentemeyer et al., 2011] to forecast the spread of sudden oak death (SOD) in California from (1990-2030). A distribution of host[4] abundance was derived from previous SOD modelling work [Meentemeyer et al., 2004] and comprised 250 × 250 grid cells weighted by the relative susceptibility to *P. ramorum* from 1-100. As a result, a high-resolution map of was produced throughout the state of California detailing the 'host index' from 1-100.

The large-scale SOD model developed by [Meentemeyer et al., 2011] included several epidemiological and longer-range transmission. Local-scale dispersal were estimated using Markov chain Monte Carlo (MCMC) methods from aerial surveys [Valachovic et al., 2008] of *P. ramorum*, and positive sites (between 2001–2007) confirmed by the California Department of Food and Agriculture were used to estimate the long-range dispersal. [Meentemeyer et al., 2004] found that a long-range Cauchy distribution fitted the data most appropriately over both spatial scales. Hence, a multi-scale kernel was given as:

$$K(d; \alpha_1, \alpha_2, \gamma) = \gamma(1 + (d/\alpha_1)^2)^{-1} + (1 - \gamma)(1 + (d/\alpha_2)^2)^{-1} \tag{2.27}$$

where $\alpha_1 = 20.57$m and $\alpha_2 = 9.5$km represent the short and long range dispersal kernels respectively, and the ratio $\gamma = 0.99$ estimates the total contribution to short and long-range dispersal. Using the multi-scale dispersal kernel in Equation 2.27, the epidemiological model between 250m × 250m grid cells assume the form:

$$\Psi_{ijt} = \beta \sum_i \big(\chi_t(f_i) m_{it} c_{it} I_{it}\big)\big(\chi_t(f_j) m_{jt} c_{jt} S_{jt}/N_{max}\big) \times K(d_{ij}; \alpha_1, \alpha_2, \gamma) \tag{2.28}$$

---

[4] In this context, 'host' refers to a wide-range species susceptible to *P. ramorum* [Tooley et al., 2004].

where Ψ_{ijt} represents the infection pressure from grid i to grid j in one week t intervals. Equation 2.28 includes multiple component-functions and parameters:

- A binary-valued function $\chi_t(f_i)$ that indicates if forest type f_i can infect and become infected at time t
- two indices m_{it} and c_{it} that indicate moister and temperate of patch i at time t
- I_i and S_j, the number of infected in grid i and susceptibles at j
- $K(d_{ij})$, the dispersal kernel from Equation 2.27
- β that models the rate of spore production per site per week.

From the model, [Meentemeyer et al., 2011] predicted which areas in California had the highest secondary infection risk of SOD over 40 years. Simulations were ensemble-averaged based on predicted weather conditions from 2008–2030. Climatic variations were classified as 'favourable', 'random' and 'unfavourable' for pathogen growth. In all variations, SOD was predicted to spread through California and effect 1000s of square kilometers. However, considerable spatial and temporal variation were witnessed across different Californian states. Additionally, [Meentemeyer et al., 2011] observed that 93% of short-range dispersal occurred within the range of a single 250 × 250 and 95% of infrequent long-range spread remain within 100km in their model. Although, most dispersal remained localised < 1km.

[Meentemeyer et al., 2011] emphasised the multi-faceted parameters and processes that one needs to consider before modelling a large-scale epidemic outbreak; these included, host data, dispersal, and climate. [Cunniffe et al., 2016] subsequently extended the analysis of [Meentemeyer et al., 2011] to assess the large-scale effects of epidemic control, and particular examined how to optimise eradication of SOD with limited resources. When resources are low, the authors suggested that small localised eradication zones around *known* foci optimise control; justified by the fact that small, but more numerous, control areas about diseased areas reduces the risk of failing to treat a high-risk site that causes many secondary infections.

Furthermore, [Cunniffe et al., 2016] examined management strategies that target: (1) hosts irrespective of disease status (the "host" strategy) (2) local areas with high prevalence ("hazard" strategy) (3) areas with high basic reproduction numbers ("susceptible" strategy) (4) regions ahead of the wavefront ('wavefront" strategy). Of all the management scenarios tested, [Cunniffe et al., 2016] found that (4), treating areas ahead of the wavefront, reduced epidemic spread the most. In this scenario, the affected area was reduced by $\sim 2400\text{km}^2$ and the optimal culling radius was determined to be 362.5m.

All the large-scale models discussed above split the population into smaller (sub)girds. As such, they share noticeable similarities to a metapopulation[5] commonly used by ecologists studying spatially-structured animal and plant populations [Hanski, 1998]. However, plant-disease modellers increasingly utilise metapopulation settings to study the effects of landscape features on disease progression [Benincà et al., 2020, Soubeyrand et al., 2009, Park et al., 2002].

2.2.3 Multi-scale: disease fronts

Through the years, numerous researchers have conceptualised the spread of disease, or more broadly, biological invasions, through the lens of diffusion, random walks, or localised dispersal. As a consequence of such ideas, we see the prediction of constant travelling waves [Skellam, 1951, Mollison, 1977, Grassberger, 1983, Ferrandino, 1993]. A classic example can be found in the Fisher Kolmogorov–Petrovsky–Piskunov (FKPP) equation [Fisher, 1937].

Although recent work has called travelling waves into question for plant-based (LDD) epidemics, as we discuss more below, the FKPP model is important from a historical and contextual perspective; the model comprises some fundamental 'reaction diffusion' properties that emerge from the interplay of a populations growth and spread. In two spatial dimensions, the FKKP model is given by a partial differential equation (PDE) of

---

[5] Metapopulation dynamics generally aim to deconstruct a spatial population into separate subpopulations contained within a 'patch'. Then, between-patch interactions aim to model population migrations, connectedness and fragmentation, while within-patch dynamics aims to model colonisation, persistence, competition, coexistence, and habitat suitability.

the form:

$$\frac{\partial u}{\partial t} = \mathcal{D}\nabla u + ru(1 - u/K) \tag{2.29}$$

where $\mathcal{D}$ is a diffusion coefficient ($m^2\ t^{-1}$), r is a growth rate (t^{-1}), and K is carrying capacity. More explicitly, $\mathcal{D}$ represents a populations mobility, r represents how quickly individuals in the population reproduce, and K represents the maximum number (or concentration) of individuals can occupy a region in space at any one time. If $K = u$, the whole growth term goes to zero and we only have outward diffusion (provided neighbouring regions are not fully occupied as well).

Hence, the two terms in Equation 2.29 describe a populations logistic growth and spread, $ru(1 - u/K)$ and $\mathcal{D}\nabla u$ respectively. The essential travelling-wave behaviour of Equation 2.29 is displayed in Figure 2.3 in both one and two spatial dimensions. In all panels, travelling waves are simulated with parameters $r = 0.10$, $\mathcal{D} = 0.10$, and $K = 1.00$, and at $t = 0$, the field is set to $u(x) = 0.10$ at the domains mid-point. All simulations shown in Figure 2.3 evolve according to a forward-time centered-different (FTCD) finite scheme to second order and admit dirichlet boundary conditions.

Figures 2.3(a-c) show a one-dimensional wave propagating under different conditions: (a) symmetric growth and diffusion (b) extending a linearly increasing diffusion gradient from left ($\mathcal{D} = 0.01$ at $x = 0.00$) to right ($\mathcal{D} = 2.00$ at $x = 1.00$) (c) extending a linearly increasing growth gradient that increases from left ($r = 0.01$ at $x = 0.00$) to right ($r = 2.00$ at $x = 1.00$). Figures 2.3(d-f) show the equivalent two-dimensional behaviour.

From these conditions, we can see that $\mathcal{D}$ and r control the thickness of the wavefront, which can be inferred by the quantity $\ell_w \sim \sqrt{\frac{\mathcal{D}}{r}}$ (as can be seen by the cancellation of units, i.e. $\frac{length^2}{t}\frac{1}{t^{-1}}$). For a large $\mathcal{D}$ and small r, the wave-front extends over a larger region, as demonstrated in Figures 2.3(b) and (e). Conversely, for small $\mathcal{D}$ and large r, we have a thin wavefront, shown in Figures 2.3(c) and (f). For all panels shown in Figure 2.3, the travelling front speed remains approximately constant, reflected through the predicted front velocity:

$$v \geq v_{min} = 2\sqrt{r\mathcal{D}} \tag{2.30}$$

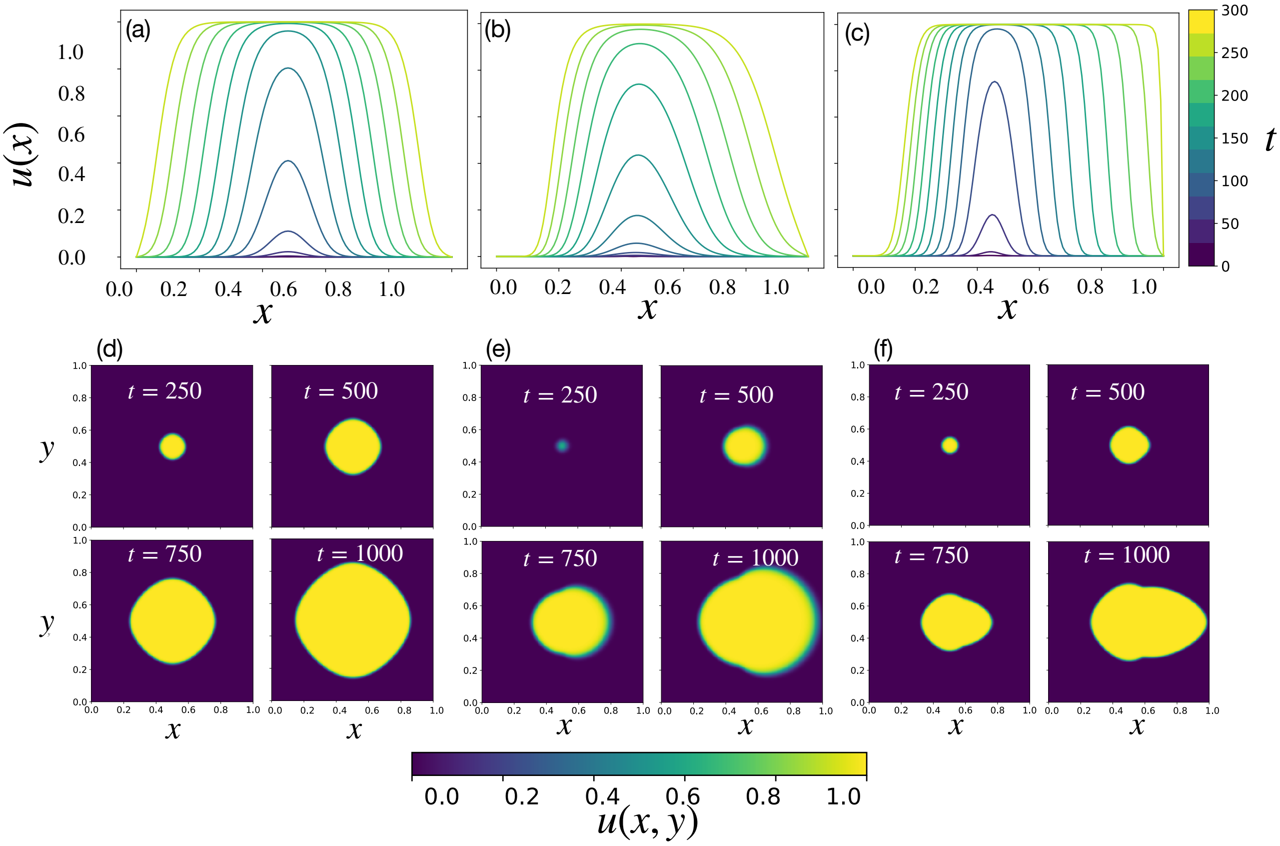


Figure 2.3: Simulating the FKPP model in one and two spatial dimensions with parameters $r = 0.10$, $\mathcal{D} = 0.10$, and $K = 1.00$. In all panels, a non-zero field value is initialised at the domains mid-point (i.e. $u(x = 0.50) = 0.10$) at time $t = 0$. Then, simulations evolve according to a FTCD finite difference scheme to second order observing dirichlet boundary conditions. (a) Symmetric growth and diffusion in 1D. (b) An asymmetric diffusion-gradient increasing from left to right in 1D. (c) An asymmetric growth-gradient increasing from left to right in 1D. (d) Symmetric growth and diffusion in 2D. (e) An asymmetric diffusion-gradient increasing from left to right in 2D. (f) An asymmetric growth-gradient increasing from left to right in 2D.

where v_{min} is the minimum wave-speed admitted by the front[6]. The numerical stability of simulations is ensured provided that the Courant–Friedrichs–Lewy (CFL) condition is met [Courant et al., 1967]. In the case of the FKPP model, the CFL condition is given by: $\mathcal{D} \times (dt/dx^2) < \frac{1}{2}$ where dx and dt represent the discretized domain and time-steps respectively.

The travelling-wave behaviour of Equation 2.29 has been used in various epidemiological applications [Britton et al., 1986, Murray, 2002, Klein et al., 2010, Bianco et al., 2013, Yano, 2017]. Indeed, the FKKP model has close connections to alternative travelling-wave models examined in the context of plant epidemics [Heesterbeek and Zadoks, 1987, Van den Bosch et al., 1988]. In the context of plant disease models, these travelling waves generally emerge from the inclusion of short-range exponentially-bounded[7] dispersal contacts between individuals—as proved theoretically by [Mollison, 1977] who compared the FKPP model against exponentially-bounded 'contact' models.

At first glance, the reaction diffusion (RD) FKPP travelling wave presents a simple, intuitive place to begin modelling the spread of disease through a population of trees. However, following earlier work on turbulent diffusion [Scherm, 1996], epidemic systems that spread through fat-tailed LDD are now thought to exhibit an accelerating 'dispersive' wavefront [Pybus et al., 2012, Cowger et al., 2005], in contrast to the constancy predicted by Equation 2.30. In this case, dispersal becomes evermore efficient as the disease front extends over larger areas.

Inverse power law dispersal kernels are thought to be of significant importance with accelerating disease fronts, following from their fat-tailed leptokurtic nature and subsequent scale-invariance. In particular, they have been successfully used to fit spore dispersal data over five orders of magnitude [Mundt et al., 2009b], from 20m to continental spatial scales. Power law scale invariance was examined by [Severns et al., 2019] for wheat

---

[6] The reader can find comprehensive derivation in [Murray, 2002], in Chapter 11 "Biological Waves". In there, we see that strictly speaking Equation 2.30 is incorrect in two spatial dimensions on account of wavefront curvature, though it still provides a reasonable estimate.

[7] Here, the choice of exponential distribution can be considered as slightly longer-range than a Gaussian dispersal kernel with the same scale parameter. Although, both kernels are still fundamentally thin-tailed.

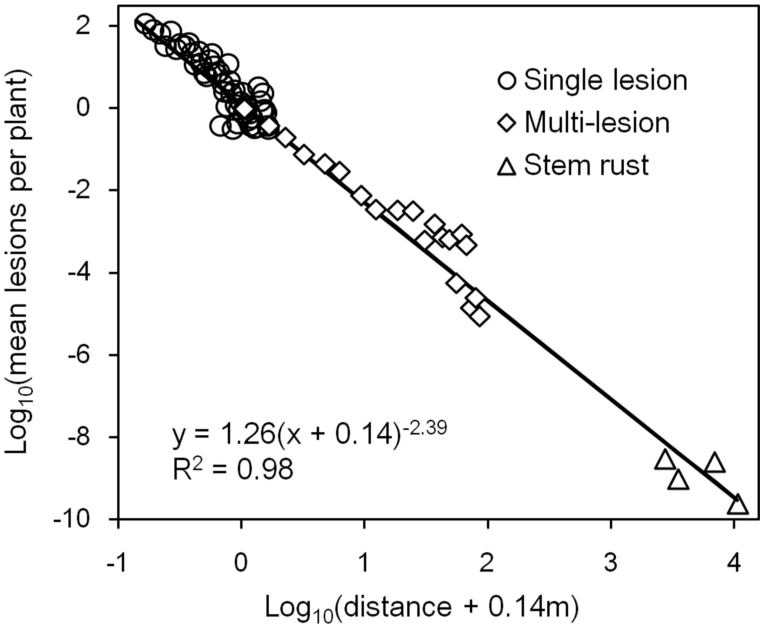


Figure 2.4: Wheat strip rust prevalence with distance shown for three data sources, as displayed in [Severns et al., 2019]. The authors collated data from different wheat strip rust studies and found that disease prevalence over different spatial scales could all be fitted to a common power law—shown in the bottom left corner.

strip rust (WSR) epidemics. In particular, the authors constructed wheat plantations and artificially introduced WSR to measure its prevalence over time. Consequently, the spatio-temporal WSR incidence rates strongly supported an accelerating, disperseive, wavefront.

In addition to dispersive waves, [Severns et al., 2019] aggregated and compared other WSR prevalence studies over different spatial scales and fitted the combined data to one common inverse power law ($y = a(x + c)^{-b}$), shown in Figure 2.4. Generally, we see that WSR induces less lesions with distance, and that infections can arise up to 10km away from the source, illustrated by triangles in the lower right hand corner. An exponent of $b = 2.39$ provided a good fit for all scales, suggesting that inverse *square* power laws may prove a useful rule of thumb to predict disease prevalence over different scales.

2.3 Ash dieback case study

Ash dieback presents an interesting case study of an emerging epidemic currently devastating ash populations throughout Europe [Enderle et al., 2019]. The history and predicted evolution of ash dieback demonstrate how an invasive, non-indigenous pathogen can spread rapidly through a foreign ecosystem that lacks evolutionary defences. Among the many factors driving the spread of ash dieback, long-distance anthropomorphic trade is the mechanism responsible for the initial introduction into Europe from the far-east [Zhao et al., 2013, Queloz et al., 2011].

The pathosystem has been the subject of much research over the years. As a result, the taxonomy, symptoms and life-cycle of the pathogen are now well-known [Gross et al., 2014b]. Understanding the spread of ADB and managing the epidemic impact on ecosystems could only be achieved by the confluence of molecular biologists, forest managers, policymakers and modellers. Although the epidemic is well underway, slowing the spread of ash dieback remains essential to allow ash populations time to adapt.

2.3.1 Historical developments

Reports of dieback on ash began surfacing in Poland in 1992, but a causal agent was not established for a decade [Kowalski, 2001, Coetsee et al., 2000]. Subsequently, [Kowalski, 2006] recognised a novel pathogenic fungus to be the causal agent, identified as an ascomycete anamorph (i.e. an asexual fungus). The fungus was named *Chalara fraxinea*, a member of the hyphomycete genus *Chalara*. The sexual teleomorphic stage of the pathogen was later attributed to *Hymenoscyphus albidus* [Kowalski and Holdenrieder, 2009], a well-known non-pathogenic fungus indigenous to Europe.

Linking the hitherto non-pathogenic *H. albidus* to the agent causing ash dieback perplexed researchers. The enigma was resolved through DNA sequencing by [Queloz et al., 2011] when a second morphologically identical ascomycete named *'Hymenoscyphus pseudoalbidus'* was identified as the pathogen responsible for widespread dieback of European ash in a process referred to as 'Cryptic speciation'.

Interestingly, the emergent epidemic caused by *H. pseudoalbidus* coincided with developments in the phylogenetic classification system of the kingdom Fungi [Hibbett et al., 2007], and dual nomenclature[8] [Wingfield et al., 2012]. Subsequently, the pathogen *'Hymenoscyphus pseudoalbidus'* was renamed to *Hymenoscyphus fraxineus* (HF).

2.3.2 Symptoms and epidemiology

European ash is highly susceptible to HF because it has no (co-evolved) evolutionary defence. Although HF is lethal to European ash, it poses little threat to its native Asian hosts *Fraxinus mandshurica* and *Fraxinus chinensis.* Once the fungus colonises a European ash leaf, it can spread through twigs, branches, the xylem, and eventually the whole tree. The symptoms include necrotic lesions, crown dieback, wilting and eventual death. In addition to leaf-infections, the pathogen can colonise the root-system [Schumacher, 2011]. Root-infections usually occur in already severely infected ash [Gross et al., 2014b]. After which, it is only a matter of time before opportunistic fungi invade and significantly accelerate mortality [Enderle et al., 2013].

The progressive symptoms of ADB, as presented by [Gross et al., 2014a], are displayed in Figure 2.5. Ascospores initially infect susceptible ash leaves (a), becoming visible after around two weeks [Cleary et al., 2013] in the summertime. In Figure 2.5, panels (b-d) show the initial infection spreading through the leaf into the rachis and the development of the first necrotic lesions—see [Mansfield et al., 2018] for further information on the precise mechanism of ascospore leaf penetration.

Over winter, the infection continues to spread through ash. Young ash develop large visible necrotic lesions, as illustrated in Figure 2.5(e-i). In spring, the infection causes shoot wilting (g) and death (h-i) before causing xylem necrosis (j). Over many seasons, large infected mature ash trees begin to die (l) as it begins forming epicormic branches, as noted by [Marciulyniene et al., 2017] and losing its canopy.

The pathogen HF is lethal to European ash of all ages. Nevertheless, research has established that small young ash trees are more at risk, and susceptibility declines with

---

[8] Originally, fungi were classified through the structure of their sexual organs. Problematically, ascomycete fungi have a complicated dual reproductive mode (both sexual and asexual) that often caused confused. However, a move toward a one-name fungi classification system has since simplified fungi taxonomy.

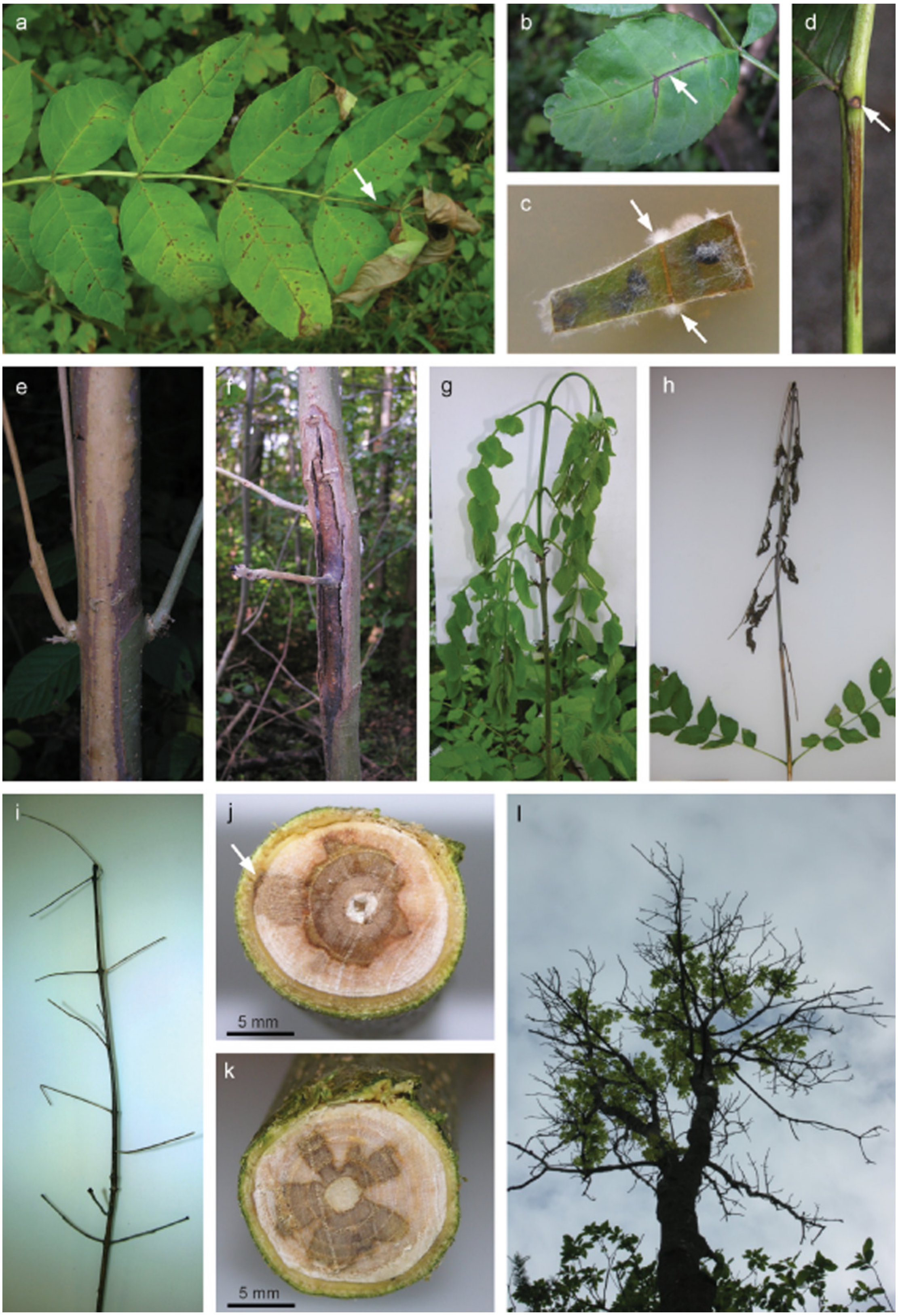


Figure 2.5: The symptoms of ash dieback, taken from the work of [Gross et al., 2014a]. The pathogen *H. fraxineus* infects the leaves of ash, leading to early onset wilting and desiccation. The fungus then reproduces asexually, spreading through twigs, branches and eventually the xylem. Symptoms include wilting, necrotic lesions, crown dieback and eventual mortality.

maturity. Surveys of ash stands conducted by [Marçais et al., 2017] in Belgium recorded that after six years of infection, small young saplings died with 35% mortality, whilst slightly larger ash (< 25cm in diameter) displayed mortality of 11%. In contrast only 3.2% of large mature ash (> 25cm in diameter) died.

Various sources of ash mortality data have been collected in different European countries. In Germany, a forest stand of planted ash trees showed a 73% mortality rate after five years [Langer et al., 2015] (as cited in a review [Enderle et al., 2017]), while observations of ADB progression in Austria suggest a low mortality rate of 5% measured over a two-year window [Keßler et al., 2012]. A study conducted at different sites throughout Great Britain suggests a time scale ranging between $3 - 15$ years of infected tree growth before death [Wylder et al., 2018].

In addition to age, ash survival also depends on the landscape. Landscape features and ADB progression were studied by [Grosdidier et al., 2020] over a sample plot of size 3.5km $\times$ 6.5km in France, where observations over two years indicated that the surrounding landscape has little impact initially in 2012. However, after pathogen establishment, later surveys in 2016-2018 showed that landscape features play an essential role. Among the results put forward by [Grosdidier et al., 2020], a highly abundant ash region increased the prevalence of collar canker and rachis symptoms in neighbouring ash. In addition, the authors found that the influence of ADB decayed exponentially up to $200 - 300$m away from the high density source, thus suggesting a density-dependency in ADB spread.

Modelling work suggests a myriad of environmental factors can also predict the vulnerability of ash and subsequent spread of disease [Dal Maso and Montecchio, 2014]. Spatial regression analysis conducted by [Chumanová et al., 2019] in the Czech Republic indicates that altitude is an important predictor of pathogen growth, which also support the strong negative temperature dependence observed by [Hauptman et al., 2013].

2.3.3 Life cycle and reproductive mode

The reproductive mode is intricate, and HF can infect hosts through the soil, water, and air [Gross et al., 2012], though the primary natural driver of disease propagation is through wind-dispersal during summertime sporulation. Sporulation typically occurs from June-September when fungal fruiting bodies on the previous litter-fall release ascospores [Grosdidier et al., 2018, Hietala et al., 2013]. Multiple ascospores sources can infect the same leaf [Gross et al., 2012]. Though the primary natural driver is wind, infection (and re-infection) of ash are also thought to be possible through the soil-borne mechanisms [Fones et al., 2016], albeit with low frequency.

Ash dieback is highly seasonal [Bengtsson et al., 2014] and follows a complex, yearly polycyclic infection cycle. Infected ash hosts will shed their leaves in the autumn, proceeded by fungal fruiting bodies growing on the dead leaf litter until summertime. In summer, fruiting body spores are wind-dispersed and continue the cycle by producing new secondary infections—together, the life cycle and symptom expressions are illustrated in Figure 2.5. It is interesting to note the cyclic similarities between yearly ADB infection/re-infection and the seasonal infections due to crop rotations, e.g. [Tankam-Chedjou et al., 2020]. Notwithstanding that infected crop removal usually coincides with harvest time instead of infected ash survival that can span years.

The life cycle of the fungus HF can be understood to have two well-differentiated reproductive modes, sexual and asexual—a common trait of phyla Ascomycota, or ascomycetes fungi [Hawker, 2016]. Initially, asexual spores (conidia) were hypothesised to only increase genetic variance and act as spermatia [Gross et al., 2014a]; however [Fones et al., 2016] called this into question, suggesting instead that asexual reproduction of the pathogen may play a role in driving the pathogen spread. Despite the potentially significant claim put forward by [Fones et al., 2016], it has gained seemingly little traction, and the role of asexual reproduction is still not fully understood.

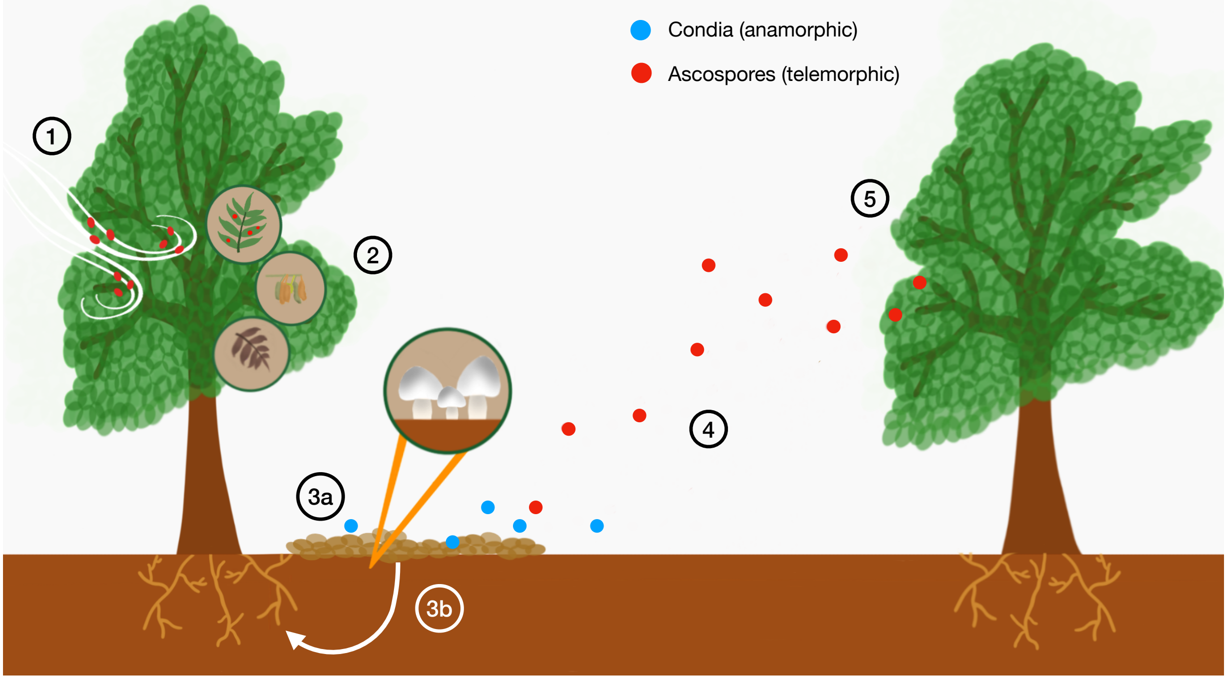


Figure 2.6: A life-cycle illustration of ADB: (1) HF ascospores disperse through wind during summertime sporulation, generally between June-September. Ascospore production constitutes HF's sexual (or 'teleomorphic') reproduction mechanism. (2) Ascospores penetrate the leaves of susceptible ash, causing the leaves to wilt. After spores infect leaves, the fungus proceeds to spread through twigs, branches and eventually the xylem. Infected leaves are shed in autumn, or from disease induced death. (3a) Over winter, a mushroom-like fruiting body grows on infected leaf-fall (usually the petiole). (3b) A proposed soil-borne infection mechanism has been proposed [Fones et al., 2016]. Here, asexually reproducing HF mycelium are thought to infect the roots of ash trees. Through both steps (3a) and (3b), asexual condia disperse from the infected litterfall, shown in blue. Condia are proposed to act as spermatia, increasing the genetic diversity of HF. (4) During the next summer period, immense numbers of ascospores are release from the fruiting body and disperse through wind. (5) Ascospore dispersal induces ADB infections in distant susceptible ash. Fungal spores in particular are known to travel large distances.

2.3.4 Dispersal

In general, fungal spores have an efficient multi-scale wind-dispersal mechanism known to travel large distances [Golan and Pringle, 2017, Wingen et al., 2013, Mundt et al., 2009a], often described by power-law kernels, e.g. [Shaw et al., 2006]. However, from an epidemiological perspective, data on spore dispersal does not necessarily reflect the dynamics of new infections, because we cannot guarantee the availability of susceptible host material. More explicitly stated, invasive spore colonisation is not guaranteed even if hosts are available to infect. Still, studies on spore dispersal shed essential light on the spatial scale of ADB dispersal.

Modern methods typically rely on 'spore trapping' and real-time polymerase chain reaction (PCR) to study spores dispersal. Data collected by [Chandelier et al., 2014] over three years using a (rotating arm) trapping system and PRC amplification. The authors reported a 10% spore trapping efficiency, and that most dispersed ascospores remained within 50m from the infectious source, with only a small number of spores exceeding distances beyond 50m. The work of [Chandelier et al., 2014] demonstrated the utility of novel PCR methods to spore trap, but undesirably collected dispersal data over relatively small spatial scales.

Among the first landscape-scale fungal spore studies were conducted by [Rieux et al., 2014]. The authors focused on a comparable ascomycete fungus, *Mycosphaerella fijiensis*, affecting banana plants. Interestingly, [Rieux et al., 2014] reported that asexual spore dispersal gradients extended a small distance 15m. As opposed to occasional, rare LDD in sexual (ascospore) dispersal up to 1000m. Contrasting sexual and asexual spore dispersal was novel, and observing a small localised asexual dispersal gradient supports the accepted idea that sexual dispersal in ADB is the dominant driver of disease spread.

Arguably the most comprehensive multi-scale study of ADB spore dispersal was performed by [Grosdidier et al., 2018]. In their paper, [Grosdidier et al., 2018] tracked the local and landscape-level dispersal of ascospores produced by *H. fraxineus* in France. The data collected relied on spore trapping and PRC, where the reported trapping efficiency was 30–47%, thus marking an improvement over previous studies e.g [Chandelier et al., 2014]. Most spores dispersal remained localised up to 50m away from the inoculum source, consistent with the results mentioned above put forward by [Chandelier et al., 2014]. Nevertheless [Grosdidier et al., 2018] set spore traps over much larger spatial scales ($\sim$ 100km), subsequently detecting ADB spores $50 - 100$km ahead of the disease front.

Two dispersal kernels were used by [Grosdidier et al., 2018], a thin tailed Gaussian and an inverse power law of the forms:

$$D(a,r) = \frac{1}{\pi a^2} \exp\big[-\frac{r^2}{a^2}\big] \tag{2.31}$$

and

$$D(a,r) = \frac{(b-1)(b-2)}{2\pi a^2}\big[1+\frac{r}{a}\big]^{-b} \tag{2.32}$$

where a and b are fitted parameters (for the Gaussian kernel, $a = \sqrt{2}\sigma$ with σ being the standard deviation.). In Equation 2.31, the fitted value was $a = 196$m, while the fitted values in Equation 2.32 were $a = 203$m and $b = 3.3$ respectively.

Equation 2.32 falls into the classical two-parameter geometric family of dispersal distributions. The scale parameter is described by a and the shape parameter by b. The mean dispersal distance described by Equation 2.32 is $\frac{2a}{b-3}$, and parameters a and b are valid for $a > 0$ and $b > 2$.

The functional form of Equation 2.32 is predicated on pollen dispersal studies, as reviewed by [Nathan et al., 2012]. The tail of Equation 2.32 is particularly well suited to describe LDD events, as noted by [Austerlitz et al., 2004] when describing the dispersal of pollen particulates. Moreover, a study by [Devaux et al., 2007] used Equation 2.32 to model pollen-dispersal (and thus plant gene-flow) over landscape-level spatial scales. Presumably, the ability of Equation 2.32 to describe LDD, and the size similarity between pollen and fungi spores, motivated [Grosdidier et al., 2018] to include it their field study.

2.3.5 Management and control

Losing abundant ash populations could have dire consequences for several ecosystem functions, including nutrient recycling, food webs, and biodiversity. As such, ADB presents a conservation challenge throughout Europe and Great Britain [Pautasso et al., 2013]. Moreover, under the rapid proliferation of ADB, various ash-dependent species risk extinction; [Hultberg et al., 2020] identified 115 at-risk (lichens, fungi, invertebrates, bryophytes/moss) species in Sweden that rely on ash. Alarmingly, many of the ash-associated species identified by [Hultberg et al., 2020] also depend on elm species, which in turn face the fungal pathogen Dutch elm disease [Brasier, 1991].

Given the ecological importance of ash in numerous temperate European forest types (e.g. floodplain, ravine, and lowland [Dobrowolska et al., 2011b]), management and pathogen control remain essential. The control of ash dieback in a well-established focus of infestation, in natural and artificial environments, is virtually impossible [Havrdova et al., 2017], and it is already well recognised that ADB will eventually wipe out the vast majority of ash in Great Britain [Hill et al., 2019].

After the first reports of ash dieback in the UK in 2012, the UK government put forward the 'Chalara Management Plan'[Defra, 2013] in 2013[9] . The report details actions taken by the government to manage the spread of ADB. In the short term, the government primarily sought to slow the spread between affected regions, whereas, in the long term, the strategy consists of replanting genetically tolerant ash trees. In addition, the government set out plans to engage landowners and industries to take part in control initiatives voluntarily and minimise the effect on the timber trade and supply lines. Finally, the report details rapid-response modelling work undertaken by Cambridge University (unpublished) to help understand which areas in GB are likely to be the most affected. Specifically, output from the model indicates which low-risk and high-value areas should be monitored preferentially. Unfortunately, however, the model developed by Cambridge

---

[9] The plan can be found at: https://www.gov.uk/government/publications/chalara-management-plan. Note, the pathogen's old name 'Chalara' ash dieback has since been replaced with *Hymenoscyphus Fraxineus*.

University cannot be found in the literature.

Controlling the spread of ash dieback reflects the spatial and temporal scale over which it spreads. In particular, ADB fungal spores are thought to be able to jump between patches of ash, even in the absence of susceptible hosts [Wingen et al., 2013]. However, LDD accounts for only a small minority of spore dispersal. Furthermore, despite many notions of LDD, no unified LDD classification system exist, which led [Golan and Pringle, 2017] to propose a definition based on the distance traversed by the top 1% of spores.

The long-term survival of ash depends on a small proportion of genetically resistant ash trees. Despite many unpublished reports, genetic tolerance studies only began surfacing around a decade after the widespread outbreak [Kjær et al., 2012, Stener, 2013, McKinney et al., 2014]. In particular, [Muñoz et al., 2016] showed the heritability of crown dieback and collar-lesion symptom expression. In the three French provenances sampled, [Muñoz et al., 2016] found no evidence for regional tolerance. Presently, genetic tolerance is widely accepted [Havrdová et al., 2016, Skovsgaard et al., 2017], and more recent work has focused on metabolomic tolerance classification, profiling which metabolite markers correlate with resistance [Nemesio-Gorriz et al., 2020a, Nemesio-Gorriz et al., 2020b, Sidda et al., 2020, Chaudhary, 2020].

A silvicultural system proposed by [Skovsgaard et al., 2017] involves visually scoring severity based on crown dieback and collar lesions. In this scheme, forest managers would inspect disease severity to record disease progression over time and record genetically resistant trees to cultivate for future timber production. In addition, [Skovsgaard et al., 2017] proposed that diseased forest stands should not be indiscriminately felled on account of high-value tolerant individuals; this suggestion stands in contrast to the idea of a ‘cull radius’, as alluded to by [Cunniffe et al., 2015b]. In general, felling infected stands is ill-advised [Chandelier et al., 2017], except when infected hosts present a risk of uncontrolled and damaging tree-fall—as explained by [Hill et al., 2019] when assessing the clean-up cost within Great Britain.

Following the establishment of genetic tolerance, numerous long-term preservation strategies rest on cultivating and replanting trees that exhibit low damage levels. Breeding programs in many European countries are currently underway, as reviewed by [Plumb et al., 2020]. In the UK, the Living Tree Project (https://livingashproject.org.uk) aims to collate tolerant UK ash for future breeding.

Chapter 3

Tree disease: a simple lattice model

Typically, models of tree disease are complicated and involve multiple parameters informed by experimental data. Moreover, modelling a specific pathosystem requires in-depth, specialist knowledge to incorporate biological realism, such as pathogen lifecycles or environmental suitability. This introductory Chapter outlines a simple model of tree disease spreading through a forest that will lay the foundations for more detailed treatment in later Chapters. Consequently, the compartmentalised SIR, percolation-based, model used by [Orozco-Fuentes et al., 2019] will be re-examined.

From first principles, the percolation model is analysed over a single tree density parameter (ρ), leading to a mathematical and conceptual definition of percolation in the context of tree-based epidemics. Then, a discussion of critical phenomena, universal behaviour, and self-similarity in the system will follow. Crucially, this Chapter demonstrates the importance of thresholds in the spread of tree-based diseases.

The one-parameter model will give way to a more involved two-parameter stochastic percolation. The two-parameter model incorporates an additional infectivity parameter (β) representing pathogen virulence. In general, measuring infectivity is challenging and subject to much spatial and temporal variation due to changing climatic/environment conditions or species-level genetic variations in susceptibilities. Accordingly, including infectivity is vital to construct representative models of tree-based epidemics. Thus, the two-parameter model will constitute a simple lattice model (SLM) of disease dynamics. Additionally, this Chapter will introduce ensemble-averaging parameter sweeps used throughout the thesis to categorise the SLM behaviour.

3.1 Percolation formalism

Consider a square lattice of size $L \times L$, where each site in the lattice can be in one of two states: open with probability ρ or closed with probability $(1 - \rho)$. Therefore, the probability ρ describes a density parameter and encapsulates the occupancy of a homogeneous distribution of open and closed lattice positions. One open site (c_p) is connected to another (c_q) if laid within the Von Neumann neighbourhood [Toffoli and Margolus, 1987].

A connected set of open nearest neighbours define a cluster denoted by C, where $c_i \in C$. Given the set C, it is possible to traverse between member sites $c_i \in C$ by moving through the lattice in either horizontal or vertical steps 'Von Neumann motion'. Given two distinct non-overlapping clusters $c_i \in C_1$ and $c_j \in C_2$, then $c_i \neq c_j$ i.e. there is no way to jump from C_1 to C_2 following Von Neumann motion. Connectivity is defined between lattice sites rather than the edges which connect them, known as 'site' percolation—as opposed to 'bond' percolation.

If ρ is close to zero, only small clusters would be present as a disordered system; conversely, if ρ is large, a connected network of open positions would dominate the domain, thus defining an ordered system. At some point for $\rho \in [0, 1]$, a critical threshold (ρ_c) would be reached and a singularity of connected sites would span an infinite sized lattice. On a finite lattice, the cluster is said to percolate and form a 'spanning cluster' (C_∞) that extends between at least two different edges of the lattice. The formation of the spanning cluster occurs abruptly between a very narrow range of ρ values. Therefore, the threshold for percolation defines a critical-point [Stauffer, 1979].

The critical point can be defined as the least value of ρ where percolation occurs with non-zero probability. This can be formalised by first considering the probability function: $\theta(\rho) = \{\rho : |C| = \infty\}$ where $\theta(\rho)$ is the probability of an arbitrary site, within a lattice of density ρ, belonging to cluster C of size ∞ (i.e. the spanning cluster). The critical value then satisfies:

$$\rho_c = sup\{\rho : \theta(\rho) = 0\} \tag{3.1}$$

At this point, we have described the spanning cluster conceptually, though nuances and technicalities still complicate the definition. For example, the percolation probability depends strongly on the lattice size. If the mean cluster size (defined by a 'characteristic' cluster length C_r) is more significant or comparable to L, percolation can occur despite the density residing in the sub-critical regime. Thus, a sufficiently large lattice is required to approximate the threshold ρ_c reliably and the spanning cluster C_∞.

When the domain size is large, the threshold ρ_c will remains approximately constant and insensitive to small changes in the domain. All proceeding simulations in this Chapter are conducted on finite-sized domains between two and three orders of magnitude larger than individual lattice points. The model presented in this Chapter (first published by [Orozco-Fuentes et al., 2019]) was found to agree with the accepted percolation threshold, when the lattice had size $L \sim 500 \times 500$; the lattice configuration used here will therefore assume the same size configuration.

Intuitively, it is clear to see the link between percolation and epidemiology: open lattice positions act as susceptible members of a population (S), and ρ defines a density of the hosts. In this paradigm, the spanning cluster describes a high-impact epidemic that spreads uncontrollably. Small to medium-sized clusters existing at (or slightly below) the threshold describe short-lived, failed epidemics where the pathogen spreads for a time before becoming extinct.

Percolation models have clear limitations when describing mobile hosts, but fortunately, the spread of disease through sessile tree populations is more appropriate. Thus, percolation theory provides a valuable, yet simplistic, framework for modelling tree diseases, mainly when the effect of long-distance pathogen dispersal is limited or reduced. Examples include soil-borne pathogens, i.e. *Rhizoctonia solani* and *Pythium spp* [Otten et al., 2004, Poggi et al., 2013]. In addition, we can envision that percolation becomes a relevant approximation to dispersal in unfavourable abiotic conditions. For example, areas of low wind/rainfall splash reducing the dispersal of citrus canker (caused by the bacterium *Xanthomonas axonopodis*) [Bock et al., 2010], or regions of lower wind speeds reducing the fungal spore dispersal of ash dieback [Solheim and Hietala, 2017]. Moreover, a percolation setting becomes more suitable as the landscape becomes more densely populated with trees, as argued by [Orozco-Fuentes et al., 2019]. The reader can find a more in-depth review on epidemic percolation models in section 2.1.5.

3.2 Percolation-based SIR

Forming a percolation-based lattice model of tree disease requires us first to combine a compartmental SIR-like model within the lattice (L) mentioned above and an appropriate transmission dynamic to model the spread of disease between lattice points. Firstly, the percolation density parameter (ρ) defines a simple host distribution, whereby ρ represents the probability of a susceptible tree S (given a numerical value 1). In contrast, empty lattice positions define an insusceptible state $\emptyset$ (with numerical value 0).

In the model, a susceptible tree will transition into the I compartment (having a numerical value of 2) if it neighbours an infected tree. Transitions between states occur with a probability of 1, and the Von-Neumann neighbourhood describes the connectivity between trees. After an arbitrary number of time-steps T, an infected tree will transition into the removed state R and die; see appendix A.1 for more information on the numerical implementation. For simplicity, the infectious lifetime will not be considered as a parameter but will remain fixed at $T = 1.0$—revisited below in section 3.3.

The initial conditions begin with a small patch of infected hosts (of size 5×5 for example) located at the origin (L_O) at $t = 0$, and percolation events describe the boundary conditions. Percolation occurs when the infection spreads from L_O to any of the four lattice boundaries, thereby defining a connected cluster of infectious-removed trees that span the domain. Each time-step through the simulation represents an arbitrary unit of time, and simulation termination occurs on one of three boundary conditions: (1) percolation is observed (2) the pathogen becomes extinct (3) the time-horizon of N steps is reached. Typical simulations on a domain of size 500×500 are shown in Figure 3.1 for three successive time-steps.

Figure 3.1 illustrates dynamic simulations of disease spread through a series of 500×500 domain. At density $\rho = 0.70$, Figures 3.1 (a-c) reveal a diamond shaped pattern of spread. The unnatural spread pattern in Figures 3.1 (a-c) reflect artefacts of the square lattice geometry, and therefore is unlikely to be realised in nature. However, the diamond-like spread begins to disappear in Figure 3.1(d-f), when simulations have density $\rho = 0.65$. At this density, a wave-like propagation of infected trees spreads radially from the epicentre outward toward the lattice boundary. Interestingly, we may understand these observations through the lens of stochasticity. A low tree density reduces the chance of spread between nearest neighbours, which produces a more noisy and disordered epidemic spread. In turn, the lower infection probability between neighbours disrupts the highly ordered wavefront (panels (a-c)) into a circular travelling wavefront (panels (d-f)). Nevertheless, it is also worth remarking that an alternate lattice geometry (e.g. triangular or honeycomb) would alter the diamond-like pattern altogether.

Figures 3.1(g-i), show the spread of disease for a lower host density of $\rho = 0.60$. As the disease spreads outward, a more fractal-like pattern begins to emerge. Moreover, the disease spreads slowly, as evidenced by the smaller area traced by the infectious-removed trees shown from white to red. In this regime of spread, ‘persisting’ simulations result from the slowly evolving epidemic as it defines a disordered cluster of infectious-removed hosts.

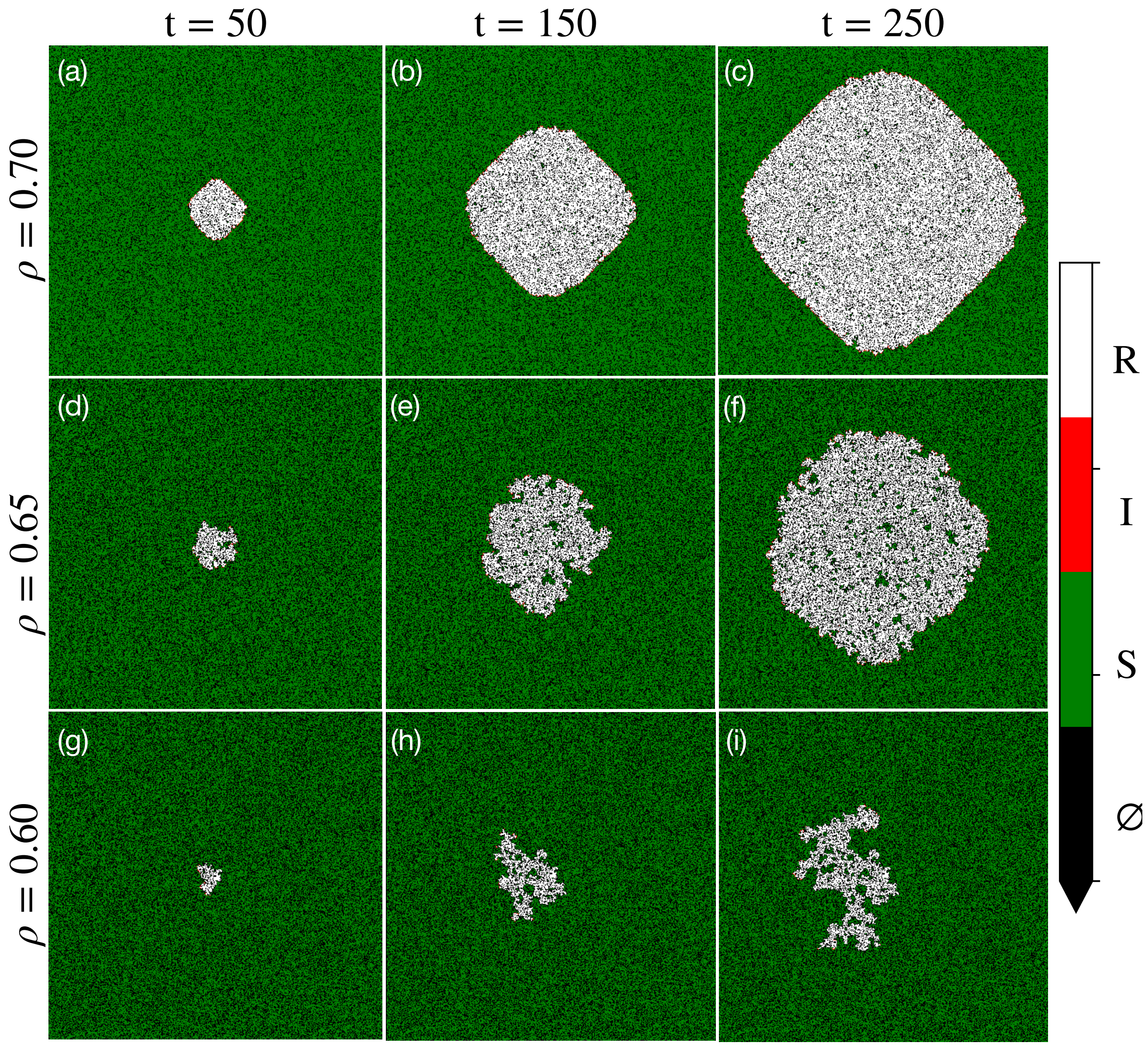


Figure 3.1: Evolving outbreaks in the one-parameter percolation-based SIR model are shown for different tree densities (ρ). From left to right, three successive time-steps are plotted on a domains of size 500 × 500. Green and black pixels represent susceptible and insusceptible (empty) host units, respectively, while white and red lattice sites depict removed and infectious hosts. (a-c) High-density simulations reveal an unnatural diamond-shaped spread, undesirably reflecting the underlying lattice geometry. (d-f) Simulations above the threshold spread radially outward from the epicentre, defining a connected cluster of infectious-removed trees in the process. (g-i) Around the percolation threshold, the disease spreads slowly and chaotically outward.

3.2.1 Percolation threshold

Consider a simulation with tree density below the epidemic threshold. In this case, an evolving epidemic is unlikely to percolate to the domain boundary, and the pathogen will become extinct. If the host density increases, we can imagine that epidemics begin percolating to the boundary for some specific value. Consequently, the probability of percolation was examined over a sweep of density parameters, shown in Figure 3.2(a). Unsurprisingly, a threshold-like behaviour is revealed. All the simulations that form Figure 3.2(a) evolved inside a 500×500 sized domain. The probability of percolation ($Pr(\rho)$) defines a critical region, highlighted in orange (where $\rho \in [0.57, 0.62]$), which separates regimes of pathogen extinction and percolation/epidemic. The threshold depicted by Figure 3.2(a) is consistent with the accepted percolation threshold for a two-dimensional square $\rho_c \approx 0.592$.

The value of $Pr(\rho)$ depends non-trivially on stochasticity, which motivated an ensemble-averaged approach—further reading on the underlying theory of ensemble-averaging can be found in [Gibbs, 1902]. In Figure 3.2(a), 100 repeated simulations obtain the probability of percolation. Simulations that percolate to the domain boundary assume the numerical value of one, while pathogen extinction events assume zero; for each value of ρ, the average value is computed, thus defining a probability $Pr(\rho)$.

At the critical density, denoted by ρ_c, we witness the emergence of some exciting phenomena. Figures 3.2(c-d) show a spanning cluster (C_∞) of infectious and removed trees (in white to red respectively) at $\rho_c = 0.592$. The cluster looks remarkably similar at all spatial resolutions, said to be 'self-similar' [Kapitulnik et al., 1983]. Within C_∞, one can identify clusters of untouched susceptible trees (in green) of various sizes, suggesting a distribution of cluster sizes occupy all possible length scales. In the literature, clusters can be described by a 'cluster number' (n_s) distribution, where n_s is the number of clusters containing s open/susceptible lattice sites. Furthermore, around the percolation threshold, there can be significant fluctuations in the size of the clusters formed[1].

---

[1] The related statistical fluctuations analysed by [Orozco-Fuentes et al., 2019] present an effective early warning system for the prediction of forest-based pathosystems—revisited in the next Chapter.

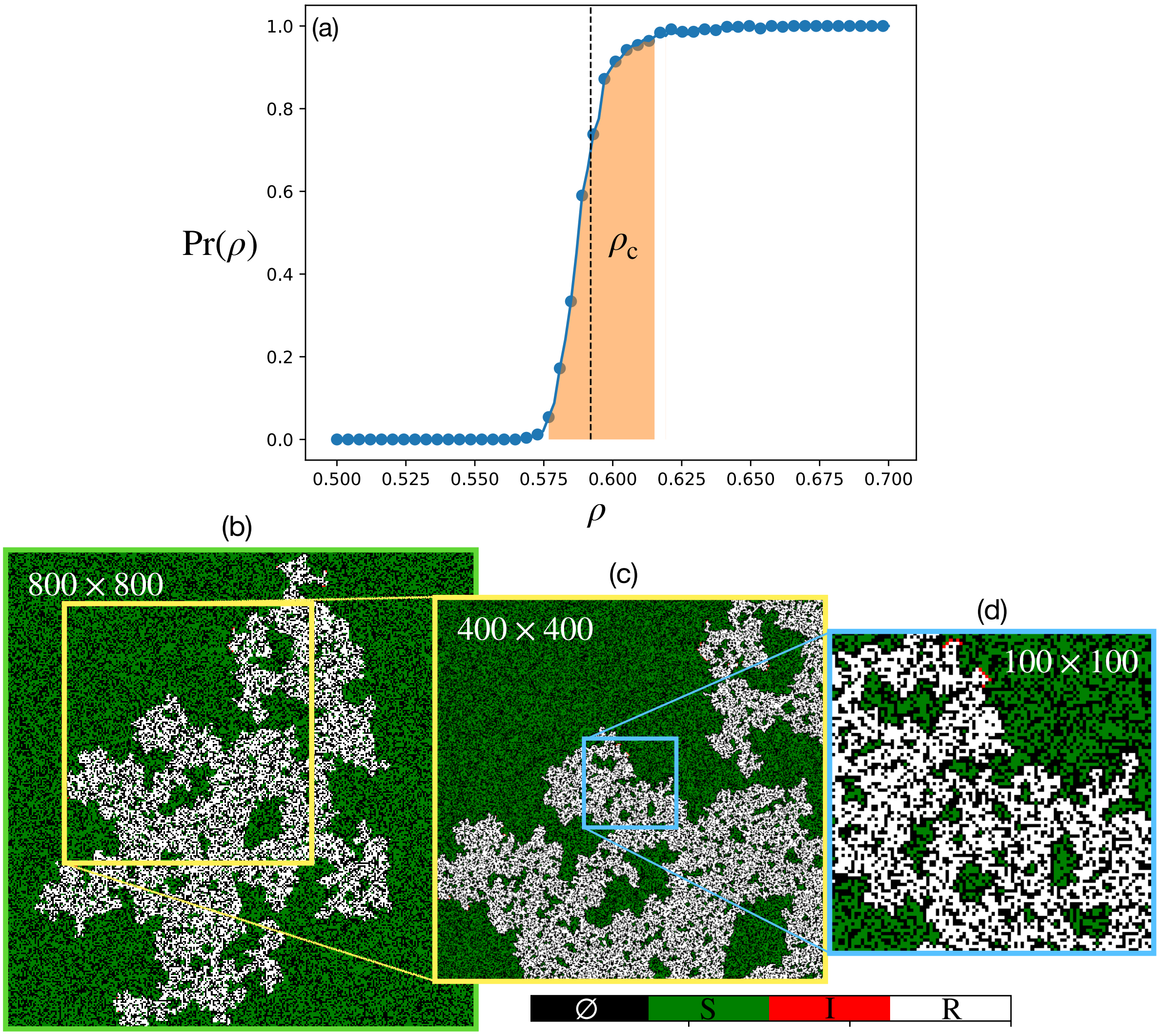


Figure 3.2: The percolation threshold is determined for the one-parameter percolation SIR model. (a) The probability of percolation (Pr(ρ)) is plotted against host density. The shaped orange region highlights a threshold consistent with results from classical percolation theory, namely $\rho_c = 0.592$, shown by the vertical dashed line. (b-c) At the critical density ρ_c, a cluster spanning the domain in (b) is assessed over progressively smaller resolutions. Similar features are observed on different scales and scale invariance is observed in the model.

Lattice	NN	**Site percolation**	**Bond percolation**
1D	2	1	1
Square 2D	4	0.593	1/2
Triangular 2D	6	1/2	0.347
Honeycomb	3	0.696	0.653
Diamond	4	0.43	0.388
Voronoi	-	0.714	0.667

Table 3.1: The site and bond percolation threshold for various lattice types—data published by [Stauffer and Aharony, 2018, Becker and Ziff, 2009]. Each lattice configuration defines a set of nearest neighbours (NN).

This Chapter rests on a simulations developed on a square lattice, though we could have considered different configurations, e.g. a triangular, honeycomb or Voronoi lattice. From the observation of Figures 3.1(a-c), high host densities would produce highly ordered wavefronts reflecting the different lattice configurations. In addition, various quantities within the model would change, especially the critical density ρ_c that changes in response to the nearest neighbour (NN) number. For completeness, Table 3.1 shows a selection of site and bond percolation thresholds. Even though ρ_c would change between lattice configurations, some universal properties of the model would remain fixed, which leads to a description of universality below.

3.2.2 Universality

At $\rho \sim \rho_c$, percolation and scaling theory explain how the system follows a power law of the form $\propto (\rho - \rho_c)^\alpha$ where α is a critical exponent that is universal across all lattice types—see [Stauffer and Aharony, 2018] page 31. More broadly, a universal critical exponent describes the behaviour of a continuous phase transition and only depends on the general properties of the system (e.g. the physical dimension or the range of interaction). All systems that possess the same exponent are members of the same 'universality class' [Fisher, 1969, Ódor, 2004].

Another important notion within the model pertains to the ‘ ’*correlation function*’. More specifically, a correlation function is defined as a function that describes the statistical correlation between random variables [Van Leeuwen et al., 1959]. In the context of percolation, the correlation functions typically describe the probability that two open lattice sites, separated by distance r, are connected within the same cluster [Stauffer and Aharony, 2018]. For example, consider the probability that one open site (at the origin) is connected to another open site a distance r away. In this scenario, the probability is described by the correlation function $g(r)$. The behaviour of this function defines a length scale, denoted ξ, that dictates how the probability of ‘connectedness’ decays with distance r. For densities close to the percolation threshold:

$$\xi \sim |\rho - \rho_c|^{-\nu} \tag{3.2}$$

where ν is the critical exponent that is universal for all lattice configurations and only depends on the dimension of the lattice used. In general, there are critical exponents for other quantities, e.g. cluster sizes—discussed more below. However, all follow similar power laws, as shown by [Stauffer and Aharony, 2018, Stauffer, 1979].

Equation 3.2 can be understood by exploring how the connectivity of open sites depend on the density and the divergence that occurs at the threshold $\rho = \rho_c$. For low densities, ξ is small because all clusters exist in singlets/triplets. However, as ρ increases, the mean cluster length increases as more sites become open and connect to form larger clusters. As we approach the critical density (from the direction $\rho \to \rho_c^-$) the spanning cluster is formed and ξ diverges towards infinity, $\xi \to \infty$. If one neglects the divergent spanning cluster, a similar picture is painted for densities just above criticality $\rho > \rho_c$. That is, the correlation length ξ decays rapidly as $|\rho - \rho_c|$ increases. This time however, mid-to-large sized clusters get absorbed by C_∞ as the density increases; thus leaving only small untouched clusters, as $\rho \to 1$ and $\xi \to 0$. See [Stauffer, 1979] for a detailed breakdown of power laws and correlation length.

Lastly, it is worth discussing well-known results on how the cluster sizes (or masses) scale with the lattice dimension. If $\rho > \rho_c$, the largest cluster present (denoted by M) would scale according to $M \propto L^2$. In contrast, if $\rho = \rho_c$, M would follow $M \propto L^{1.9}$, where $d_F = 1.9$ describes the cluster fractal dimension. If we normalise the cluster by the size of the lattice (L^2), the mean density of M will decay as the lattice size increases, i.e. $L^{1.9}/L^2 = L^{-0.10}$. Broadly, the critical phenomena found in percolation theory, thermodynamics and magnetism have close ties and are described by similar power laws underpinned by scaling theory [Essam, 1980].

3.3 Pathogen infectivity

We have established a percolation-based model of tree disease described by a one-dimensional parameter-space over tree density ρ. We now extend the parameter-space to include an 'infectivity' parameter, denoted by β. Previously, we made an implicit assumption about pathogen transmission. Namely, that infected trees will transmit the infection to susceptible nearest neighbours with perfect fidelity, that is, a probability of 1. In reality, a pathogen may display a range of virulence depending on the environmental suitability or host susceptibility. For example, the fungus *H. fraxineus* causes more severe ash dieback in natural forest ecosystems [Marciulyniene et al., 2017] and releases more spores conditional on temperature [Chandelier et al., 2014].

Here, the parameter β is introduced to model infectivity. Now the probability of a susceptible tree becoming infected during a single time-step is given by: $Pr(S \rightarrow I) = \beta$. Appendix A.1 contains more descriptive information on the computational implementation. The infectivity parameter describes a transmission 'rate' (i.e. per time-step) and is closely linked to the infectious lifetime T of the tree. If a susceptible host falls within a von Neuman neighbourhood of an infected host, it will remain susceptible with probability:

$$Pr(S \rightarrow S) = \rho(1 - \beta)^T \tag{3.3}$$

where T is the number of an infectious lifetime. As T increases, the likelihood of a tree remaining susceptible decreases. Equation (3.3) sets the scene for a predictive mean-field theory. Subsequently, appendix A.2 outlines steps toward a novel continuum model of tree disease.

Previously in the one-parameter SIR variant, the infectious lifetime played a minor role in determining wavefront because pathogen transmission occurs with a probability of one. However, a susceptible host will now survive for $t = T$ time steps before transitioning into R. Although the model is ultimately non-dimensionalised with arbitrary units of time T, we envision a single time step on the order of years. Thus, at this stage, we have recovered the two-parameter model used by [Orozco-Fuentes et al., 2019], henceforth referred to as the '*simple lattice model*' (SLM).

Figure 3.3 shows three SLM simulations, spreading for different infectivity parameters at different time steps. All simulations in Figure 3.3 are governed by a fixed value of $T = 10$, and density $\rho = 0.70$. The colour bar in Figure 3.3 represents different steps through the infectious period, from yellow to red. Higher values of β yield a faster spreading velocity, as expected. Figures 3.3(g-i) indicates that introducing β has altered the percolation threshold. This can be seen by $\beta = 0.25$ outlining a more noisy and fractal-like spreading pattern despite being well beyond the standard percolation threshold of $\rho_c = 0.592$.

Figure 3.4 shows how variations in the infection lifetime can change the wave-front properties. The value of β predominantly controls the speed of the wave-front, whereas T controls the lag-time on the removal front. Therefore, increasing T yields an increase in the wave-front thickness; this is valid for $\rho > \rho_c$. Around the percolation threshold $\rho \sim \rho_c$, the relationship between T and spreading velocity is less obvious. Close to the percolation threshold, variations in T have more importance, as it could lower or raise pathogen transmission below or above percolation thresholds, respectively. If T is held fixed, the critical threshold definition can now be generalised from Equation (3.4), to include the parameter β:

$$\rho_c = sup\{\rho, \beta : \theta(\rho, \beta) = 0\} \tag{3.4}$$

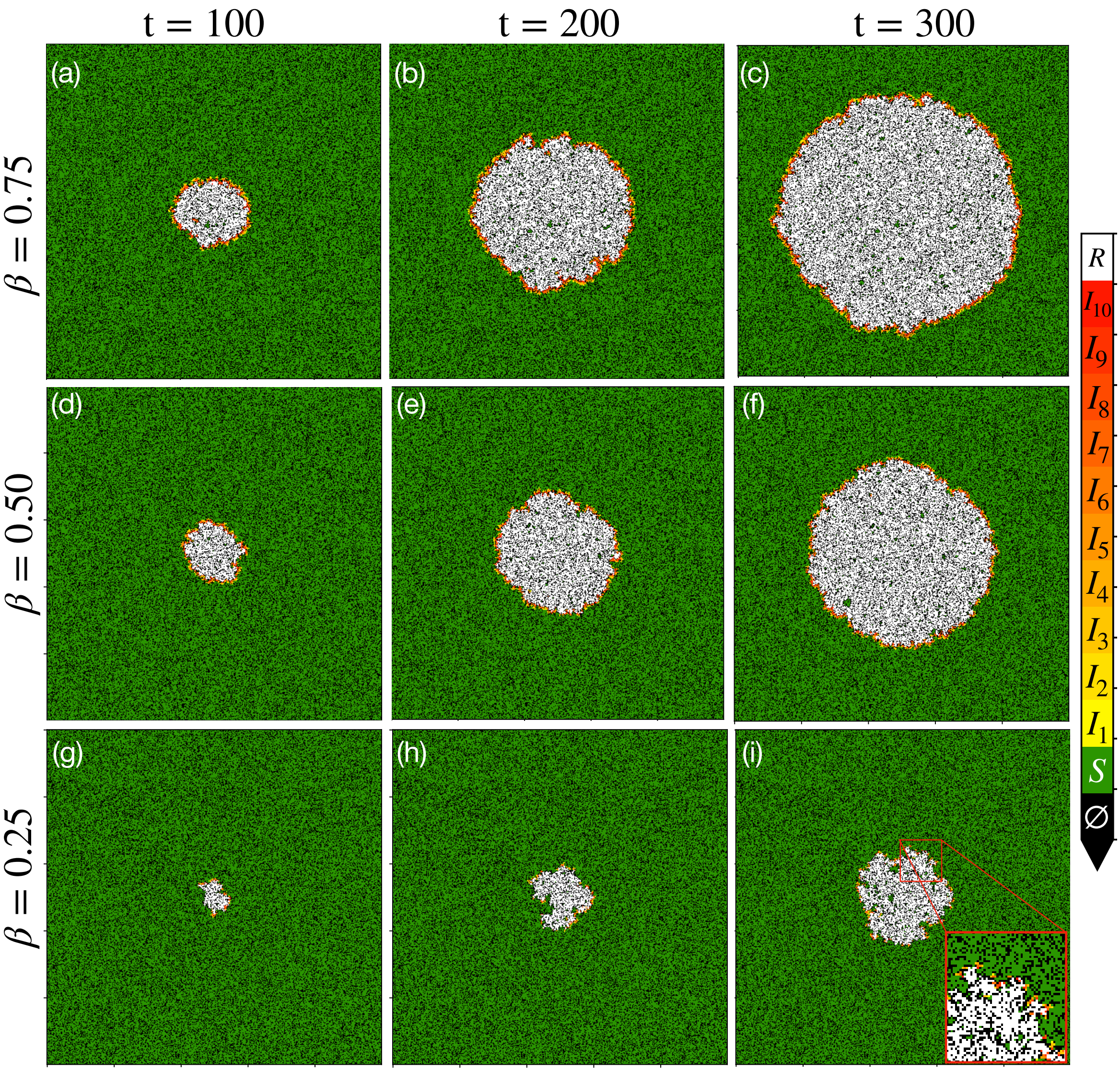


Figure 3.3: Introducing an infectivity parameter β. The SLM is shown running on a domain of size 500×500 for fixed $T = 10$ and density $\rho = 0.70$. Simulation reveal that β has an impact on the wave-front speed and changes the percolation threshold.

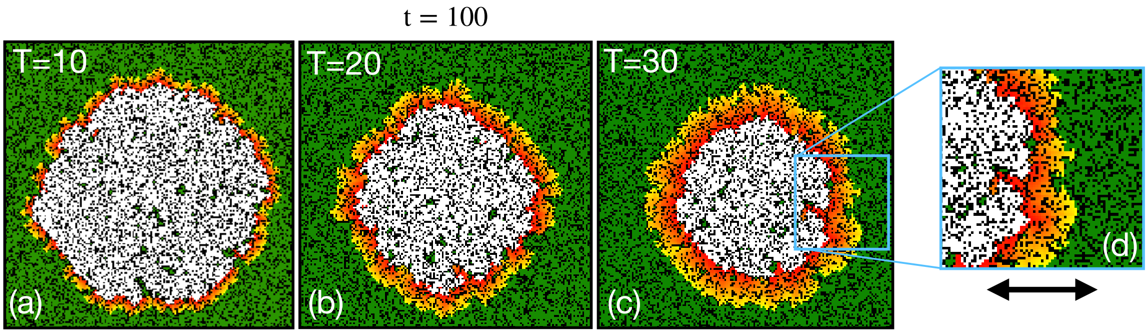


Figure 3.4: Simulations with varying infectious lifetimes T, shown through the time-step $t = 100$. For fixed ρ and β above the threshold, varying T has no impact on the speed of the wavefront. However, increasing T leads to a thicker wavefront, as the posterior interface (between R and I) lags behind the evolving wave-front (between S and I).

3.4 Epidemic thresholds

Previously in section 3.2.1, the percolation probability was examined over one parameter of host density. A threshold occurred over a narrow range of densities and separated the regime of epidemic and pathogen extinction. Notwithstanding, we did not examine the rate of pathogen propagation nor any additional dynamic information. In order to capture spreading rates, a metric comprising the 'spreading velocity' is introduced, captured by:

$$v_{radial(t)} = \sqrt{N_I(t+1) - N_I(t)} \tag{3.5}$$

where, N_I is the number of infected trees in the domain at time-step t.

Equation 3.5 is chosen because the difference between $N_I(t+1)$ and $N_I(t)$ gives an 'effective' radial velocity, where the number of spatial units progressed by the pathogen averaged over all angles per unit time. Moreover, as an infectious wavefront travels outward over an area, the number of infected trees generally grows quadratically. Hence, we effectively normalise Equation 3.5 by taking the square root. Otherwise, we might expect misleading increases in the velocity metric. However, strictly speaking, Equation 3.5 is valid only for $\rho > \rho_c$. Understood because the wavefront assumes a fractal dimension (i.e. < 2) at the percolation threshold, and averaging over two dimensions becomes unreasonable.

The time-series, $v_{radial}(t)$ is shown in Figure 3.5(a) for three combinations of (ρ, β). Unsurprisingly, higher-valued combinations produce a higher velocity. Additionally, initial instability is most significant through the first ~ 200 time-steps, a consequence of the initial conditions which suggests the system has yet to reach a steady state. Here, the system is said to be in a transient state. In Figure 3.5(a), a simulation average $\overline{v}_{radial}(t)$, can be determined and plotted as a horizontal line, repeating the measurement multiple times over an 'ensemble', gives a probability distribution.

Figures 3.5(b-d) display the probability density functions[2] of the mean velocities. In all

---

[2] To reduce artefacts of initial transience, any simulation that became extinct before the initial tran-

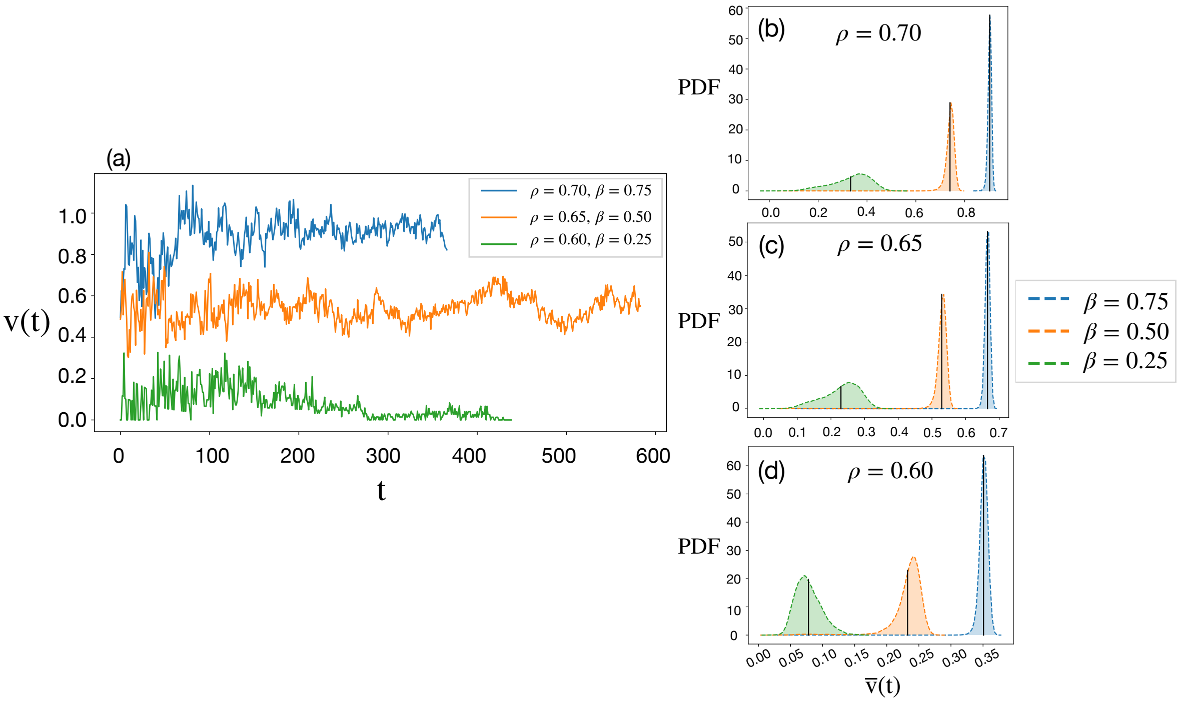


Figure 3.5: (a) The velocity metric time-series $v_{radial}(t)$ is shown for three typical simulations, higher values of ρ and β give higher mean values of propagation speed. (b-d) The probability density function of mean spreading velocity $\langle \overline{v}_{radial}(t) \rangle$ for variations in ρ and β.

panels, the vertical black line inside each distribution depicts the ensemble mean velocity. For higher host densities, the mean velocity increases and distributions appear somewhat narrower. Additionally, Figures 3.5(b-d) suggest that β plays a role in determining the velocity variance, as the distributions become wider with lower infectivity, cf. the green and blue distributions.

For now, the important statistic is merely the ensemble mean, denoted by $\langle \overline{v} \rangle$. although, from Figures 3.5(b-d), we can begin to access the third-order statistical moment of skew. Namely, the green distributions become progressively right-skewed as density is lowered. Statistical measures over the ensemble have exciting applications and can detect an early warning signal, a topic covered in the next Chapter.

Figure 3.6 contrasts the epidemic threshold according to both percolation and velocity-based metrics. In Figure 3.6(a), a one-dimensional line through the parameter-space of ρ reveals the percolation probability for the three different infectivity parameters

---

sient period of $t_{tr} \approx 200$, was excluded from calculations of the ensemble mean.

shown. The vertical dashed line shows the accepted percolation threshold, ρ_c, for a two-dimensional square lattice. For $\beta \in [0.50, 0.75]$, the probability of percolation is identical to that shown in Figure 3.2. However, a lower value of $\beta = 0.25$ decreases the percolation threshold, as evidenced by the green line shifting to the right. Figure 3.2(a) intuitively demonstrates that a pathogen with a low value of infectivity requires a more significant tree density to spread.

The ensemble-averaged radial velocity $\langle \overline{v} \rangle$ (as per Equation 3.5) mirrors the percolation threshold, notwithstanding with some differences. In Figure 3.6(b), we witness a significant increase in the propagation speed when the density crosses the threshold density ρ_c, notably for $\beta = 0.75$ and $\beta = 0.50$ shown in blue and orange. In addition, a higher β-value predictably yields a higher radial velocity, both before and after the epidemic transition.

The green curve ($\beta = 0.25$) in Figure 3.6(b) alludes to an important regime in the model. Namely, a regime of 'persistence' [Gilligan and van den Bosch, 2008]. The vertical dashed green line highlight when epidemics begin to propagate to the domain edge reliably (95% percolation). Thus, the chance of percolation in Figures 3.6(b) is high, yet the velocity remains close to zero. Together, these observations reveal a region in parameter space where epidemics can survive for long periods, barely above the threshold. Nonetheless, host population growth is omitted from the model so persistence is only approximate.

The equivalent two-dimensional plots over the full parameter-space of ρ and β are shown in Figures 3.6(c-d) for percolation and radial-velocity, respectively. Percolation in Figure 3.6(c) reveals an abrupt transition between two stable regimes of extinction and epidemic, separated by a narrow range of critical parameters ρ_c and β_c. In Figure 3.6(d), the ensemble-averaged velocity reveals a smoother transition between states. As the distinction between epidemic states is more apparent in Figure 3.6(c), one might argue that percolation captures the threshold with greater clarity. When infectivity is lower ($0.15 < \beta < 0.30$), the percolation threshold can be crossed by raising the density, reflected in both velocity and percolation metrics. However, percolation has no dependence

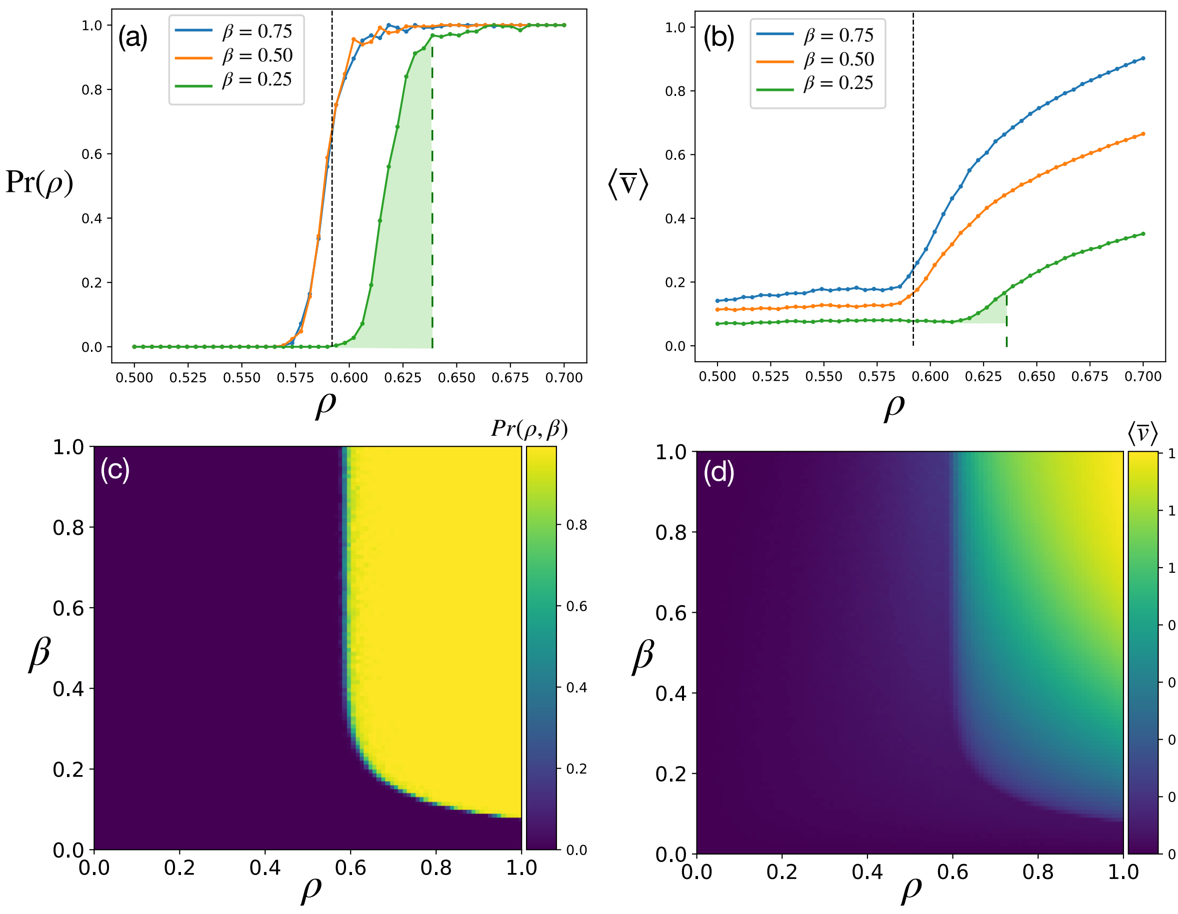


Figure 3.6: A parameter sweep of ρ and β over the 500×500 domain. (a-b) The percolation probability is shown alongside the radial velocity over a one-dimensional parameter-space of ρ. For the lower-valued infectivity of $\beta = 0.25$, the density threshold is slightly higher than classical percolation $\rho_c = 0.592$, indicated by the vertical dashed line. A gradual increase occurs when using the radial velocity, whereas percolation shows an abrupt transition. (c-d) A two-dimensional parameter sweep paints a similar picture.

on infectivity beyond $\beta \approx 0.40$, in contrast to the radial velocity that tends to increase for higher infectivities, revealing a subtle difference between metrics. Here, the parameter-sweeps can be used to portray an '*epidemic phase diagram*', which details regions in parameter space where epidemics are possible or not.

3.5 Discussion

This Chapter introduced percolation theory, leading to a one-parameter percolation SIR model of tree disease spreading through a forest. The one-parameter model exhibited stochasticity on account of the randomly populated host distribution inside a square lattice. Assessing the model on one lattice configuration signifies a limitation in our analysis, and a comprehensive treatment might consider examining the spread of disease on alternate lattices, e.g. the triangular, honeycomb or Voronoi lattice. Nevertheless, although we examined behaviour on one lattice type, the universal properties of the model would remain fixed. For example, the critical exponent describing the correlation length ξ—as discussed in section 3.2.1.

In section 3.3, an infectivity parameter (β) was included in the model. The infectivity parameter introduced temporal stochasticity into the system because infections spread probabilistically between NN, as opposed to the one-parameter model that spreads between hosts with a probability of one. Moreover, the infectivity parameter slightly altered the critical density. Still, higher values of β recovered the standard percolation threshold for a two-dimensional square lattice ($\rho_c \approx 0.592$) whereas lower β values increased the critical density needed to support an epidemic.

In nature, a pathogen might interact with diverse hosts over numerous environmental conditions. A relevant and interesting example includes the algae-like oomycete *Phytophthora ramorum* (*P. ramorum*). The pathogen *P. ramorum* affects a wide host range [Grünwald et al., 2012], including Larches (deciduous conifers of the genus *Larix*) and southern beech (*Nothofagus*), and some non-native oaks such as red oak (*Quercus rubra*) [Grünwald et al., 2008]. Moreover, different genetic variants of *P. ramorum* affect different hosts with varying severities. For example, North American oak species appear more susceptible to the NA1 lineage [Rizzo et al., 2002], while UK larch trees appear slightly more highly susceptible to the European lineage EU1 [King et al., 2015]. Accordingly, although the integration of β in the SLM is undoubtedly simplistic and general, it permits a description of varying degrees of pathogen virulence nonetheless.

Two metrics categorised behaviour in the model, a pathogens average radial velocity and a probability of percolation. The percolation probability defined a sharper, more reliable transition between epidemic and pathogen extinction. Conclusively, the percolation probability proves more accurate to classify epidemic regimes. Nevertheless, the radial velocity captured a similar threshold-like behaviour besides additional dynamic information in the form of a time series. Subsequently, the time series velocity was ensemble-averaged and examined over a sweep of ρ and β values.

Interestingly, comparing percolation to the radial velocity revealed a parameter region (of persistence) above the threshold where the pathogen survives but slowly propagates to the domain boundary. However, as host population growth is neglected in the SLM, long-term host-pathogen coexistence between is prevented and persistence is merely approximate [Gilligan and van den Bosch, 2008].

As we look to construct a more representative model, it is worth remarking on the most significant assumptions that underpin the SLM so far:

1. **Local NN interactions**: in reality, many dispersal mechanisms exist to propagate the spread of disease—see section 2.2.1 for more information. Undesirably, local interactions in the SLM revealed unnatural artefacts of lattice geometry, witnessed in Figures 3.1(a-c). Moreover, as the pathogen could not jump over empty lattice sites, epidemics were only possible above the percolation threshold, i.e. when connected clusters of hosts span the domain.

2. **Uniform dynamics:** in general, pathogen-host interactions transpire with varying degrees of severity. For example, age-dependent severity in ash dieback and the species-level virulence of *P. ramorum.* In contrast, the SLM considers that: (A) the transition probability into R occurs with a uniform number of steps (B) the probability of transition into I is identical between hosts (C) β is simplistic, with no spatio-temporal (environmental) dependence. More broadly, we have constructed a general, abstract model with no specific pathogen in mind and uniform, simplistic interactions.

3. **Host distribution**: natural tree populations present rich spatial structures with varying degrees of clustering and aggregation [Plotkin et al., 2002]. Indeed population clustering has been confirmed over various spatial, from the tree-level to the field-level [Wiegand et al., 2007]. Therefore, the flat randomly distributed host population considered in this Chapter falls short of a realistic description. In addition, the SLM describes the spread of disease in a densely populated forest. Hence, as we look to model the spread of disease over Great Britain, realistic tree canopy cover data should be incorporated into the framework.

Despite the assumptions and simplicity of the SLM, it provides a general setting upon which to elaborate. The next Chapter outlines all the applications that were investigated using the SLM.

Chapter 4

Simple lattice model: applications

Previously, a percolation-based SIR model of tree disease spreading through a forest was outlined, named the 'simple lattice model' (SLM). The SLM provides a flexible foundation to generalise as we look to model the spread of disease over realistic landscapes focused in Great Britain. This Chapter aims to examine some applications of the SLM. In particular, two applications divide the Chapter, beginning with early warning signals (EWS) for forest management and ending with a toy model of landscape-level spread.

Firstly, the system for EWS detection put forward by [Orozco-Fuentes et al., 2019] is extended. The original publication considered a fixed infectivity parameter ($\beta = 0.50$), here we generalise the analysis to the entire β parameter space. In addition, we employ an alternative metric that permits a more precise EWS detection.

Secondly, the SLM will be adapted to construct a toy model of landscape-level epidemics in Great Britain. Several tree distribution datasets in Great Britain are presented and compared, before coupling the SLM with predicted abundance data given by [Hill et al., 2017]. More specifically, units of individual trees in the SLM are re-scaled to 1km × 1km patches and projected onto a predicted oak abundance. The toy model denotes the first step towards a more representative framework over realistic landscapes.

4.1 Early warning signals

[Orozco-Fuentes et al., 2019] investigated applications of EWS for forest management and ecosystem services. The manuscript considered a lattice with fixed infectivity β and a one-dimensional parameter space of tree density ρ. Notably, the study observed significant changes in the statistical measures of variance, skew and auto-correlation in the spreading pattern just before a transition into the regime of epidemic spread. More specifically, the authors found compelling indications of 'early warning' signals within the model, achieved by calculating the moment-generating functions (i.e. variance, skew and auto-correlation) of a time series velocity. Consequently, knowing when to pre-empt epidemics in tree populations could help prevent the spread of emergent infectious diseases and conserve resources.

Here, we offer a small extension to the work presented by [Orozco-Fuentes et al., 2019]. In particular, EWS are detected more precisely using an alternate domain and velocity metric that mitigate artefacts of domain geometry—discussed more in-depth below. Moreover, the analysis is generalised to two dimensions in the parameter-space of ρ and β. After introducing these alternative concepts, we discuss problems and complexities encountered when ensemble-averaging simulations. Ultimately, these results may provide helpful information to aid forest and woodland managers in preserving tree health.

4.1.1 Cylindrical geometry

The metric [Orozco-Fuentes et al., 2019] employed to quantify EWS had a similar form as Equation 3.5, though it included both the infected *and* removed (N_{I+R}). In contrast, Equation 3.5 rests solely on N_I. Unfortunately, a radial velocity based on N_{I+R} can lead to unintuitive metric observations when geometrical effects seemingly cause the rate to in as the wave spreads outward, which becomes particularly visable for later times when the wavefront extent is considerable[1]. Thus, we will use an alternate lattice geometry to

---

[1] See Figure 4d in [Orozco-Fuentes et al., 2019], the apparent rate of increase in the velocity metric is purely due to geometry artefacts and the metric definition.

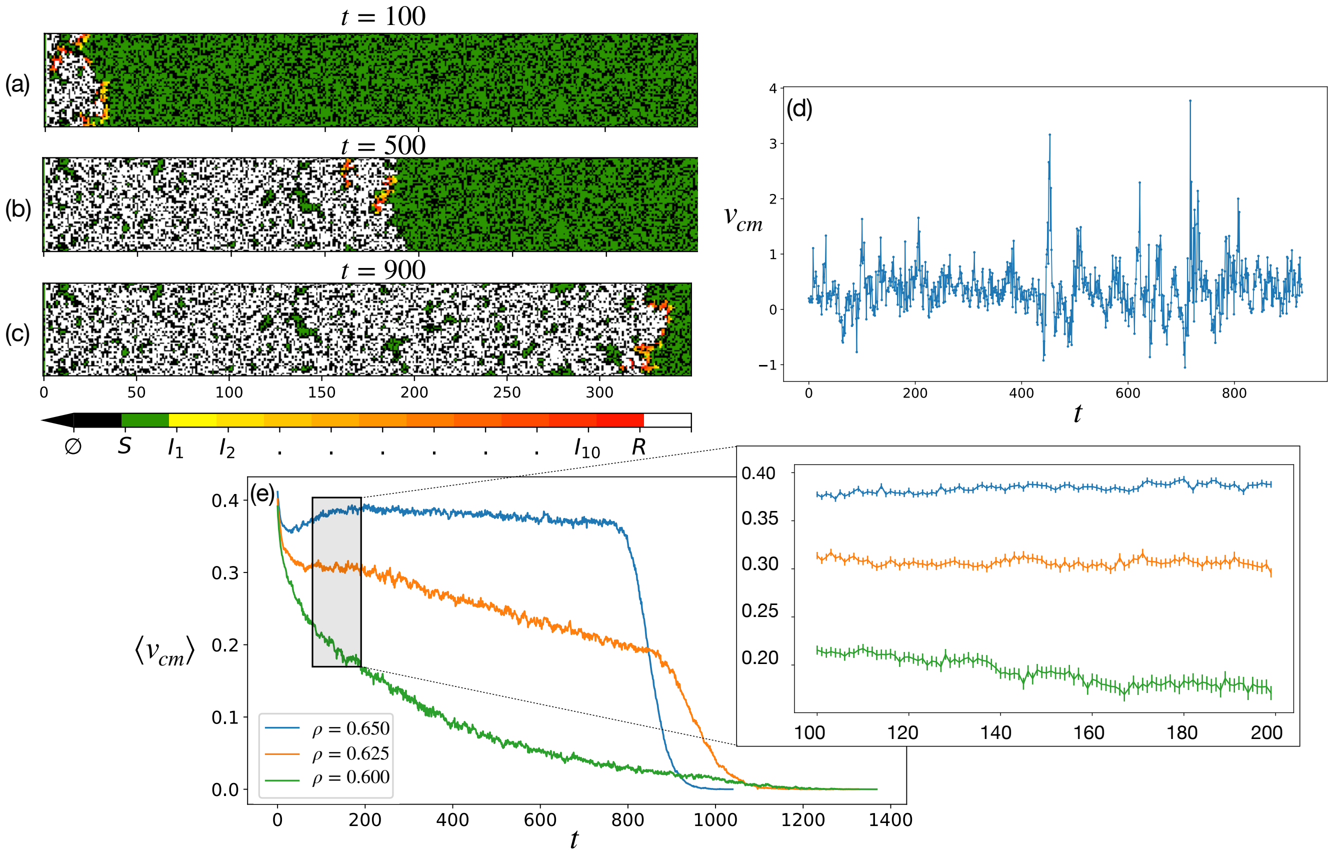


Figure 4.1: (a-c) A channel domain of size 50 × 350 is shown over three time-steps for model parameters $\rho = 0.65$ and $\beta = 0.50$. The centre of infectious mass is recorded for each time-step. (d) Plots of the centre of mass time-series for the simulation illustrated in panels (a-c). (e) The mean centre of mass time-series (of 10^4 repeats) for three variations in density and $\beta = 0.50$. Time-series begins to decay around the mean simulation run-time. The zoomed inset shows the ensemble averaged time-series for $t \in [100, 200]$ and reveals increases in error bars lower density parameters.

mitigate geometric effects.

Consider a rectangular domain of size $[L_x, L_y]$ where $L_x > L_y$ and (x, y) represent dimensions of width and height respectively. Figures 4.1(a-c) shows spatio-temporal spread in the domain, henceforth referred to as the 'channel' domain, over three time-steps. Epidemic parameters in the channel were arbitrarily chosen, yet lie above the threshold ($\rho = 0.65$ and $\beta = 0.50$). The initial conditions required an adjustment from the square domain in Chapter 3, namely, the first column is infected and taken as the origin (denoted by x_0). Additionally, the boundary conditions permit the disease to propagate freely in the $+x$ and $\pm y$ directions, thus realising a cylindrical geometry with periodic boundary conditions in y direction and fixed boundary conditions in x.

We register a percolation event in the channel domain if an infected tree reaches the last column, denoted by x_m. The percolation threshold shifts toward higher densities if the domain is sufficiently narrow (having a high aspect ratio), as the dimensionality of the travelling wave to somewhere between one and two dimensions[2]. The higher percolation threshold can be understood by considering gradual increases in the domain aspect ratio, where, in the limit $L_y = 1$ and $L_x \gg 1$, a one-dimensional domain is realised and the critical host density increases to $\rho_c = 1$. In other words, increasing the aspect ratio eventually leads to the (higher-valued) one-dimensional percolation threshold.

Unsurprisingly, studies have confirmed that site and bond percolation (in $\mathbb{R}^2$ and $\mathbb{R}^3$) depend non-trivially on a cylinders aspect ratio [Sangare and Adler, 2009]. Therefore, we choose an aspect ratio that approaches the percolation threshold for a two-dimensional square to keep consistent with the numerical results of Chapter 3. In practice, a channel of size $(L_x, L_y) = (350, 50)$, was sufficient, as shown through Figures 4.1(a-c).

4.1.2 Centre of infectious mass

The channel domain provides an advantageous setting to capture EWS by avoiding two-dimensional geometrical effects. Moreover, the channel permits an improved (more intuitive) metric, based on the mean infective displacement from the origin:

$$v_{cm}(t) = \frac{\sum^i x_i(t)}{N_I(t)} - \frac{\sum^i x_i(t-1)}{N_I(t-1)} \tag{4.1}$$

where $x_i(t)$ is the spatial location of the i^{th} infected tree along the x axis at time t and $N_I(t)$ is the total number of infected trees.

The form of Equation 4.1 displays intrinsic similarities to the Newtonian centre of mass, given by:

$$x_{cm} = \frac{\sum^i x_i \times m_i}{\sum_i m_i} \tag{4.2}$$

where the analogous quantities in Equation 4.1 and 4.2 are $m_i = 1$ and $\sum^i m_i = N_I$.

---

[2] Additionally less space in computer memory is occupied and simulation time is lowered

Thus, Equation 4.1 averages the displacement of infected trees precisely as Equation 4.2 averages the displacement of Newtonian mass. Accordingly, Equation 4.1 will be denoted as the 'centre of infective mass' (COM) velocity $v(t)_{cm}$. Figure 4.1(d) displays the time series $v(t)_{cm}$ of Figure 4.1(a). Notably, the time series in Figure 4.1(d) allows for negative values and looks different to that shown previously in Figure 3.5(a).

Ensemble-averaged COM time-series $\langle v_{cm} \rangle$ are shown in Figure 4.1(c) for the three values of tree density. The blue time-series has epidemic parameters well beyond the threshold and begins to decrease around $t \in [800, 900]$, coinciding with percolation to the domain edge. Contrarily, the green time series lies slightly above the percolation threshold and decays more gradually from $t = 0$, due to a higher extinction probability. A small number of long-lasting simulations (exceeding $t > 1000$ steps) occurred in the green time-series, indicating criticality in the system—previously likened to a regime of persistence in section 3.4. The inset of Figure 4.1(c) shows the ensemble-averages between $t \in [100, 200]$, including the standard error for each time-step. Notably, error bars are most significant for the lowest-valued density shown in green, reflecting a more chaotic spread.

4.1.3 Ensemble averaging method

The findings of [Orozco-Fuentes et al., 2019] were gathered by first producing a distribution of mean time-series velocities $\overline{v}_t$. In this scheme, EWS were detected from statistical moment-generation functions over the mean velocity, e.g. $\big\langle var(\overline{v_t}) \big\rangle$. However, altering EWS detection by computing the mean simulation variance $\big\langle \overline{var}(v_t) \big\rangle$ was found to reveal a clearer signal. Consequently, we proceed by detailing an ensemble method that permits the capture of 'within-simulation' variance.

Before the ensemble averaging method is elaborated, it makes sense to first define a relevant mathematical notation. Suppose a simulation with parameters ρ, β propagates for f time-steps, Equation 4.1 describes the time series as: $v_{cm}^{t=1}, v_{cm}^{t=2}, ..., v_{cm}^{t=f} \in V^{\rho\beta}$. Then, a set of N independent time series are generated by repeating N ensemble realisations. For an arbitrary point in the parameter-space (ρ, β), the set of N replicate simulations can be described by the set: $\{V_1^{\rho\beta}, V_2^{\rho\beta}, ..., V_N^{\rho\beta}\} \in \mathcal{V}_{\rho\beta}$, where each $V_i^{\rho\beta}$ describes an individual simulation time series and $\mathcal{V}_{\rho\beta}$ describes the entire ensemble for parameters (ρ, β). Thus, an EWS is detected by calculating the mean time-series variance, defined by:

$$\langle \overline{var}(v_{cm}^{\rho\beta}) \rangle = \frac{1}{N} \sum_{i=1}^{N} var(V_i^{\rho\beta}) \tag{4.3}$$

A proper time-series analysis requires the same number of observations within each ensemble. If not, we risk mistaking statistical fluctuations/errors for an EWS. There are two observations to consider: (A) the number of time steps within simulations (B) the number of repeated simulations, N.

In general, stochasticity will prevent two simulations from having the same number of time steps, $|V_i^{\rho\beta}| \neq |V_j^{\rho\beta}|$. Thus, calculating time-series variance for simulations with a small number of time steps might be more error-prone than long-lived simulations. As such, we introduce a fixed window of time-steps ($t_O \leq t \leq t_F$) in a bid to fix the number of observations. Provided the window length $t_F - t_O$ captures a sufficient number of time-steps, we avoid significant fluctuations in $V_i^{\rho\beta}$. Variance inside this window is defined by:

$$\langle \overline{var}(v_{cm}^{\rho\beta}) \rangle = \frac{1}{N} \sum_{i=1}^{N} var(V_i^{\rho\beta}\Big|_{t_O}^{t_F}) \tag{4.4}$$

Initial transience constrains the particular choice of t_O in the channel, which distort calculations of the time-series variance. The lower bound was set to $t_0 = 100$ because initial instability occurred most over the first 100 time-steps, and the window upper-bound to $t_F = 200$ (a two-fold increase of t_O) so that simulation variance is measured over a sufficient number of steps. Lastly, the simulation boundary conditions required a slight alteration to fix the number of variance measures to N. Previously, one of three events terminated a simulation, either: percolation, pathogen extinction, or when the number of time-steps exceeds the time horizon. Here, we relax the condition that terminates simulations when the pathogen dies off; otherwise, short-lived simulations ($t < t_O$) would be omitted from the ensemble and reduce the number of variance measures below N. Additionally, the time horizon is fixed to t_F, thereby mitigating the cost of simulating unnecessary steps.

4.1.4 EWS parameter-sweeps

Here, an EWS is defined as‘*statistically significant changes in the variance of* $v(t)_{cm}$ *prior to a critical shift into the epidemic regime*’, calculated according to Equation 4.4. Specifically, EWS are calculated by assessing the ensemble-averaged variance of velocity $v(t)_{cm}$ over each simulation between time steps $t_O = 100$ and $t_F = 200$, justified by the reasons laid out above in section 4.1.3. However, it is worthwhile to remark that, in general, EWS are defined in terms of *any* statistical signals which indicate a system is approaching a critical tipping point [Scheffer et al., 2009]. Notwithstanding, Figure 4.2 displays the ensemble-averaged time-series variance, as per Equation 4.4. The colour bar shows variance over the entire parameter space of density and infectivity from white to back. In Figure 4.2, the lower and upper red lines indicate a percolation probability of $Pr(\rho, \beta) = 0.05$ and $Pr(\rho, \beta) = 0.95$, respectively. Inside these regions, the system transitions into an epidemic, and we witness a considerable rise in the variance.

The observations of Figure 4.2 agree with the results of [Orozco-Fuentes et al., 2019]. Although measuring EWS over a two-dimensional parameter space reveals some additional information not captured in the original analysis, i.e. when infectivity is low ($\beta < 0.40$), EWS preempt the epidemic by a more significant margin—indicated by the red arrows. In addition, variance over the epidemic transition appears sharper when β is high, as indicated by the darker shade in the upper right quadrant of Figure 4.2.

The following thought experiment can explain the EWS asymmetries in Figure 4.2: suppose infectivity is high, and density lies just below criticality ($\rho \lesssim \rho_c$), so susceptible clusters do not percolate. In this case, disease transmission is high on account of β, and all susceptible hosts become quickly infected, although an outbreak will come to a halt due to insufficient hosts. Therefore, in this model, an aggressive pathogen spreading through low tree densities may propagate rapidly but have a short, chaotic signature. Hence the transition is steep, and the simulation variance is significant, as indicated by the smallest

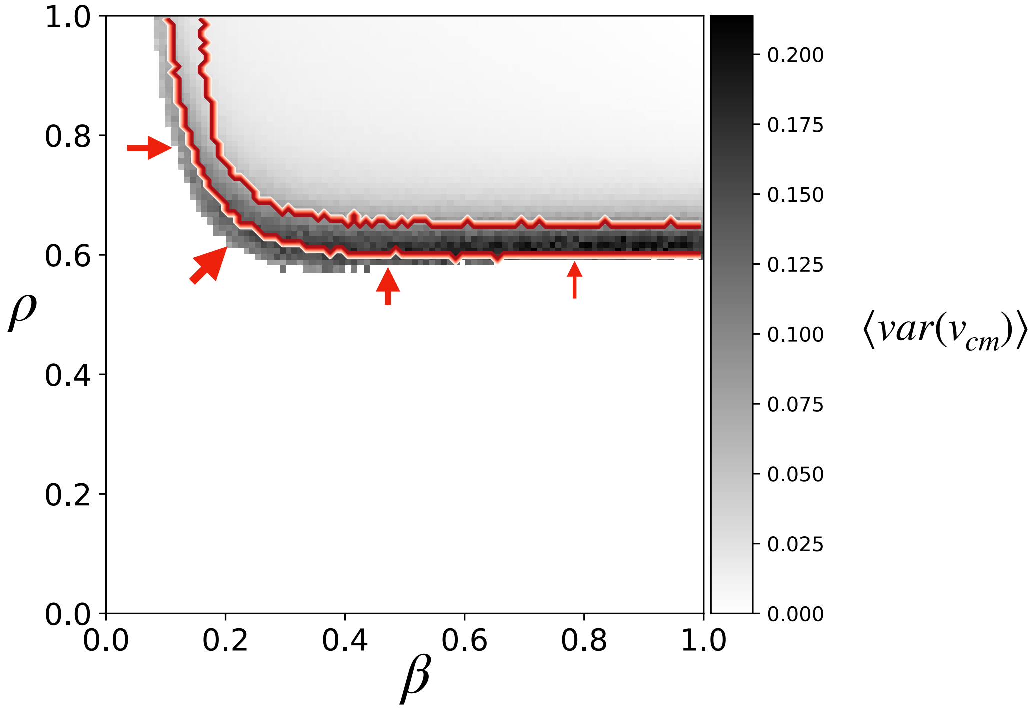


Figure 4.2: The ensemble-averaged variance of $v_{cm}(t)$ over a two dimensional parameter sweep of ρ and β. Red contours show the lower and upper bound of percolation (i.e. between 5% and 95% probability). The epidemic regime is pre-empted by increases in variance more clearly for certain parameter values, indicated by the arrows.

arrow in Figure 4.2. In this region, detecting an EWS is hard because transitions occur most rapidly.

Now we consider the converse, a less infectious pathogen with an abundance of susceptible hosts, located in the top-left region of Figure 4.2. Here, transitions into the I compartment are slower because the pathogen is less infectious, although this time, hosts are abundant. Thus, a pathogen is likely to spread predictably for longer times, giving rise to a slightly less abrupt variance signature, as indicated by a lighter colour before the transition in the top left quadrant of Figure 4.1. Although the variance spike is not as significant, it preempts the epidemic transition by a more considerable degree—indicated by the larger arrows in Figure 4.2. Altogether, Figure 4.2 indicates that the strength of an EWS depends on the particular combination of epidemic parameters, though fundamentally, an EWS is detectable for all parameter combinations.

4.2 Tree distribution datasets in Great Britain

Large-scale epidemic models of tree disease rest on robust, high-quality host data. Data-driven approaches are crucial for predicting disease spread over country-wide scales, though collecting high-quality host data involves myriad challenges. Most notably, large-scale species distributions require vast datasets that demand significant economic resources and person-hours to assemble and maintain over time. However, satellite-based remote sensing technologies pose an attractive solution—see [Camarretta et al., 2020] for a recent review of remote sensing technologies. Despite the significant advances of remote sensing technologies, most freely available data sets still rely on traditional surveying methods to collect data throughout Great Britain (GB). Consequently, the most widely known and widely used datasets are reviewed below. Following this, statistically-generated species distribution models, typically based on surveyed data, are reviewed.

4.2.1 National Surveys

Surveyed data predominantly describes either: abundance, presence-only, presence-absence data. Generally, abundance data describes percentage canopy cover per km^2. In contrast, binary-valued presence-only and presence-absence data simply record if a species is present or present and absent, respectively. Abundance captures significantly more information than presence-only data, yet unfortunately, they are in short supply.

Countryside Survey

The countryside survey (CS) is a long-running, national survey of diversity and species abundance in GB [Wood et al., 2017]. The UK Centre of Ecology and Hydrology (UKCEH) undertakes the surveys, primarily funded by the Natural Environmental Research Council alongside other government agencies. Individual surveys have been undertaken in: 1978, 1990, 1998, 2007, and 2019. Random stratified sampling captures a representative species abundance[3] over of all land cover compositions, e.g. lowland acid grassland, freshwater,

---

[3] A useful (unpublished) project merged abundance data from CS with myForest. The abundance data can be found at the Oxford University research archive: `https://ora.ox.ac.uk`.

and broad-leaf forest.

Abundance data is collected for numerous dominant species, including trees, shrubs, ground flora and soil type, making the scope of CS data vast. Moreover, long-running records spanning decades reveal ecosystem trends imperative for ecological monitoring. More recently, 100 1km^2 plots of vegetation and soil data were collected[4] [Smart, 2020]. The dataset constitutes the first of five planned surveys, part of a rolling monitoring strategy collected every five years.

National Forest Inventory

The National Forest Inventory (NFI) collects and maintains forest and woodlands data in GB. Originally, the NFI was established to help restore and expand Britain's woodlands following the First World War [James et al., 1990]. Regular programs (10-15 year intervals) implement surveys of woodland and forest size, distribution, composition and condition across GB. Records cover areas over 0.5 ha and 20% coverage. As of 2019, $622,381$ individual records exist, spanning 2.9×10^6 ha over 13% of the total land cover within GB. NFI data comprise ESRI shape files[5], that outline numerous forest types, e.g. broadleaved, conifer, mixed-predominantly broadleaved or mixed predominantly conifer. Despite an extensive coverage, publicly available NFI surveys describes presence-only data—with no proportion or species coverage. Although, additional datasets are available to purchase, including: (1) Tree species percentage per region by woodland type (2) Tree species proportions within the upper canopy of each NFI sample plot, without supplying the exact location of the individual sample plot.

Botanical Society of Britain and Ireland

The Botanical Society of Britain and Ireland (BSBI) has recorded species presence-only data since 1950. BSBI datasets are publicly available[6] upto a resolution of 2km × 2km, though records upto 100m × 100m are available to registered members. As of 2020, BSBI

---

[4] The data is free to download on the UKCEH website: `https://catalogue.ceh.ac.uk`

[5] Free to download at: `https://data-forestry.opendata.arcgis.com`

[6] BSBI data can be downloaded from: `https://database.bsbi.org`.

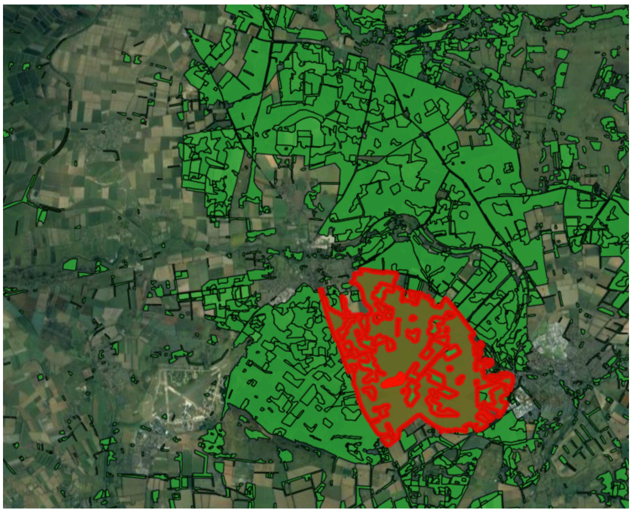


Figure 4.3: NFI data super imposed onto a Google earth image, taken from a report (unpublished) by S. Orozco-Fuentes et al. NFI data covering Thetford Forest Park (16.684km^2) is shown as a polygon in the NFI 'woodland' category. Data is interpreted as the conifer forest type. Here, surveys comprises presence-only data, and no tree species percentage cover is reported. NFI data extends throughout ∼ 13% of land coverage in GB and large non-woodland areas remain un-surveyed.

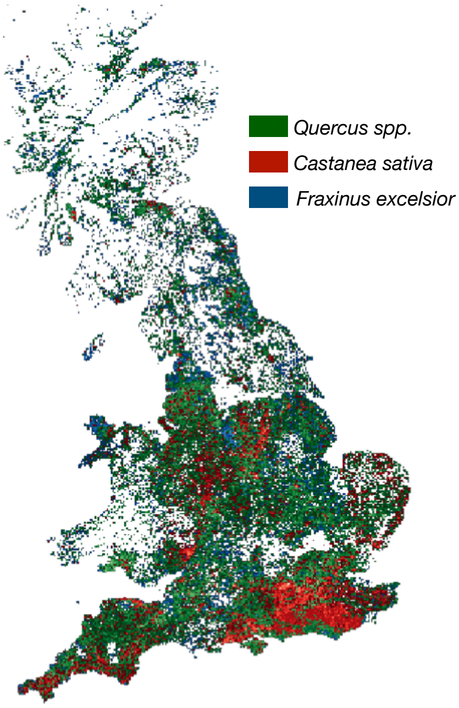


Figure 4.4: BSBI presence-only datasets—as reconstructed by S. Orozco-Fuentes et al. (unpublished). Three important large deciduous tree species, European ash (*Fraxinus excelsior*), Oak (*Quercus spp.*), and sweet chestnut (Castanea sativa), are overlaid onto the same map at 2km × 2km resolution. The BSBI datasets are extensive and report presence-only data over a country-wide scale.

records are collected at 1km × 1km square resolution or better—making the datasets among the highest-resolution surveys collected by traditional methods. The BSBI distribution database contains records of plants and charophytes as reported by users and conservationists with MapMate[7]. Despite the availability of high-resolution data, observations are collected ad hoc by users and not curated scientifically. Moreover, the distributions contain both well-surveyed and poorly-surveyed plots of land likely to carry uncertainties. As such, BSIBI data is helpful to reconstruct several baseline tree distributions across GB, as demonstrated by [Hill et al., 2017].

4.2.2 Species distribution models

In the absence of extensive host data, species distribution models (SDMs) aim to generate synthetic data, typically from less-extensive surveys. First developed in the 1990s, SDMs have become fundamental to ecological and biogeographical inference studies. SDMs rest on statistical algorithms (discussed more below) that process and correlate environmental data to predict species habitat suitability in space and time, as depicted in Figure 4.5. Thus, the overall SDM approach differs significantly from traditional mechanistic (e.g. compartmental, agent-based, and PDE) spread models, as discussed in section 2.2.

Synthetic distributions have numerous applications for examining biodiversity, conservation, resource management, ecology, and climate change [Franklin, 2013, Skov et al., 2016, Wittmann et al., 2016, Hunter et al., 2012, Zhang et al., 2019]. Although SDMs have many applications, few studies focus specifically on constructing SDMs for tree-based pathogens. However, recently [He et al., 2019] coupled remote-sensed data with an SDM to produce disease-induced mortality maps of sudden oak death in California. Namely, by coupling remote-sensed mortality data to an SDM of the tree pathogen *P. ramorum*, [He et al., 2019] was able to narrow down areas in the landscape where mortality was disease induced—in contrast to areas affected by other causes of stress and mortality.

The vast majority of SDMs fall into two categories: correlative [Srivastava et al., 2019],

---

[7] MapMate is software designed to aid users to share ecological data: `https://www.mapmate.co.uk`

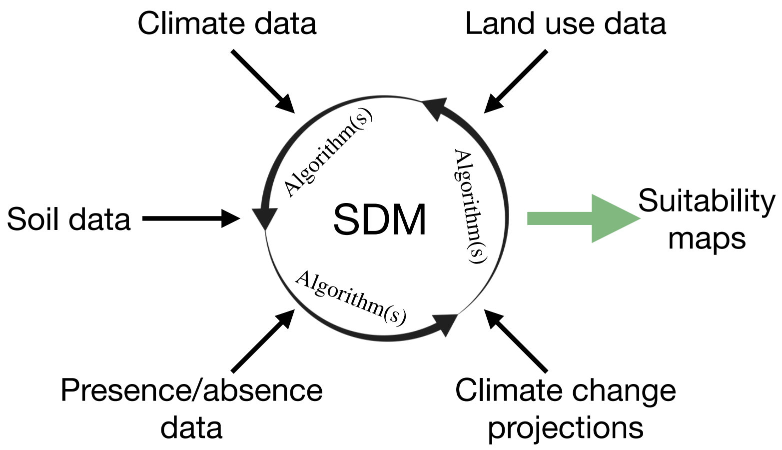


Figure 4.5: A graphical species distribution model (SDM) illustration, adapted from [Pecchi et al., 2019]. A variety of predictor variables and input data sources are used in conjugation with modelling algorithms to produce a habitat suitability map.

and mechanistic [Shabani et al., 2016]. Correlative SDMs relate (widely available) presence-only, or presence-absence, data to several environmental predictor variable datasets. For tree species, predictor variables include temperature, precipitation, altitude, and soil type [Ray et al., 2021, Hill et al., 2017]. Following this, a species distribution map can be predicted, albeit with uncertainties and errors. Commonly used statistical methods include Regression (i.e. General Linear Models, General Additive Models, Multivariate Adaptive Regression Splines) and Machine Learning (i.e. Artificial Neural Networks, Classification And Regression Tree, Random Forest). The general correlative approach is reflected in Figure 4.5; for a more in-depth review of correlative SDMs, see [Pecchi et al., 2019].

Correlative SDM approaches require little to no prior knowledge of the physiological processes that link organism and environment. Hence, mechanistic methods aim to incorporate an organisms behavioural, physiological, and morphological constraints to the environment, as reviewed by [Kearney and Porter, 2009]. However, linking a species physiological response to the environment comes with significant computational challenges, as it typically relies on vast, multi-variable time-series datasets [Shabani et al., 2016].

A review paper by [Guillera-Arroita et al., 2015], revealed a diverse use of SDMs. Out of 100 publications reviewed by [Guillera-Arroita et al., 2015], SDMs were applied to: (1) managing threatened species (16% of articles) (2) predicting climate change impacts (13%) (3) understanding phylogeographic patterns (9%) (4) controlling threatening processes (8%) (5) landscape management (8%) (6) biological invasions (7%). However, no epidemic applications were reported.

Nevertheless, a thorough literature search revealed a variety of epidemiological SDM applications. The crossover between ecological SDM methods and epidemiology has been referred to as ‘Infectious Disease Cartography’ [Kraemer et al., 2016]. With Infectious Disease Cartography, one seeks to map the likelihood, or risk, of infectious disease outbreaks and produce risk-maps. A number of publications have applied SDMs livestock diseases [Hollings et al., 2017], and human-based diseases including the global distribution of Dengue Fervour [Bhatt et al., 2013] and Zika virus [Messina et al., 2016], and Anthrax in Kenya [Otieno et al., 2021]. However, SDM applications for tree disease epidemics appear absent from the literature.

Predicting species abundance

SDM-generated tree occurrence data have limited applications to ecologists and forest managers. This motivates statistically-generated abundance data that includes significantly more information about the population occupation/density and ecosystem. Modellers have examined numerous approaches to predict tree species abundance; including, linking the abundance-occupancy relationship [Gaston et al., 2000] and the scaling pattern of species occupancy over progressively smaller spatial scales [Hui et al., 2009].

An interesting, and highly relevant, approach to predict the abundance of common tree species in Great Britain was put forward by [Hill et al., 2017]. At a high level, BSBI presence-only data were combined with a series of environmental covariates using a species distribution model to produce a map of predicted occurrence data. Then, random forest regression was employed with a training sample (70%) of less extensive abundance data (consisting of CS, myForest and Bluesky’s National Canopy Map). Results were then cross-validated with the remaining (30%) abundance data; Figure 4.6 displays a flow-diagram of the method presented by [Hill et al., 2017].

A more detailed explanation of the treatment proposed by [Hill et al., 2017] follows:

Stage (1)

- *Presence-only BSBI data was downloaded for* 25 *common species of trees in GB, and all records less than* 2km × 2km *resolution were discarded. Next, the presence-only data was converted into presence-absence data by considering 'well-surveyed' records that Hill et.al defined as having a minimum of two survey between* 1950 *and consisting of* 50 *species. Species missing from these well-surveyed areas were assumed truly absent.*
- *Using biomod2 [Thuiller et al., 2016], a SDM was then fitted against a cohort of 15 environmental variables, e.g. soil type (European Soil Database), temperature (Worldclim), precipitation (Worldclim), altitude (Worldclim), type of land cover (Countryside Survey), among others. The net result was a map of predicted occupancy at* 1km × 1km *resolution.*
- *For each species, predictions from a suit of models—GLM, GAM, CTA, GBM, RF—were repeated and combined into an ensemble distribution model. Each model was cross-validated against* 30% *of the well-survyed BSBI presence-absence data using the receiver operator curve (ROC) [Jiménez-Valverde, 2012] and the true skill statistic (TSS). [Hill et al., 2017] then selected the best performing predicted occurrence for each species.*

Stage (2)

- *Abundance data from CS and myForest were both expressed as hectares covered per kilometer squared (*$\mathrm{ha/km^2}$*). This entailed using woodland cover from the NFI dataset to multiply the percentage cover of each species within a woodland patch, with a proportion of woodland cover per kilometer.*
- *Random forest (RF) regression then modelled the relationships between (CS and myForest) abundance data with the SDM-generated map of predicted occupancy. In addition, RF regression used four covariates, three of which consisted of Bluesky's National Canopy data (i.e. total tree cover, woodland tree cover, non-woodland tree*

cover) and NFI edge broadleaved woodland (i.e. within 50m *of non woodland) data.*

The abundance datasets produced by [Hill et al., 2017] combine several mainstream tree datasets in GB; moreover, constructing the ensemble model involved a variety of statistical models. The predicted occurrence data was examined against the ROC. Most species demonstrated functional ROC scores between 0.71 and 0.96 and performed exceptionally well for ash (0.96) and oak (0.90).

Although there were numerous assumptions that underpinned the methodology. Primarily, the BSBI dataset used by [Hill et al., 2017] exists through ad-hoc user and volunteer self-reports. Thus, some regions are more surveyed than others over time, which led Hill et al. to make the 'well surveyed' recorded assumption (i.e. only considering records surveyed twice since 1950 containing a minimum of 50 species). The assumption permitted the conversion of raw presence-only to presence-absence, at the cost of overestimated absence in these regions. That is, even supposing 50 species are reported within a subset of the (2km × 2km tetrad) record, other large regions could remain unsampled—the authors did not appear to scrutinise this assumption sufficiently.

The RF regression used by [Hill et al., 2017] marked a novel approach to modelling species abundance. The authors chose to argue in favour of RF regression because of its insensitivity toward the data distribution, which worked well with the less comprehensive abundance data sources (as the map of abundance had a high percentage of zeros from missing records). The abundance model quality was examined against 10-fold validation (explained in [Refaeilzadeh et al., 2009]). The root-mean-square error (RMSE), between predicted and observed abundance, generally ranged between 5 and 10, where the RMSE scale reflects the response variable units (ha/km^2), i.e. 5% and 10% respectively. The result was country-wide predicted abundance, with noticeable, yet sufficiently low inaccuracies.

The low amount of available abundance data in GB significantly impoverished the abundance maps produced by [Hill et al., 2017]. Consequently, of the 25 species datasets considered, all will contain numerous (small-scale) errors and uncertainties. Nonetheless,

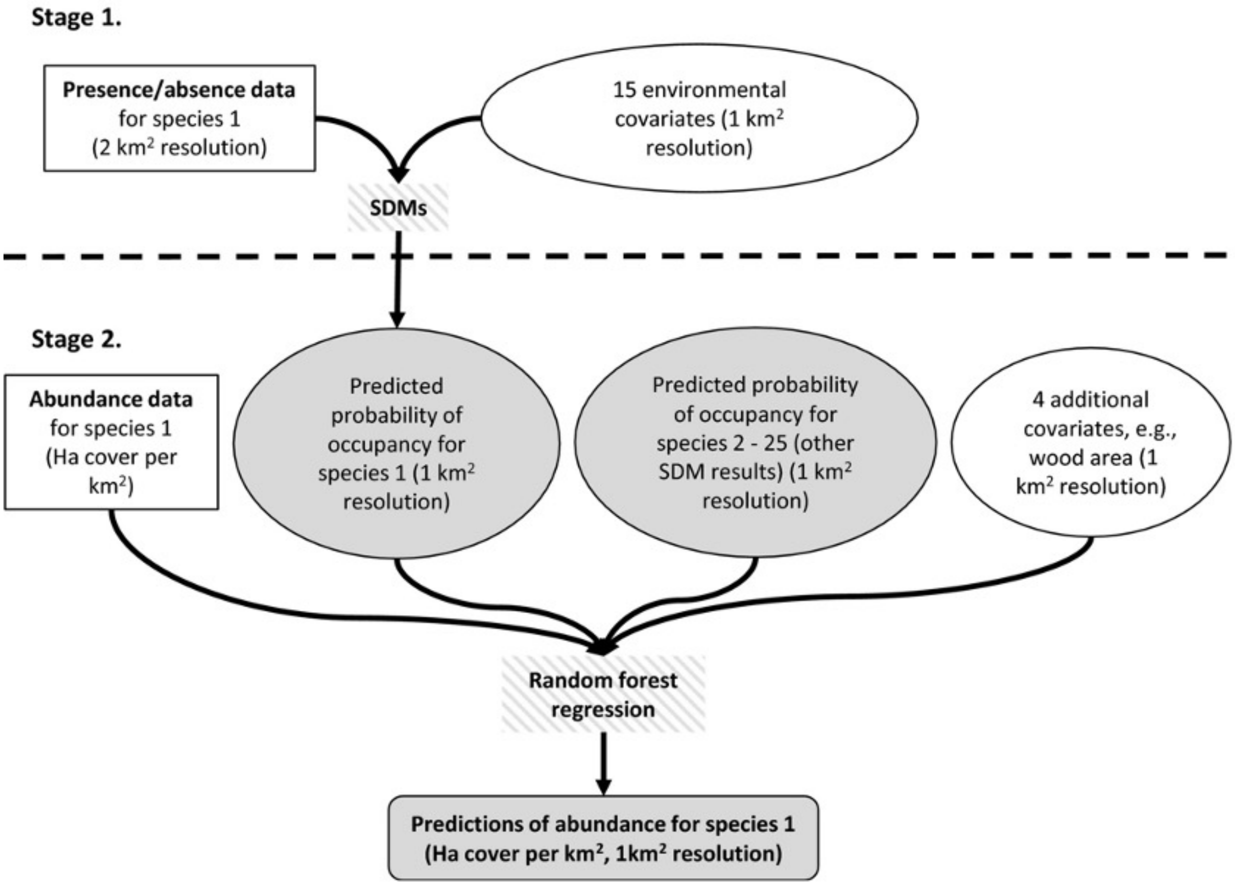


Figure 4.6: A flow diagram of the two-stage abundance method put forward by [Hill et al., 2017] to model tree species abundance (taken from the publications' materials and methods section).

the modelled abundance maps captured more large-scale spatial structures than both the BSBI and CS distributions.

More recently, [Ray et al., 2021] produced a similar SDM as Hill et al. for oak in GB using biomod2 [Thuiller et al., 2016]. The authors focused on mapping high-density oak woodlands (with 60% canopy cover or above) to predict which NFI map polygons (by forest type) were most likely to contain oak stands. However, [Ray et al., 2021] did not make their oak maps publicly available, nor did they produce a general-purpose abundance map relevant for epidemiological studies. To date, the data sets produced by [Hill et al., 2017] constitute the best publicly available country-wide maps of abundance in GB, despite their limitations.

4.3 A toy landscape-level SLM

So far, this work has focused on a homogeneous distribution of hosts randomly populating an ideal domain. In this section, a toy landscape-level SLM is constructed and simulated on the map of Great Britain (GB). Modelling the spread of disease over GB entails a large change of scale within the SLM. Specifically, host units now comprise 1km × 1km '*patches*' of land. Collecting high-quality abundance data over large areas is challenging and expensive—reviewed in section 4.2. Therefore, the SLM is combined with a map of predicted oak abundance produced by [Hill et al., 2017].

4.3.1 Realistic boundaries

Realistic host distributions describe complicated, irregular and heterogeneous domain boundaries known to influence the spread of disease [Madden and Hughes, 1995]. In contrast, the SLM spreads through a square lattice with regular domain boundaries and homogeneously distributed hosts. Moreover, the initial conditions (ICs) in the SLM have been limited to a small *focus* of infected hosts located at the domain centre. However, an epidemic propagating from the centre of a square lattice might look very different from an epidemic emerging from an arbitrarily located epicentre inside a domain with complicated irregular boundaries. As a consequence, we examine the interplay of initial and boundary conditions in the SLM, compelled further by articles highlighting the importance of domain shape [Mikaberidze et al., 2016] and critical domain size [Abad et al., 2020, Reimer et al., 2017].

As a first step toward modelling with realistic host distribution, SLM domain edge effects were examined inside the shape of GB. Figure 4.7(a) shows the map populated with a random homogeneous distribution 'patches' at resolution 1km^2. Here, we change the scale, so host units represent patches of land, as opposed to individual trees in Chapter 3. The green and black pixels represent susceptible and insusceptible patches of land, respectively.

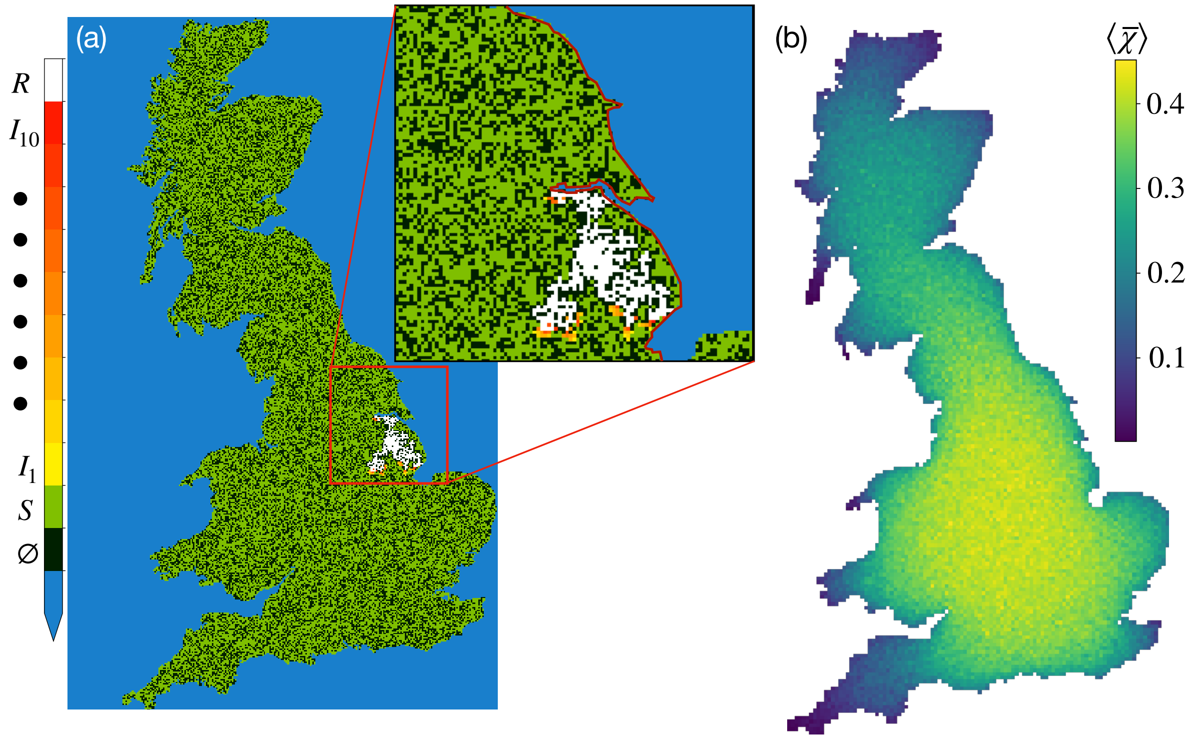


Figure 4.7: (a) The SLM spreading on a map of GB. The domain geometry and epicenter location are non-trivial aspects likely to influence the spread of disease. The colour bar indicates susceptible S, infected I_i (where i is the i^{th} state), and removed R states. Moreover, the zoomed inset shows an example of the Humber estuary preceding an infectious wave-front. (b) A map of ensemble-averaged mortality-ratios χ, shown by colour for each spatial location with SLM model parameters $\rho = 0.65$ and $\beta = 0.25$.

In Figure 4.7(a), we see that coastline edge effects might reduce the probability of an epidemic. For example, epidemics emanating from epicentres located just below the Humber estuary (shown on the inset) could encounter more impedance when compared to more centrally-located positions. Accordingly, SLM edge effects and domain BCs were examined by assessing the tree mortality from every possible epicentre, as shown in Figure 4.7(b).

Figure 4.7(b) reveals the ensemble-averaged mortality ratio, denoted by χ. The toy landscape-level SLM assumed parameter values just above threshold in a 2D square lattice, where each simulation began from a single epicentre located at patch (i, j). Here, the mortality ratio captures the total epidemic impact, defined by:

$$\chi = \frac{R_f}{S_0} \tag{4.5}$$

where R_f is the final number of patches in the removed state at T_f and S_0 is the number of susceptible patches at time $T = 0$. Given the stochasticity, it is necessary to repeat simulations for each epicentre and calculate the mortality ratio $\overline{\chi}$. Iterating over the whole of the GB in this way permits visualisation of the spatial-susceptibility of the pathogen β, depicted by colour in Figure 4.7(b).

Figure 4.7(b) shows the result of the ensemble averaging χ for each patch of land[8]. As expected, the domain BCs and map geometry change the resulting epidemic scale, mainly in Scotland and the south Eastern leg of GB towards Exeter and Plymouth. Regardless, centralised regions show a roughly constant susceptibility. Higher epidemic parameters increased the mean mortality ratio and reduced spatial variations in the mortality ratio, whereas lowering epidemic parameters had the opposite effect. Although Figure 4.7(b) paints an ideal epidemic scenario involving one infected patch at $t = 0$, we can assume an emerging epidemic within the SLM model depends non-trivially on the interplay of epicentre location and domain boundary.

---

[8] Ensemble averages for each patch ($\approx 2 \times 10^5$ in total) was computationally costly. Hence, the domain resolution was coarse-grained to 5km × 5km sized-patches to reduce the number of simulations. Even so, simulations were ensemble-averaged on the Leeds high-performance computing facility (ARC) using a task array. The ensemble took approximately two hours on 25 cores.

4.3.2 Realistic abundance data

Figure 4.8(a) shows the distribution of oak in GB given by [Hill et al., 2017]. The corresponding probability density function (PDF) of canopy cover is shown in Figure 4.8(b). Each pixel in Figure 4.8(a) depicts the predicted hectares of oak canopy cover per kilometre squared of land, ha/km^2, represented by colour. Therefore, each measure of abundance correlates to host density (justified by the fact that 100ha = 1km^2), denoted by[9] ρ. Therefore, the spatial map in Figure 4.8(a) displays irregularities and population heterogeneity across GB visually.

The PDF displayed in Figure 4.8(b) reveals that most land patches occupy lower canopy cover values. Moreover, a small number ($\sim 5\%$) of high abundance values in the interval $\rho \in [10, 35]$ ha/km^2 skewed the distribution. These high-valued density patches were reduced to 10ha/km^2, thereby capping the highest value to 10% of oak canopy cover. The PDF describes a continuous distribution, which contrasts the binary-valued distribution of hosts within the SLM. Accordingly, we introduce a threshold function to navigate the problem:

$$\phi(\rho) = \begin{cases} 1 & \rho_{i,j} \geq \rho \\ \\ 0 & \rho_{i,j} < \rho \end{cases} \tag{4.6}$$

where the canopy cover $\rho \in [0, 10]$ ha/km^2 acts as a threshold value, above which, all patches become susceptible and assumes the numerical value of unity (i.e. $\rho < \rho_{i,j} = 1$). Insusceptible states are described when $\rho \geq \rho_{i,j} = 0$. The inset of Figure 4.8(b) visually depicts the abundance threshold function. Here, the vertical black line is an arbitrary threshold ρ, and all canopy cover values less and greater than ρ describe insusceptible and susceptible, respectively. In this interpretation, susceptible patches have enough hosts to support disease survival, growth, and spread. Conversely, patches of land below the abundance threshold are presumed insusceptible because of insufficient hosts.

---

[9] Although the spatial scale is larger, ρ still represents the same underlying tree density for any given 1km × 1km patch. Thus, the same notation for density ρ is kept the same as Chapter 3

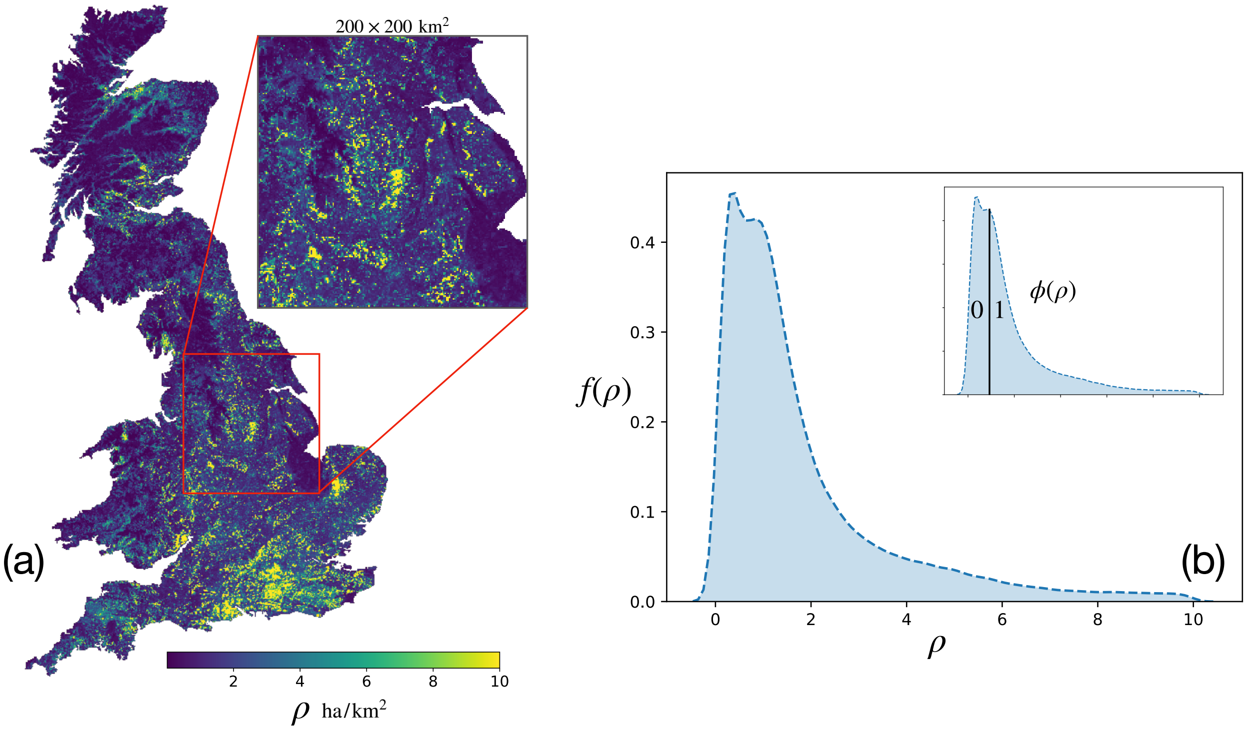


Figure 4.8: (a) The abundance distribution of common oak (*Quercus robur*) given by [Hill et al., 2017]. Each pixel describes a predicted value of abundance with units hectares of canopy cover per kilometer squared of land (ha/km^2). (b) The probability density function of oak abundance in GB, $f(\rho)$ ha/km^2. The zoomed inset illustrates the process of generating threshold function ϕ.

Earlier in Chapter 3, hosts distributions were easily characterised by $\rho \in [0, 1]$, describing a uniform density throughout a square lattice. Now, however, host heterogeneity prevents a simple description of density. Nevertheless, by considering the percentage of susceptible patches, one can define an effective landscape density ρ^*:

$$\rho^* = \frac{\sum^{i,j}(\rho_{i,j} \geq \rho)}{|\mathcal{L}_{GB}|} \tag{4.7}$$

where $\mathcal{L}_{GB}$ represents host distribution over Great Britain. The terms $\sum^{i,j}(\rho_{i,j} \geq \rho)$ and $|\mathcal{L}_{GB}|$ represent the total number of susceptible patches and total landmass respectively. Given an increase in the threshold ρ, the effective density ρ^* decreases; likewise, a decrease in ρ increases ρ^* as more patches become susceptible. Thus, ρ^* presents a convenient, although crude, agent to adjust the host distribution to higher or lower densities.

Following the same arguments in Chapter 3, the landscape-level percolation model becomes an appropriate description of disease spread in tree populations if long-distance dispersal events (over $\geq$ 1km) are omitted. Thus, the model becomes helpful when either the frequency of long-distance dispersal is low or the pathogen disperses through small scales. For example, smaller dispersal distances result for fungal spores in low wind speeds [Viljanen-Rollinson et al., 2007, Solheim and Hietala, 2017] or when the pathogen spreads through mycelial growth in soil, i.e. *R. Solani* and *Pythium spp* [Otten et al., 2004, Poggi et al., 2013].

4.3.3 Epidemics in heterogeneous landscapes

Applying the effective density (ρ^*) of Equation 4.7, we can initialise a set of binary-valued SLM heterogeneous domains. Figure 4.9 shows three variations of effective density $\rho^* \in [0.40, 0.50, 0.60]$, alongside the corresponding thresholds of abundance canopy cover shown below. In Figure 4.9, the SLM is simulated with infectiviy $\beta = 0.25$, until pathogen extinction, shown through through four time-steps. Between panels (a) (e) and (i), the differences in the domain density are visible, as larger values of abundance thresholds produce lower density maps. For the three simulations shown in Figure 4.9, epicentres were placed in the south, where the canopy cover is most dense.

Previously, density was uniform in all directions, but now heterogeneity unevenly distributes susceptible hosts. Notwithstanding, we may still expect an epidemic to emerge if density and infectivity parameters satisfy a critical threshold, as explored in Chapter 3. For all ρ^* values shown in Figure 4.10, initial outbreaks ($0 < t < 250$) spread above the threshold. Although at $\rho^* = 0.40$, we notice a significant drop in disease progression beyond $t = 250$ in (c-d), approximately extending from Oxford to Buckinghamshire due to a low density; in contrast, $\rho^* = 0.50$ and $\rho^* = 0.60$ spread above the threshold for all panels.

In Figure 4.9 we can observe a spatial dependence in the SLM threshold, where above-threshold regions (e.g. the in south) depend on the host density ρ^* parameter. In particular, Figure 4.9 demonstrates that increasing ρ^* can lead to channels opening between different above-threshold regions, thus permitting disease to invade new regions—i.e.

compare Figure 4.9(d), (h), and (l), where only higher density simulations spread progressively further North.

Unfortunately, the definition of percolation (as per Equation 3.4) becomes obscure in light of the irregular domain configuration shown in Figure 4.10, primarily because the extent between borders is now non-uniform and dependant on epicentre location. Furthermore, simulation time series are subject to additional noise on account of heterogeneity, making velocity-based metrics (as shown in Equation 3.5) hard to use. Subsequently, the examination of disease progression will focus on the mortality ratio χ, as its measurement depends only on the final value of removed trees.

4.3.4 Spatially dependent ensembles

The observations of spatially varying thresholds, as revealed by Figure 4.9, motivates an examination of epidemic impact as a function of epicentre location. As such, we apply the same ensemble averaging method discussed previously (in Figure 4.7) to the oak abundance data set. That is, treating each location (i, j) in GB as a potential epicentre and ensemble averaging host mortality over replicate simulations. Here, the mortality ratio $\langle \overline{\chi} \rangle \in [0, 1]$ describes the ratio of removed host patches to the total number of susceptible patches at $t = 0$ and expresses the final sized epidemic in the SLM.

The result of spatial ensemble averages are shown in Figure 4.10 for three different effective densities and one value of infectivity $\beta = 0.25$. The values of effective density in Figure 4.10 are arbitrary, but crucially exhibit the heterogeneous SLM behaviour. Unsurprisingly, increasing the effective density yields a higher mortality ratio (as defined by Equation 4.5). In turn, a higher mortality increases the magnitude of colour bars in Figures 4.10(a-c). Alongside a more severe mortality ratio, a higher ρ^* value also permits disease propagation over more extensive regions, witnessed by comparing Figures 4.10(a) and (c).

In Figures 4.10(a-c), yellow regions highlight where the pathogen is most likely to spread through susceptible hosts. In the toy model, mortality is approximately independent

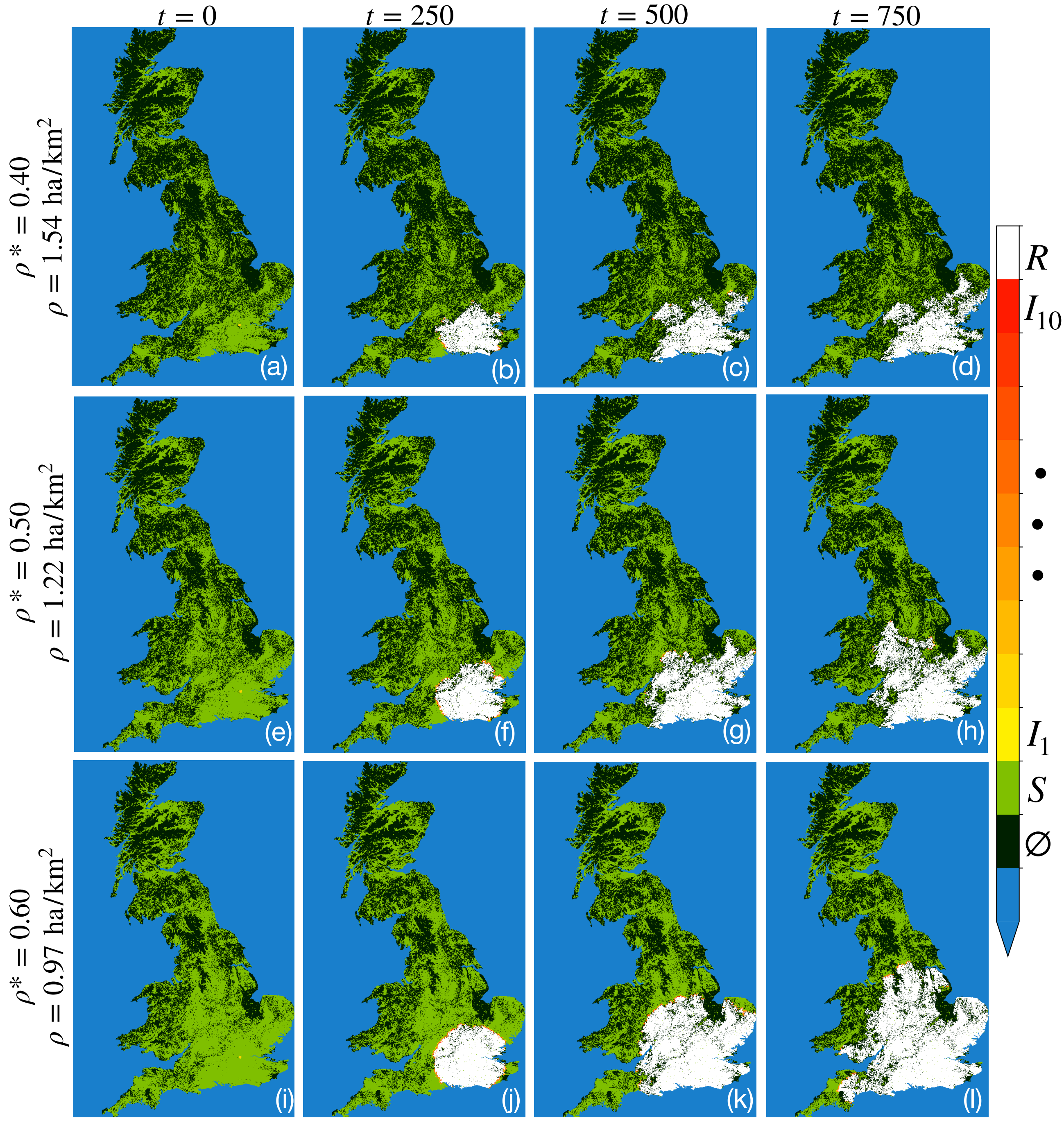


Figure 4.9: The simple lattice model running on a binary-valued Oak domain with infectivity $\beta = 0.25$ for three variations of effective density ρ^*.

of epicentre location, provided sufficient (Von Neumann) connectivity between patches, supported by uniform yellow contours within susceptible regions. In panel (a), the spatial locations encircled in dashed red highlight a region of instability that appears to separate two susceptible regions. In these regions, the epidemic impact could vary significantly, consequently prompting the corresponding plots of mortality variance in Figures 4.10(d-f).

Mapping the ensemble averaged mortality variance reveals uncertainty, where regions support the possibility of both epidemic and extinction. Figure 4.10(d) captures a region of uncertainty, highlighted in red. In this region, the toy model may or may not give rise to a large scale epidemic, as opposed to the most southerly, high-mortality region beneath the dashed red lines. Panels (e) and (f) show a variance only in the edges of the centrally located susceptible region. It is alluring to consider the implications of high variance regions in Figures 4.10(d-f) in the context of epidemic control. Namely, we may suppose that epidemic control through high variance regions could be an effective strategy in stopping the spread of disease. Although the mortality ratio (χ) categorises the overall epidemic scale in the SLM, it fails to reflect any information about how far an epidemic is likely to propagate. Recording the maximum distance travelled by the pathogen proved a helpful method to illustrate spatial progression, maps of ‘maximum distance’ are displayed in appendix A.2.1.

4.3.5 Heterogeneous parameter sweeps

Thus far, spatial ensemble analysis rests on a fixed infectivity ($\beta = 0.25$) and has focused on mapping mortality with a varying effective density parameter ρ^*. In this section, we investigate the entire parameter space of ρ^* and β. Figure 4.11 depicts a full parameter sweep of the toy landscape SLM. Due to more host units in computer memory, simulating the spread of disease in the toy landscape SLM is more computationally challenging when compared to the ideal square SLM in Chapter 3.3. Even though epidemics in the toy SLM are conditioned on epicentre location (confirmed in section 4.3.4), analysing a single epicentre is sufficient to capture the essential toy model behaviour. Ergo, we

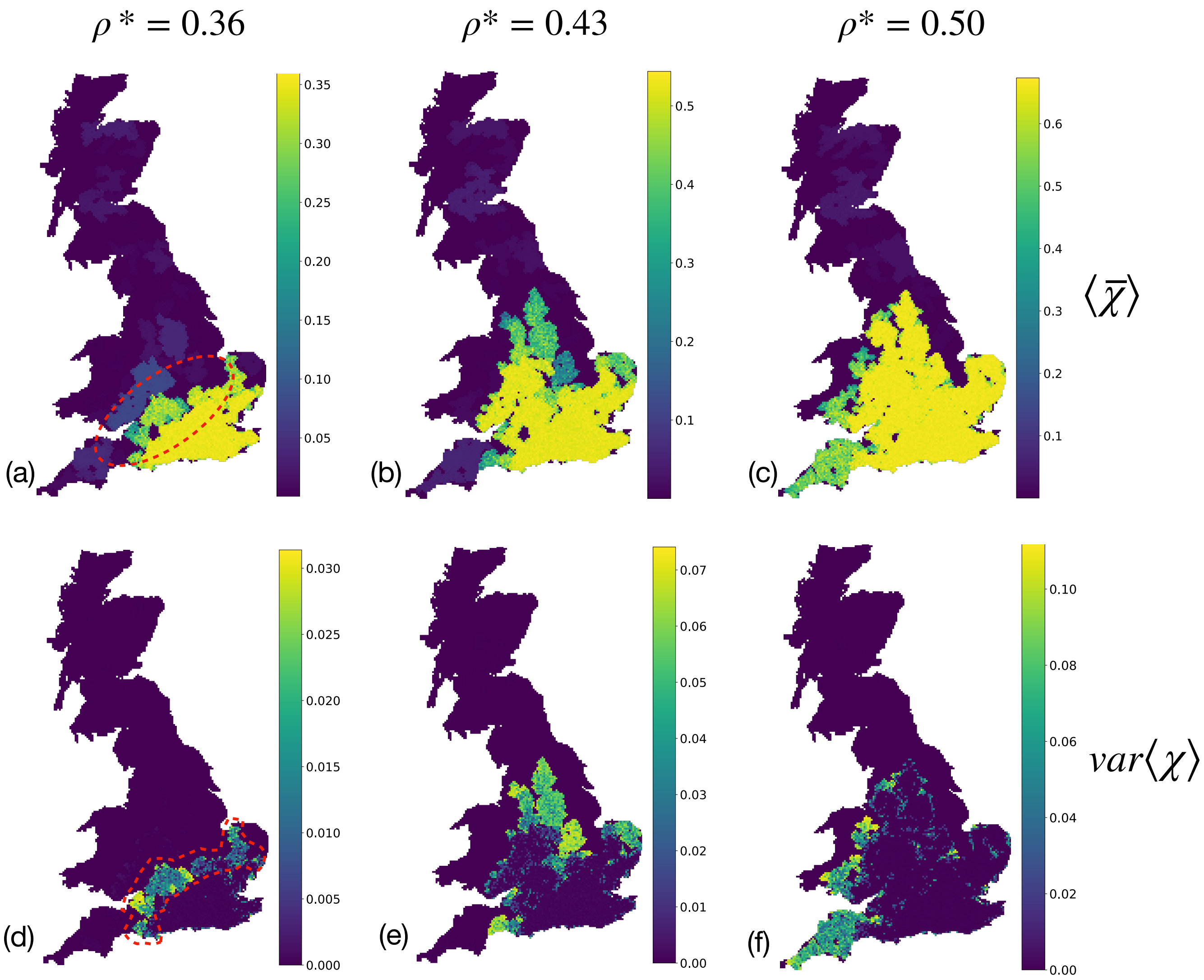


Figure 4.10: Spatial phase showing ensemble statistics over the oak data-set for three variations of density threshold $\phi(\rho) : \rho \in [0.37, 0.43, 0.50]$ and fixed infectivity $\beta = 0.25$. (a-c) The ensemble mean of mortality ratio χ measured for each pixel epicenter. The dotted red circle in (a) shows two neighbouring susceptible regions. (d-f) Ensemble variance over χ. The dotted shape in (d) highlights an unstable region of high variance and uncertainty separating two susceptible areas of the population in (a).

present an analysis through a single epicentre.

Figure 4.11 shows the ensemble-averaged epidemic phase space of the toy landscape SLM through a epicentre—indicated by the red point in Figure 4.11(a). The Parameter sweeps of ρ^* and β demonstrate multiple discontinuities and sharp increases in χ for particular combinations of ρ^* and β; this contrast with the parameter sweeps inside a homogeneous square domain. Moreover, Figure 4.11(b) reveals a large asymmetry between ρ^* and β as more discontinuous jumps appear most when ρ^* is increased, i.e. moving horizontally through Figure 4.11(b). Hence, host heterogeneity gives rise to distinct behaviours for both ρ^* and β axes, as opposed to the (approximately) symmetric ensembles in a homogeneous square domain.

Figures 4.11(c-d) contrast behavioural differences between ρ^* and β axes. Specifically, we compare one-dimensional slices through both parameters ρ^* and β. Figure 4.11(d) details how variations of ρ^* effect the model behaviour through β-space. Interestingly, Figure 4.11(d) depicts the same infectivity threshold of $\beta \sim 0.10$, identical to the SLM evolving on a uniform square domain. When β increases, fewer discontinuities arise when compared to ρ^*, as evidenced by smoother curves.

For each value of density in Figure 4.11(d), the mortality remains fixed beyond $\beta \sim 0.30$. We can understand the independence between χ and infectivity through a numerical example: the probability of a susceptible patch remaining susceptible when it encounters an infected neighbour is given by Equation 3.3 as $Pr(S \rightarrow S) = (1 - 0.30)^{10} = 0.03$. Therefore, on average the pathogen transmits successfully to susceptible neighbours with probability $Pr(S \rightarrow I) = 0.97$, e.g. if a particular epicentre belongs to a susceptible region containing 100 patches, only three patches remain susceptible. In this instance, most patches in the cluster become infected, and further increases in β do not affect the mortality[10]. When $\beta \sim 0.30$, only increases to the domain density have the potential to raise the final epidemic size, indicated by the increases in the height of the curves in Figure 4.11(d).

---

[10] Increasing the infectivity to $\beta = 0.40$ yields a $Pr(S \rightarrow S) = 0.006$, leading to negligible changes in the final epidemic size—however, the rate of progression is still faster.

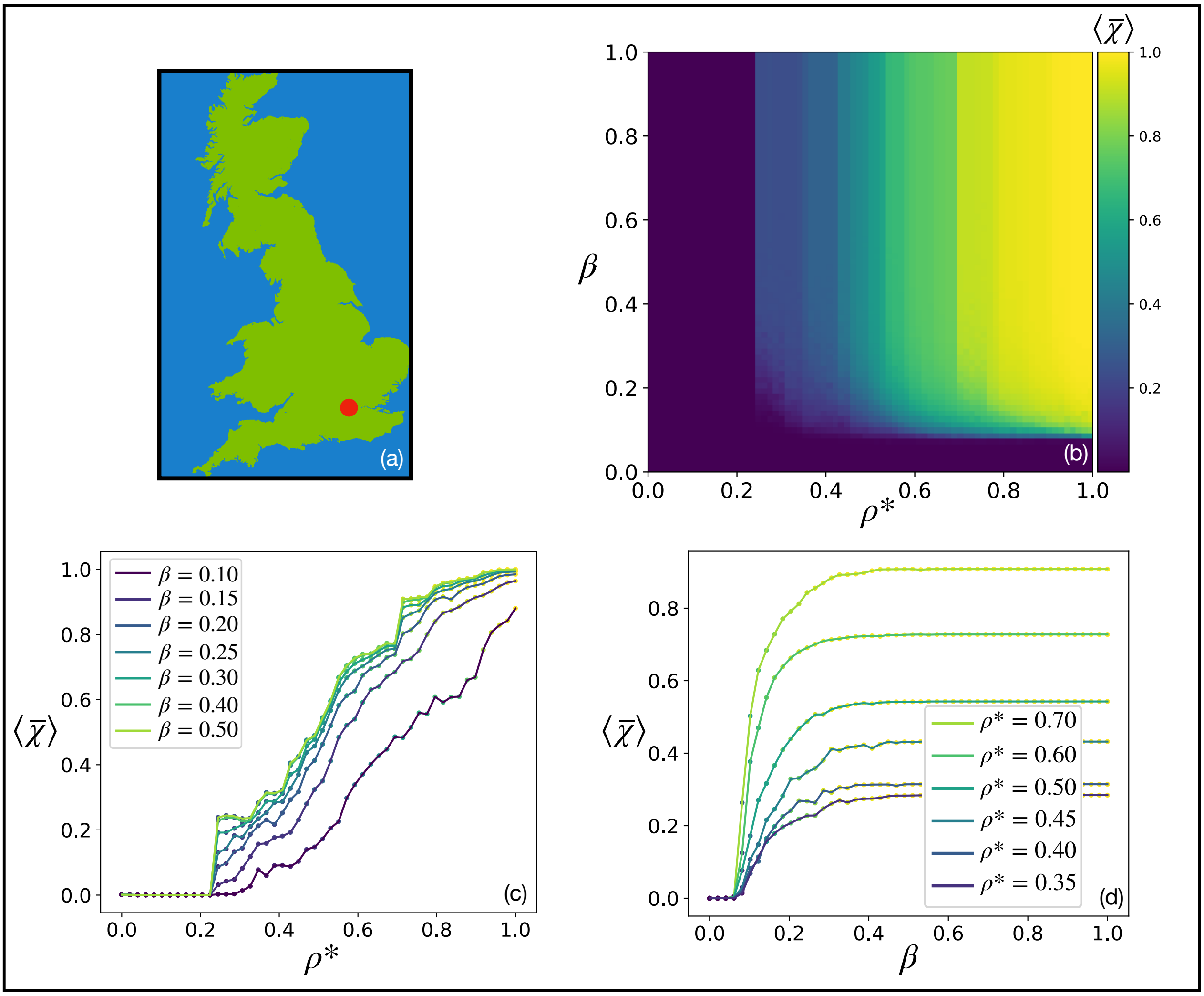


Figure 4.11: Ensemble-average parameter sweeps of the toy model display comparable, albeit noisier, epidemic thresholds compared to the SLM. (a) parameter sweeps were performed over replicate simulations beginning from the red dot. (b) The mortality ratio plotted over the two-dimensional parameter space of ρ^* and β (c) A one-dimensional plot of mortality as a function of host density is shown alongside several slices of infectivity, indicated by colour. (d) The mortality ratio is found over infectivity β for different values of effective density ρ^*.

4.4 Discussion

In this Chapter, we explored two applications of the SLM viz; early warning signals and a toy landscape-level model. The original analysis conducted by [Orozco-Fuentes et al., 2019] relied on a velocity metric based on the number of infected and removed trees, N_I and N_R respectively. In this scheme, the number of infected and removed hosts scales as $\propto (N_I + N_R)^2$ above the threshold, giving rise to an effective increase in velocity for later times; as we discussed, these undesirable artefacts of domain geometry lead to confusing increases in the time series velocity, despite a constant rate of progression. Therefore, EWS were detected using an alternate (COM) time-series metric and abstract cylindrical domain configuration that negated geometrical effects. A two-dimensional investigation was undertaken, sweeping the entire parameter space of tree density ρ and infectivity β.

After setting up the EWS framework, the two-dimensional parameter sweep revealed that preemptive EWS detection is more obtainable when infectivity is lower. Observing these asymmetries in EWS detection, conditioned on infectivity β, highlights the possible challenge of preempting progressively infectious pathogens; in particular, given that host susceptibility is likely to increase as a consequence of climatic stressors [Garrett et al., 2006]. Subsequently, we may hypothesise the heightened challenge of early warning indicator detection for forest-based pathosystems in the face of climate change.

EWS have found applications in a variety of ecological processes, e.g. aquatic ecosystem function [Kramer and Botterweg, 1991], forest desertifications [Yang et al., 2005] and species-level extinctions [Drake and Griffen, 2010]. Nevertheless, few sources focus explicitly on EWS from tree epidemics akin to the dynamic (velocity-based) approach used by [Orozco-Fuentes et al., 2019]. Instead, most research has focused on the more general class of forest health and tree mortality[11] based on tree growth rings [Rogers et al., 2018, Mamet et al., 2015]. As such, the EWS method presented in section 4.1 differs from the wider literature, and more work is required to scrutinise the utility of dynamic, velocity-based, EWS detection methods.

---

[11] See [Torres et al., 2021] for a related review on remote sensing technologies and forest-health

Secondly, we constructed a toy landscape-level SLM spreading through an example distribution oak, as generated by [Hill et al., 2017]. The model of landscape-level epidemics neglects several essential features of invasive disease; most importantly, it inherited the nearest neighbour contact assumption, as discussed in Chapter 3. Hence, we labelled the landscape-level interpretation a 'toy' model. Although many limitations underpin the toy model (e.g. the omission of long-distance dispersal [Rieux et al., 2014] and cryptic infections [Gilligan et al., 2007b]), it highlighted the inability of the SLM to describe the spread of disease through lower tree densities ($\rho \in [0.01, 0.10]$), typical throughout GB.

As the SLM could not describe the spread of disease through more realistic host densities, we introduced an effective density parameter ρ^*, predicated on an arbitrarily chosen threshold. Introducing an additional density threshold parameter is undesirable, unnatural and speculative. Therefore, we are motivated to change direction and construct a non-local dispersal model in the proceeding Chapter, in line with more contemporary dispersal-based approaches, e.g. [Parnell et al., 2009a, Meentemeyer et al., 2011]. In this setting, transmission between trees can occur over larger length scales and permit the spread over lower tree densities.

Notwithstanding the inherent toy SLM shortcomings, its construction demonstrates the use of a novel predicted abundance dataset provided by [Hill et al., 2017]. The predicted abundance distribution is partly generated from numerous data sets[12], as we reviewed in section 4.2. However, predicted (statistically regressed) abundance data contains uncertainties and inaccuracies alongside the loss of small-scale host spatial structure $< 1\text{km}^2$. As argued by [Cunniffe et al., 2015a], capturing host spatial structure, even when data are limited, is essential, and methods are required to assess the impact of incomplete or inaccurate host data.

Following this argument, host data accuracy presents a notable assumption in the toy model. That being said, density parameter sweeps over GB (as shown in Figure 4.11) could form a simple procedure to assess the effect of host error, i.e. contrasting epidemic

---

[12] Including ancient woodland shapefiles, BSBI distribution database, Countryside Survey data, myForest and the National Forest Inventory Great Britain 2014

outcomes between two upper and lower density error bounds. Evaluating landscape-level parameter-sweeps are uncommon, and most large-scale models repeat simulations over numerous control scenarios and rest on fitted parameters e.g. [Cunniffe et al., 2016, Harwood et al., 2011]. Assessing disease outbreaks over a range of landscape-level density-based parameters could describe a risk-based approach, as articulated by some authors investigating epidemics through smaller spatial scales [Hyatt-Twynam et al., 2017].

Chapter 5

Including dispersal

In the previous Chapter, we employed a simple lattice model (SLM) based on percolation. The percolation setting provided a tractable starting place. Nevertheless, several limitations underpinned the model. Chiefly, percolation models rest on local, nearest-neighbour (NN) contacts and cannot describe epidemics at lower, more realistic tree densities ($\rho \sim 0.10$). Subsequently, this Chapter will generalise the SLM and incorporate non-local NN interactions by introducing a generic Gaussian dispersal kernel, allowing epidemics at far lower tree densities.

The thin-tailed Gaussian kernel introduced here represents an intermediate step between the NN interactions in the SLM and the fat-tailed inverse power-law dispersal present in Chapter 6. The ease of integrating a Gaussian kernel also proved helpful when deriving an analytic expression of R_0, as discussed more below. Other choices of simple one parameter dispersal kernels are possible, including the slightly longer range negative exponential—discussed at length in [Nathan et al., 2012]. Regardless, any developing epidemic that follows a similar ranged kernel will ultimately approximate similar spreading patterns [Bullock et al., 2017].

Firstly, we will combine dispersal-based interactions within a simplistic SIR framework to construct a non-local model (NLM) of tree disease. The model behaviour is then examined under various dispersal length scales and fitted against the standard SIR to contrast disparities between spatial and non-spatial models. After establishing the non-local dispersal model, a spatially explicit (analytic) expression for the basic reproduction number will be derived for the model, denoted by R_0.

Next, we compare analytical predictions of R_0 against the total tree mortality, equivalent to the final-sized epidemic. Lastly, the expression of R_0 is scrutinised against the 'actual' number of secondary infections, computed by contact-traced individual tree-to-tree infections. Notably, the analytic and contact-traced methods of calculating R_0 define a threshold at $R_0 = 1$. As before, the analysis is kept generic, with arbitrary units of time and distance, before incorporating more biological realism in the next Chapter.

5.1 A small-scale non-local SIR model

As before, we begin with a model fixed inside a square lattice of size $\mathcal{L}$ and host units refer to individual trees. Host distributions are initialised by a Bernoulli trial with probability ρ according to a binomial distribution. Thus, the probability of host occupation (ρ) can be seen as a tree-density parameter and interactions between hosts are modelled over a flat and randomly distributed population. The state of a tree can be in one of three conditions: susceptible, infected, or removed (SIR). We assume all trees are equally susceptible, and trees that become infected transition through the states $S \to I \to R$ without the possibility of recovery.

From first principles, the probability of infection at a distance r can be described by an unnormalised Gaussian function $g(r;\ell)$, where ℓ is a distance that sets the scale of dispersal. If two trees—one susceptible (S_x) and one infected ($I_{x'}$)—are separated by a distance $r = |x - x'|$, then a transition probability between the states $S_x \to I_x$ can be defined by $g(r;\ell)$ multiplied by infectivity β:

$$Pr(S_x \to I_x;\ I_{x'}) = \beta \exp\left[\frac{-r^2}{2\ell^2}\right] \tag{5.1}$$

where β is interpreted as a probability, i.e. $\beta \in [0, 1]$. Equation 5.1 generalises the SLM to include non-local interactions, hence referred to as the non-local model (NLM). Each probability of transition is assessed against a sample drawn from a continuous uniform distribution $U(0, 1)$ following a Poisson construction [Cook et al., 2008]. Probabilities are then calculated for each times step while host S_x remains susceptible, and repeated for each susceptible tree in the domain (i.e. $\forall S_x \in [\mathcal{L}, \mathcal{L}]$). See Appendix B.3 for more information on the computational implementation. A table of parameters for the NLM is shown below in Table 5.1.

The same uniform lifetime dynamic (used previously in the SLM) controls the period hosts remain infectious. That is, a host transitioning into the I compartment will remain infectious for T time steps before uniformly transitioning into the R compartment. In section 3.3, the infection period was shown to alter the wave-front thickness and the threshold value of infectivity β required for an epidemic. However, an arbitrary number of $T = 100$ infectious time steps remains fixed throughout this Chapter. Uniform transitions into the R compartment help to keep the model simple but present an assumption that goes against the grain of more common exponential lifetime dynamics—discussed more below in section 5.2.2.

The work presented in this Chapter is purposefully kept generic, with no specific pathogen in mind. Therefore, each Monte Carlo step through the simulation has arbitrary units of time and distance. Nevertheless, the units of time and distance can be envisioned to be on the order of days and meters to reflect the approximate spatio-temporal scale of a general tree pathogen. As demonstrated later in Chapter 6, spatial scale within the model can be calibrated by choosing a suitable lattice constant, denoted by α, that reflects the size of host units.

Model parameter	Description	Typical value(s)
ρ	Tree density	$0.00 - 0.10$
β	Infectivity probability	$0 - 10^{-3}$
β^*	Auxiliary infectivity	$0 - 10$
ℓ	Gaussian dispersal parameter	$0 - 100$
t	Simulation time step	1 Au
T	Infectious life time	100
α	Lattice constant	1 Au
$\mathcal{L}$	Square lattice dimension	200 - 2000
R_0	Basic reproduction number	$0 - 20$
$R_0^{(i)}$	Generational reproduction number	$0 - 20$

Table 5.1: Parameters used in the generic NLM, time and distance are given in arbitrary units and host densities are informed from by [Hill et al., 2017].

5.2 Model behaviour

Spatio-temporal epidemic progression within the NLM is depicted in Figure 5.1. Figures 5.1(a-b) depict an epidemic spreading through the domain at two time steps; simulation parameters are given by $\rho = 0.01,\ \ell = 25,\ \beta = 1.0 \times 10^{-3}$ on a domain of size 500×500. All panels in Figure 5.1 begin from a small number of infected hosts at the domain centre at $t = 0$. A tree density of $\rho = 0.01$ approximately mirrors the median canopy coverage of a large deciduous tree distributed throughout the GB, according to the predicted oak abundance data given by [Hill et al., 2017]—presented previously in Figure 4.8. Unsurprisingly, extending the neighbourhood of interaction to non-nearest neighbours permits an epidemic for much lower tree densities in comparison to the SLM percolation threshold studied in Chapter 3.

Figures 5.1(a-b) suggest an approximate wave-front-like behaviour, as infections spread out radially from the epicentre. The corresponding ensemble-averages of Figures 5.1(a-b) are shown below in Figures 5.1(c-d), and confirms a travelling wave-like spread. For 200 repeated simulations, the spatial locations of infected trees were recorded and plotted as a two-dimensional frequency distribution. The upper and lower marginal plots of Figures 5.1(c-d) show the one dimensional horizontal and vertical frequency distributions, respectively. Disease progression in Figure 5.1(d) reflects the radial propagation of a travelling wave and a disease gradient of approximately $\approx 3\ell$. Thus, choosing a small value of $\ell = 25$ in comparison to the domain effectively recovers the essential wave-like behaviour exhibited by the SLM.

The thin-tailed Gaussian kernel does not permit the pathogen to jump large discontinuous distances, particularly for a small value of $\ell = 25$, as shown in Figure 5.1. Therefore, we can present an analogy to percolation provided that the ratio $\frac{\ell}{\mathcal{L}}$ is small, allowing us to calculate a wavefront similarly to the method described previously in Chapter 3. Accordingly, Figure 5.1(e) reveals the largest distance an infected host will likely reach over 500 steps. Boundary conditions in Figure 5.1(e) terminate simulations upon three conditions: (A) The simulation time step exceeds 500 steps (B) No infected trees remain in the domain (C) An infected tree I_x falls within a distance $\mathcal{L} - 3\sigma_{ga} \leq I_x \leq \mathcal{L}$ away from the epicentre.

After t time steps, the maximum distance reached by the pathogen is $D_{max}(t)$. Figure 5.1(e) shows the ensemble-averaged maximum infectious distance for four infectivity parameters, along with a 95% confidence interval about the mean for each time step. The time-series data shown in Figure 5.1(e) indicates whether or not the pathogen dies off or survives to the domain boundary. Furthermore, scatter plots in the upper marginal of Figure 5.1(e) mark when the pathogen arrives at the domain boundary. A cluster of infectious-removed trees spans the domain whenever the pathogen survives long enough to propagate to the edge.

Unsurprisingly, Figure 5.1(e) reveals that, on average, pathogens with higher infectivities propagate to the domain boundary quickest, illustrated by comparing the first (red) and second (green) highest infectivity line plots. In contrast, the lowest two infectivity values (blue and orange time series, respectively) fail to reach the domain boundary in most

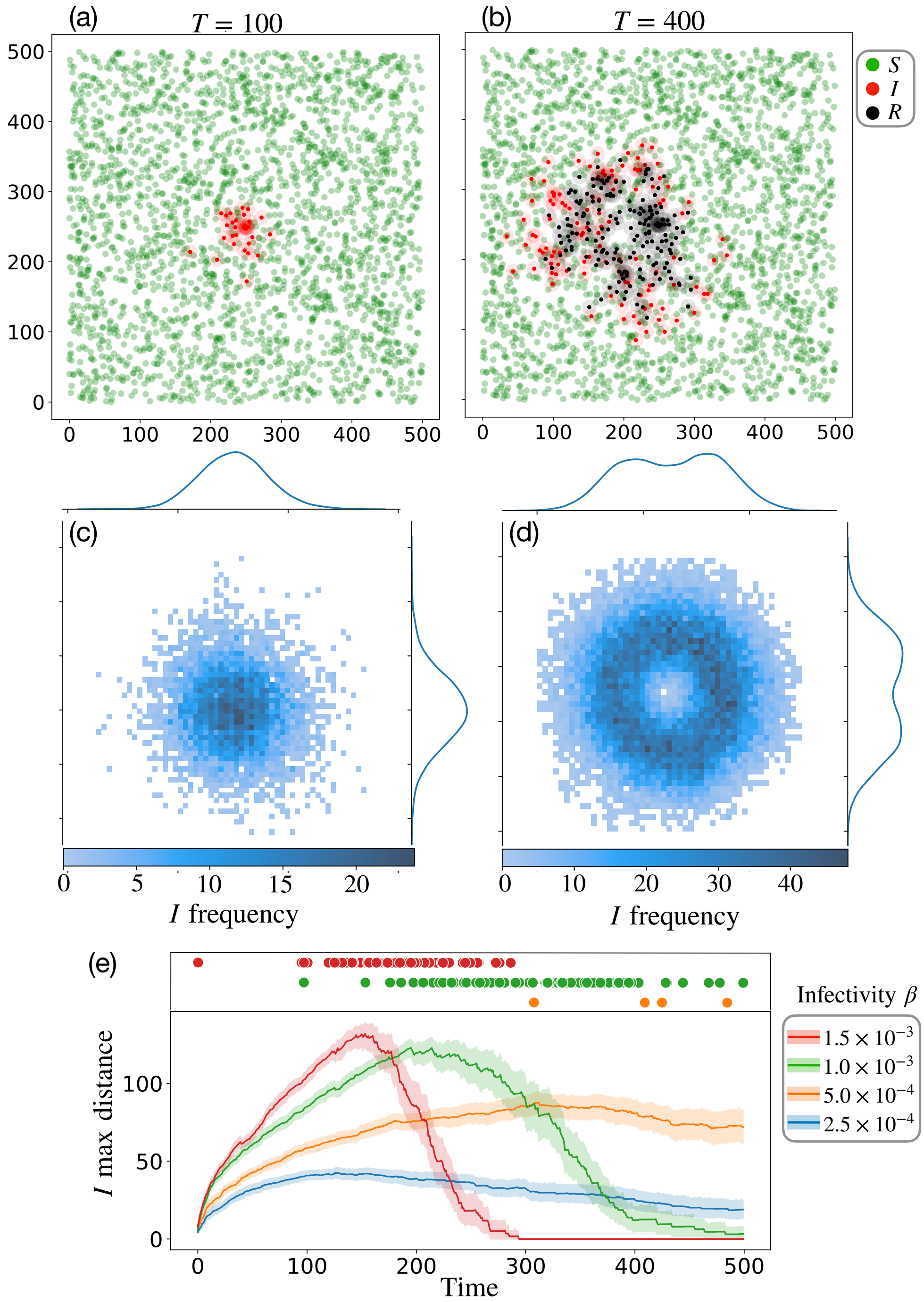


Figure 5.1: Percolation-like disease progression of the dispersal-based SIR model with a small dispersal length scale of $\ell = 25$. All figures were accessed in a domain of size $\mathcal{L} \times \mathcal{L} = 500 \times 500$ and fixed host density $\rho = 0.01$. (a-b) An evolving epidemic with infectivity $\beta = 1.0 \times 10^{-3}$ is shown through two time-steps. Green pixels represent susceptible trees in S, while red pixels represent infected trees at different steps in the I category. Infected trees uniformly transition into the removed compartment at $t = 100$, shown in black. (c-d) An ensemble-averaged spatio-temporal frequency distribution (of 200 repeats) representing the number of trees in the infected compartment. The probability of an infected tree located at row x and column y is represented by a kernel density estimate (ascertained via Seaborn, a statistical/plotting package in Python) in the upper and horizontal marginals. (d) The maximum infectious distance ensemble-averaged over 100 repeats for four infectivity parameters shown by colour.

simulations, notwithstanding the small number of orange scatter points shown in the upper marginal.

Modelling the spread of disease for small $\frac{\ell}{\mathcal{L}}$ is helpful to understand the NLM and presents a clear connection to the percolation-based SLM. However, realistic dispersal-based tree disease is unlikely to exhibit slow marching travelling waves at these small spatial-scales. In particular, as $\frac{\ell}{\mathcal{L}}$ becomes larger, tracking the maximum distance inside a finite domain of size L becomes ill-defined; undoubtedly this becomes even more relevant with fat-tailed dispersal kernels. Therefore, although the time-series $D_{max}(t)$ is justified for pathogen progression when $\frac{\ell}{\mathcal{L}}$ is small, an alternative metric is required to understand epidemic progression as we look to increase the scale of dispersal. Consequently, the reproductive ratio R_0 is introduced later in section 5.3.

5.2.1 Normalising infectivity β

Before we move onward, a tool to fix epidemic impact over various dispersal length scales is outlined below. Using β to control infectivity in the model, as governed by equation 5.1, is pragmatic for single parameter of ℓ—as shown in Figure 5.1. However, suppose the dispersal parameter ℓ in equation 5.1 is increased. Undoubtedly, a larger area under the unnormalised kernel, $\exp\big[-r^2/(2\ell^2)\big]$, would produce more secondary infections. In turn, more secondary infections produced by each infected tree affords a more severe epidemic. In this setup, infectivity, as defined in Equation 5.1, depends strongly on the scale of dispersal ℓ and epidemic-impact would vary significantly under different dispersal parameters. Ideally, the infectivity (i.e the strength of interaction between trees) should not depend strongly on the dispersal kernel because this makes model comparisons over different length scales ℓ difficult. This motivates an updated scheme.

At the very least, β could be manually varied to match the approximate epidemic-impact between different ℓ valued simulations[1]. However, scaling β differently for each ℓ pa-

---

[1] One may suggest that directly normalising the kernel poses a solution to fix the epidemic scale for all values of ℓ. Although, this is ultimately incorrect from an implementation perspective. If the Gaussian kernel in 5.1 is normalised, it will cease to be a yield a probability describing individual tree-to-tree interactions and the transition between states.

rameter is cumbersome and ultimately untenable for many simulations. A mathematical sleight-of-hand can resolve the dilemma by simply factoring out the dispersal normalisation from β:

$$\beta = \frac{\beta^*}{2\pi\ell^2} \tag{5.2}$$

where β^* is an 'auxiliary' infectivity parameter that isolates infection pressure to a single parameter that remains fixed between simulations with different ℓ values. In this manner, infectivity and the dispersal remain probabilities (i.e. $\beta = \beta^*/2\pi\ell^2 \in [0, 1]$, and $\exp\big[-r^2/(2\ell^2)\big] \in [0, 1]$ respectively), and epidemic severity will be matched between different ℓ-valued simulations. Some limitations of this method are discussed below in section 5.3. So, henceforth, unless otherwise stated, the remainder of this Chapter will employ the normalised infectivity. That is, the right-hand side of equation 5.2 will be substituted into equation 5.1 (and the analytic expression of R_0 outlined below) to permit model comparisons over the parameter space of ℓ.

5.2.2 SIR fitting: dispersal-mediated contact-mixing

This section aims to shed light on whether or not the spatial NLM can recover a non-spatial process and if so, answer which parameter regime is required approximate the non-spatial process. Understanding when to include spatial dynamics becomes particularly important given the increased computational cost of spatial simulations. In a nutshell, why bother to include spatial dynamics if we could use a simpler non-spatial model? Consequently, we will fit the SIR model given by [Kermack and McKendrick, 1927] to simulated data from the NLM. Notwithstanding the spatially-structured host distribution, it makes sense to compare the NLM with predictions from the SIR framework given the same compartmental transitions $S \rightarrow I \rightarrow R$.

Two parameters control the epidemic evolution in the SIR model, an infectivity rate β and a removal rate γ. Both rates β and γ could be manually varied to match data from NLM simulations, though the task can be simplified significantly by considering I as a function of S, accomplished as follows:

$$\frac{dI}{dS} = \frac{dI/dt}{dS/dt} = \frac{\beta IS/N - \gamma I}{-\beta IS/N} = -1 + \frac{\gamma}{\beta}\frac{N}{S}$$

Letting $\alpha = \frac{\gamma}{\beta}$ we have:

$$dI = \Big(-1 + \frac{\alpha N}{S} \Big) dS$$

Now using integration by separation of variables:

$$I = -S + \alpha N \ln(S) + C \tag{5.3}$$

where C is a constant of integration. Thus, we have reduced the task of trying to match two parameters to considering a single one, namely α. Before we can fit the NLM to equation 5.3 we need to determine C. Initially $S_0 + I_0 + R_0 = N$ and $R_0 = 0$ i.e. R_0 is the number of removed at $t = 0$, not the reproductive ratio. Thus, evaluating equation 5.3 at $t = 0$ gives:

$$I_0 + S_0 = N\alpha \ln(S_0) + C$$
$$\implies C = N - N\alpha \ln(S_0)$$

Upon substitution back into equation 5.3 we have:

$$I = -S + N\big(1 + \alpha \ln(S/S_0)\big) \tag{5.4}$$

more information on the behaviour of equation 5.3 is given in the appendix B.2. By fixing the initial conditions in the differential SIR equations to match the NLM simulations—i.e. with $I_0 = 1$ and $S_0 = \rho\mathcal{L}^2$ —we can compare the NLM to the SIR model.

The canonical SIR model is non-spatial and rests on a 'well-mixed' population assumption, as described in section 2.2. In a well-mixed population, each individual is equally likely to make contact (or pass on the infection) with any other individual. Conversely, a spatially-structured dispersal-based model will generally not describe well-mixed contacts between individuals in the population because the infection probability decreases with distance—see [Cook et al., 2008].

Nevertheless, we can expect an approximation to contact mixing when the scale of dispersal, set by ℓ, becomes comparable to the domain. In this scenario, secondary infections are prone to disperse more uniformly over larger areas, in stark contrast to the localised (wave-like) transmission previously witnessed when ℓ is small. Hence, the comparisons below pay attention to the interplay of parameters ℓ and $\mathcal{L}$. Accordingly, Figure 5.2 shows the NLM fitted to Equation 5.3 for two domain sizes and two dispersal parameters.

Figure 5.2 contrasts SIR and NLM models. Ensemble-averaged data from the NLM is plotted in black and is shown alongside the 25 individual simulations in light blue. NLM simulations contrast the SIR model plotted in red. Tree density and infectivity in the NLM are fixed to $\rho = 0.01$ and $\beta^* = 4$, respectively. Notably, the parameters were large enough to ensure epidemics were witnessed in all simulations. Using least-squares—specially, the Levenberg-Marquardt algorithm [Moré, 1978] implemented in Python—the ensemble mean was fitted to equation 5.3, shown in red. In all panels, the arrow of time is from right to left, i.e. initially, the number of trees in S starts high and decreases as the number of trees in I rises then falls.

Interestingly, for all but one panel in Figure 5.2, epidemics progress faster in the NLM than predicted by the SIR model—indicated by the NLM having a steeper gradient beginning from $S_0 = \rho\mathcal{L}^2$. One possible cause of disparity between models is due to infectious lifetime dynamics. Exponentially distributed lifetimes are implicit within the SIR model. Whereas, the NLM relies on uniform transitions into the R compartment that understood by examining Figure 5.2(a), i.e. 400 hosts are present in the domain at time $t = 0$ and it takes precisely $t = 100$ steps to elapse before the first transition into R.

On the other hand, trees evolving with SIR dynamics will gradually transition into the R compartment at all time steps according to an exponential distribution. For the same initial conditions, it follows that more infectious trees might be expected in the NLM between times $t \in [0, T]$, leading to more secondary infections that, on average, increase

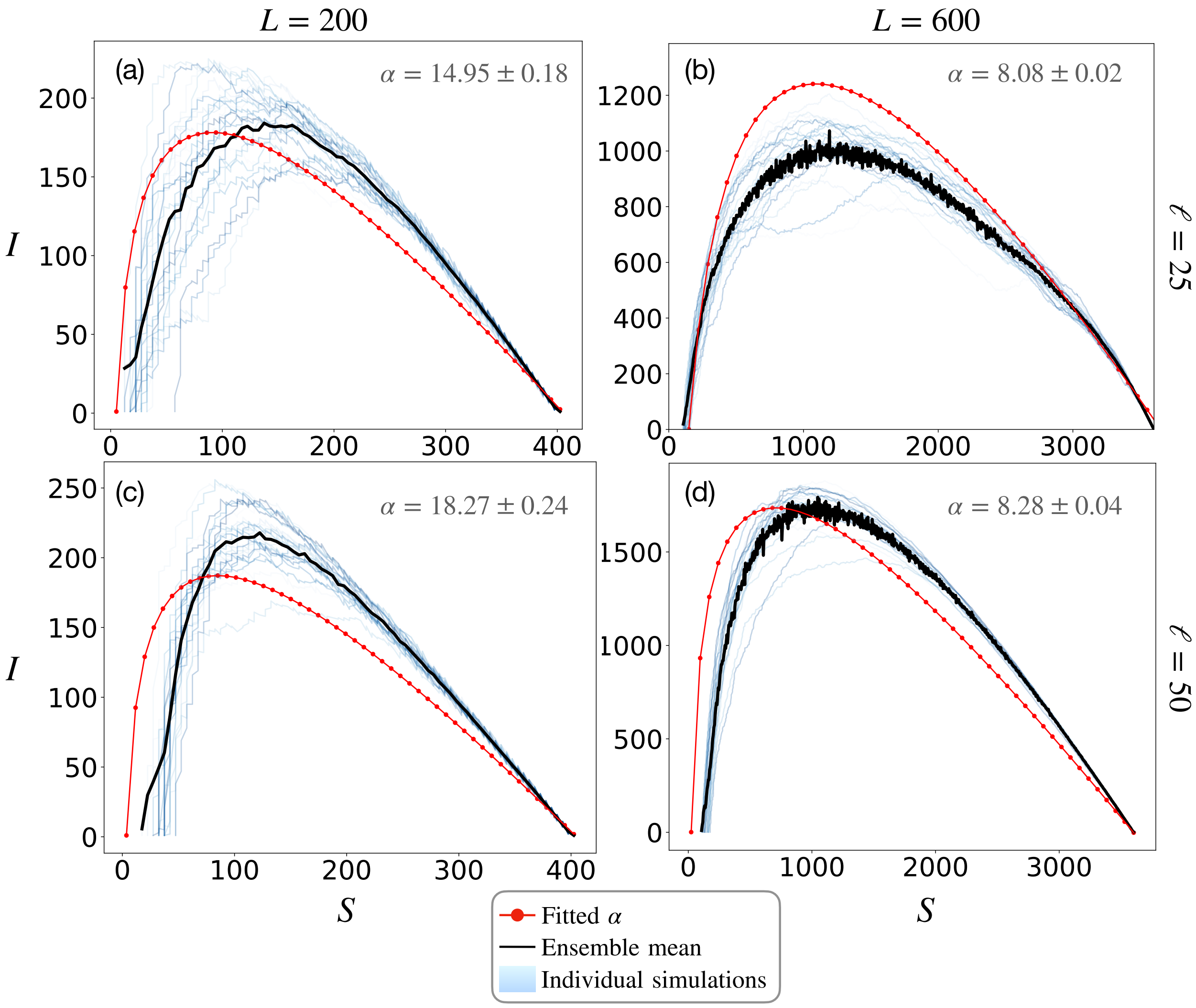


Figure 5.2: Fitting the non-local dispersal model (NLM) to the traditional SIR model given by [Kermack and McKendrick, 1927]. All simulations evolved with parameters $\beta^* = 4$ and $\rho = 0.01$ above the threshold for spread. (a) A small localised dispersal kernel of $\ell = 25$ fitted against the canonical SIR. On a small domain of size 200×200, the NLM spreads faster than the fitted SIR model. (b) On a larger patch of size 600×600, the SIR model predicts a faster rate of spread in comparison to the NLM, illustrated by the disparity between red and black lines. (c) On 200×200 sized domain, increasing the dispersal parameter to $\ell = 50$ results in a similar trend to panel (b), albeit with slightly less agreement between NLM and SIR models. (d) Increasing the dispersal parameter to $\ell = 50$ reduces the large disparity between SIR and NLM, shown in Figure (b).

the scale of epidemic. In appendix B, Figure 5.2 is replicated, although equation 5.3 was fitted to an exponentially distributed variant of the NLM. Consequently, Figure B.2 in appendix B shows, a closer fit to equation 5.3.

Figure 5.2(b) demonstrates another important aspect of NLM behaviour related to contact mixing in the spatial host distribution. Looking at Figure 5.2(b), a faster rate of spread is predicted for the SIR model, demonstrated by the divergence between red and black lines. In this regime, $\frac{\ell}{\mathcal{L}}$ in the NLM is small, and contact-mixing in the host distribution can be assumed low —supported by Figure 5.1 that demonstrated a wave-like spread. Moreover, the disparity between models is reduced by increasing the dispersal parameter to $\ell = 50$, illustrated in Figure 5.2(d). Figure 5.2 therefore indicates that if the system is approximately well-mixed, the NLM spreads comparatively to the SIR, albeit slightly skewed because of uniform lifetime dynamics. Whereas, if the system is not well-mixed, a slower epidemic marches across the domain in a wave-like manner that deviates significantly from the SIR model. Altogether, these results point toward the inability of non-spatial models, such as the SIR model, to describe a spatially structured model of tree disease. That being said, in a parameter regime where the ratio $\ell/\mathcal{L}$ is sufficient for population mixing, the SIR model can describe the system with some accuracy[2].

5.3 A spatially-explicit reproduction number

As remarked earlier, percolation-based distance metrics become ill-defined when the pathogen can jump on long distances and a more robust metric is required to examine the model going forward. As such, a basic reproduction number will be outlined for the NLM. The concept of R_0 is widely used (and widely misinterpreted [Delamater et al., 2019]), and multiple methods of calculation exist in the literature [Heffernan et al., 2005]. Although, to recap, R_0 is fundamental to understand epidemic thresholds in human and animal populations.

---

[2] Well-known results from percolation theory present a simple model analogy to the observations from Figure 5.2. Consider a large domain (of size L) below the percolation threshold, and sub-dividing the domain into boxes of size ξ, where ξ/L is small. Percolating clusters could be observed in each box i.e. at length scales comparable to ξ, but not L see [Stauffer and Aharony, 2018] pages 64-65.

Crop-based reproduction ratios have been examined extensively [Gubbins et al., 2000, Park et al., 2001, van den Bosch and Gilligan, 2008, van den Berg et al., 2011], yet the concept remains less explored in tree-based diseases. In general, R_0 is complicated, and may vary in response to numerous abiotic factors such as temperature, humidity and wind speed. Notably, the threshold $R_0 = 1$ should separate regimes of epidemic and confinement for any definition of R_0. Furthermore, when defining an R_0 value for tree-disease, the importance of spatial structure cannot be ignored [Park et al., 2001].

5.3.1 Approximating R_0 analytically

In this section, an idealised, spatially explicit expression of R_0 is derived for the NLM. Defining an informative R_0-value for tree-based pathosystems is not simple, and care is needed when defining an R_0 value. The following thought experiment outlines an approach to approximate reproduction number:

Consider a single primary infected tree at time $t = 0$, surrounded by a distribution of susceptible neighbouring trees. Throughout its infectious lifetime, the primary infection will lead to R_0 secondary infections. If secondary infections do not produce other tertiary infections, the neighbourhood around the primary infection remains untouched by other diseased trees, and the reproductive potential can be approximated by R_0.

The thought experiment simplifies the epidemic branching process by neglecting tertiary (quaternary, and so on) infections, illustrated by comparing Figures 5.3(a-b). In turn, simplifying the system will help keep the mathematics tractable and permit an analytic derivation of R_0 without advanced mathematics. However, the derivation will be idealised and likely to overestimate the actual reproduction ratio in specific epidemic regimes.

Figures 5.3(c-d) illustrate two hypothetical epidemic systems, with higher and lower R_0 values. When the scale of dispersal is smaller (but still larger than the average distance between trees) and infectivity is high, as in Figure 5.3(c), we expect a coupled system with a large number of secondary infectious induced inside a smaller area. In this case, the R_0 approximation would deviate from model simulations because secondary/tertiary infections

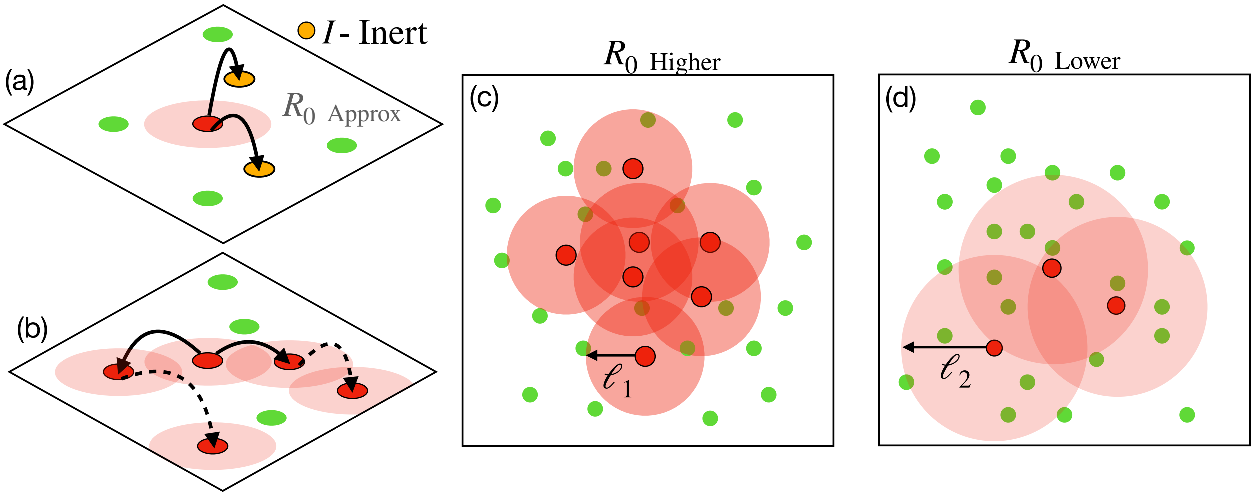


Figure 5.3: Approximating a spatially explicit value of R_0 as a function of ρ, β, ℓ and T. (a) An idealised scenario where secondary infections do not produce tertiary infections, but instead transition into an inert state—shown in amber. (b) The usual epidemic branching process where secondary infections produce tertiary (and so on) infections about the primary infection, shown by the solid and dashed arrows respectively. (c) Depictions of a highly infectious regime in the NLM, where R_0 is large and the dispersal parameter (ℓ_1) is smaller. (d) Illustrations of an alternate, less invasive system when the scale of dispersal (ℓ_2) happens to be larger. The R_0 derivation aims to compute the scenario shown in (a), and becomes accurate in the epidemic regime illustrated in (d).

would reduce the host density about the primary infection. In other words, the finite-sized population gives rise to negative spatial correlations when other infected trees reduce the local density of susceptible hosts. The reader is referred to [Tildesley and Keeling, 2009] for a helpful discussion of R_0, heterogeneity, spatial correlations and finite-size effects.

Conversely, Figure 5.3(d) depicts a scenario when the pathogen induces a lower number of secondary infectious over a larger area. Hence, in the limit where secondary/tertiary infected trees do not influence the local density around the primary infection, the thought experiment (mentioned above) is expected to become increasingly accurate. However, even with a larger dispersal kernel, the approximation illustrated by Figure 5.3(a) would become more inaccurate for highly invasive systems with a large R_0.

For example, incredibly high values $R_0 \in [30, 70]$ have been estimated for wheat stripe rust (WSR) epidemics [Severns et al., 2019, Mikaberidze et al., 2016]. Nevertheless, a field of crops is significantly dense in comparison to an average tree population distributed over the landscape—discussed extensively in section 4.4 concerning oak. Consequently, we argue that, in general, the reproduction ratio of a tree pathogen is unlikely to reach super epidemic regimes that parallel WSR epidemics; supported further by estimates of $1 \lessapprox R_0 \lessapprox 6$ for the Dutch elm disease epidemic in GB [Swinton and Gilligan, 1996a] and $1 \lessapprox R_0 \lessapprox 4$ for oak processionary moth epidemics in London [Wadkin et al., 2022].

5.3.2 Derivation

Suppose the domain is uniform with density ρ_0 at time $t = 0$, and one singular infected tree exists inside a large domain. In such a configuration, the mean number of secondary infections expected over the first time-step can be calculated by integrating equation 5.1 as follows:

$$R_0(t = 0) = \beta\rho_0 \int_{-\infty}^{\infty} \exp\Big(-\frac{r^2}{2\ell^2}\Big) dr = 2\pi\beta\rho_0\ell^2 \tag{5.5}$$

If host (re-)growth is neglected, less trees are available to infect at time-step $t + 1$. Tree density can therefore be seen as a monotonically decreasing function of time $\rho(t)$, leading to the expression:

$$R_0(t) = 2\pi\beta\ell^2\rho(t) \tag{5.6}$$

this expression presents a convenient relationship between tree density and the number of expected secondary infections. In a large but finite domain, of size $\mathcal{L}$, tree density approximately follows:

$$\frac{d\rho}{dt} = -\frac{R_0(t)}{\mathcal{L}^2} = -\frac{2\pi\beta\ell^2\rho(t)}{\mathcal{L}^2} \tag{5.7}$$

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

$$\rho(t) = \rho_0 \exp\Big(-\frac{2\pi\beta\ell^2}{\mathcal{L}^2}t\Big) \tag{5.8}$$

If density reductions from other secondary and tertiary infections are neglected, the final number of expected secondary infections after T infectious time-steps is given by:

$$R_0(T) = \mathcal{L}^2\big(\rho_0 - \rho(T)\big) = \mathcal{L}^2\rho_0\Big[1 - \exp\big(-\frac{2\pi\beta\ell^2}{\mathcal{L}^2}T\big)\Big] \tag{5.9}$$

Equation 5.9 can be used as a first approximation toward an R_0 value—the reader can find an equivalent, discrete-time derivation in Appendix B.1. Nonetheless, uniform density reductions (in equation 5.7) assume that secondary infections are equally likely at all spatial locations about the primarily infected tree. On average, neglecting spatial variations overestimates the number of secondary infections induced by the tail-ends of the dispersal kernel, thus giving rise to a greater R_0 value. A more accurate, albeit more complex, equation can be derived, allowing for Gaussian spatial variations

For transparency, the above derivation is solely based on β, and not the normalised infectivity parameter. However, from the exponential exponent in equation 5.9, it is clear to see how normalising the infectivity ($\beta^*/2\pi\ell^2$) prevents the number of expected secondary infections dependency on ℓ. That is, substituting $\beta^*/2\pi\ell^2$ into the equation 5.9 leads to an exponent, (i.e. $T\beta^*/L^2$) independent of ℓ.

Incorporating Gaussian dispersal

Equation 5.9 neglects a spatially varying transmission probability, though it can be refactored by first re-writing the density as:

$$\rho(r,T) = \rho_0 \exp\big(-\beta T g(r;\ell)\big) \tag{5.10}$$

where $g(r;\ell)$ is a Gaussian kernel. Equation 5.10 can be interpreted as the generalised form of equation 5.8, this time incorporating Gaussian spatial variations into R_0. Upon substitution back into equation (5.9), the total number of secondary infections after T time-steps is given by:

$$R_0(T) = \int_0^\infty 2\pi r\big(\rho_0 - \rho(r,T)\big)dr = \int_0^\infty 2\pi r\rho_0\Big[1 - \exp\big(-\beta T g(r;\ell)\big)\Big]dr \tag{5.11}$$

Note, the finite lattice square of size $\mathcal{L}^2$ has been replaced with integration in polar coordinates over dr. Equation 5.11 is hard to solve directly, but it can be integrated by performing a series expansion on the exponential term. One may then proceed to integrate on a term-by-term basis:

$$\begin{aligned}
R_0(T) &= \int_0^\infty 2\pi r\rho_0 \Big[1 - \exp\big(-\beta T g(r;\ell)\big)\Big] dr \\
&= 2\pi\rho_0 \int_0^\infty r\Big[1 - \sum_{n=0}^{\infty} \frac{\big(-\beta T\big)^n}{n!} \exp\big(-\frac{r^2}{2\ell^2}\big)^n\Big] dr \\
&= 2\pi\rho_0 \int_0^\infty r\Big[\sum_{n=1}^{\infty} \frac{(-1)^{n+1}\big(\beta T\big)^n}{n!} \Big(\exp(-\frac{nr^2}{2\ell^2})\Big)\Big] dr \\
&= 2\pi\rho_0 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}(\beta T)^n}{n!} \int_0^\infty r\exp(-\frac{nr^2}{2\ell^2}) dr \\
&= 2\pi\rho_0\ell^2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}(\beta T)^n}{(n)(n!)}
\end{aligned} \tag{5.12}$$

if β and T are small, the 1^{st} order term in equation 5.12 is sufficient to approximate R_0 as a linear function of T—confirmed by numerical simulations in the next section. Finally, equation 5.12 can be summed to give:

$$\begin{aligned}
R_0(T) &= 2\pi\rho_0\ell^2 \sum_{n=1}^{\infty} \frac{(-1)^{n+1}(\beta T)^n}{(n)(n!)} \\
&= 2\pi\rho_0\ell^2\big(E_1(\beta T) + \ln(\beta T) + \gamma\big)
\end{aligned} \tag{5.13}$$

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. Intriguingly, the jump between equations 5.12 and 5.13 is well known in the field of complex analysis—see [Abramowitz and Stegun, 1948], page 228. Alternatively, the summation in equation 5.12 could equivalently be given as:

$$\mathrm{Ein}(\beta T) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}(\beta T)^n}{(n)(n!)} \tag{5.14}$$

where Ein is known as the ‘entire’ function, leading to the well established relation:

$$E_1(z) = -\gamma - \ln(z) + \mathrm{Ein}(z)$$

The next section will investigate the suitability of 5.13 to describe the NLM.

5.4 R_0 behaviour: analytics vs numerics

In Figure 5.4, the analytic predictions of R_0 from equation 5.13 are compared against numerical simulations. Figures 5.4(a-b) plot the total number of secondary infections due to the primary infection over its lifetime, denoted by $\sum_{t=0}^{T} R_0(t)$. In both panels (a-b), NLM simulations were ensemble-averaged $N = 50$ times for four infectivity values, indicated by the solid coloured lines. The final value of R_0 is observed when the infectious lifetime of $T = 100$ steps is concluded. Figures 5.4(a-b) show two epidemic scenarios, with lower ($\ell = 35$) and higher ($\ell = 100$) dispersal parameters. As expected, equation 5.13 tends to overestimate R_0 for both dispersal parameters when infectivity is high, illustrated by the dotted scatter plot.

The time-series in Figure 5.5(a) reveals that lower β^* parameters produce a constant infection rate, indicated by linear relationship in blue-green. Equation 5.13 agrees well with these lower infectivity parameters. However, large deviations from model output can be seen for $\beta^* = 10$ at later times. Here, the number of new secondary infections plateau for $\beta^* = 10$ because other secondary/tertiary infections reduce host density around the primary infection. Subsequently, at later times, fewer and fewer trees in the primary infections neighbourhood are available to infect, causing R_0 to level off. Thus, equation 5.13 describes constant transition rates accurately but deviates from model simulations when infection rates decrease because other (secondary/tertiary) infections reduce host local densities.

In Figure 5.5(b), the dispersal parameter is increased to $\ell = 100$. A larger dispersal kernel encompasses a larger neighbourhood. Subsequently, $R_0(t)$ saturates less for higher R_0 values, in contrast to Figure 5.5(a). Despite a surprising degree of simplicity, the linear relation for lower β^* parameters is predicted by equation 5.13. The linearity can be understood by noting that when $\mathrm{Ein}(\beta T)$ is small in comparison to ℓ^2 (and βT is small), the first-order term inside equation 5.13 reduces to a linear equation in T. More

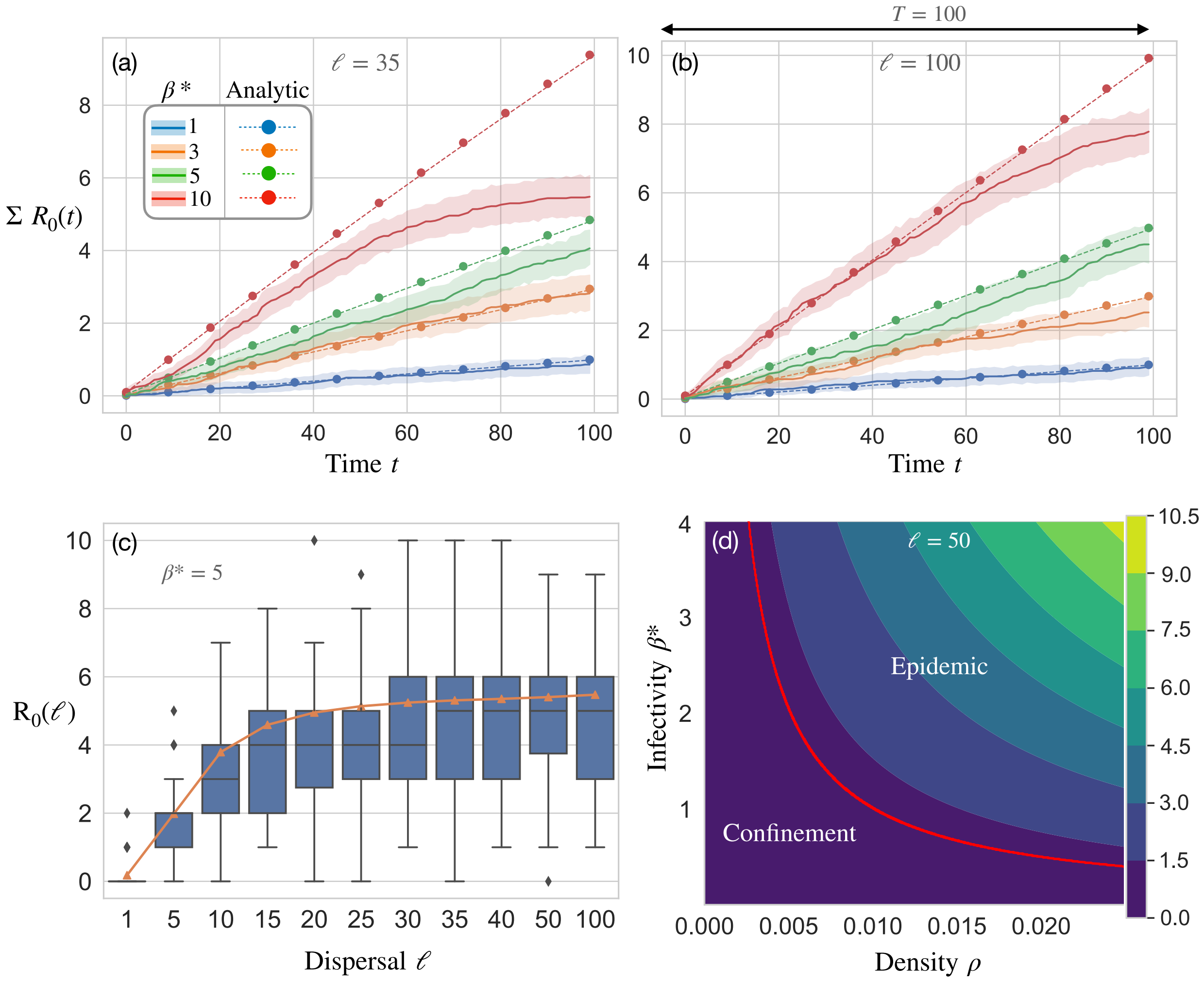


Figure 5.4: Comparisons between the analytical expression for R_0, according to equation 5.13, and numerical simulations. In panels (a-b), The cumulative sum of secondary infections are ensemble-averaged (50 repeats) over the infectious lifetime of $T = 100$ steps. (a) When the scale of dispersal is smaller ($\ell = 35$), a high infectivity (in red) causes R_0 to saturate to a maximum over the infectious period. For smaller infectivities (blue-green), R_0 increases at a constant rate and fails to saturate. (b) For a larger dispersal kernel ($\ell = 100$) and high infectivity, R_0 increases at a more constant rate, yet deviations still exist at the conclusion of its infectious lifetime. (c) R_0 is shown as a function of the dispersal parameter for fixed infectivity $\beta^* = 5$. For small length scales, the normalised infectivity produces a less infectious outbreak. However, R_0 is approximately fixed for larger ℓ values. Analytic predictions, shown in orange, tend to overestimate the spread for small ℓ. (d) The 2D R_0 phase plane predicted by equation 5.13. A threshold, given by $R_0 = 1$, is plotted in red that predicts the separation between confinement and epidemic.

interestingly, however, we can observe that R_0 is similar when β^* is low, revealed by comparing the blue-green lines in Figures 5.5(a-b). Whereas, for higher infectivities, R_0 tends toward a smaller value in Figure 5.5(a) comparison to panel (b). Observing R_0 deviate with different ℓ parameters compels a study of R_0 over the space of ℓ, which leads to Figure 5.5(c).

In Figure 5.5(c), the basic reproduction number R_0 is assessed over a range of dispersal kernels. Tree density in Figure 5.5(c) is fixed fixed to $\rho = 0.01$ together with infectivity $\beta^* = 5$. Predictions from equation 5.13 are shown in orange, and R_0 can be seen to increase with the dispersal kernel up to around $\ell \in [25, 30]$ before saturating to $R_0 \sim 5$. When ℓ is small, R_0 is low; the reason for this are two-fold: (A) neighbourhoods defined by a small ℓ parameter are likely to become fully occupied by secondary-infected trees (B) the scale of dispersal is less than the average distance between trees, thereby preventing the spread. Then, as the kernel is increased, R_0 asymptotically increases to a maximum value beyond which increasing ℓ bares no impact on R_0—provided that the number of secondarily infected trees is small in comparison to the number of hosts in the neighbourhood[3]. Therefore, if ℓ is large enough, the normalised infectivity can be seen to effectively constrains the epidemic severity to a limiting value.

From the reproduction number, a transmission threshold can be defined by $R_0 = 1$, predicting the separation of states between confinement and epidemic. In Figure 5.5(d), the threshold predicted by equation 5.13 is marked in red, overlaying a two-dimensional phase plot of R_0 over tree density and infectivity. According to Figure 5.5(d), when model parameters satisfy $R_0 > 1$, the pathogen may propagate for a time before dying off or culminate in an epidemic. Below the threshold, the pathogen has little chance of spreading to neighbouring trees and little chance of causing a large-scale epidemic. Equation 5.13 provides a computationally efficient means to categorise model behaviour. The following section assesses how the total tree mortality, or final-sized epidemic, relates

---

[3] Insight into the underlying behaviour could be gleaned by looking at equation 5.13. That is, by accessing the growth of $2\pi\rho_0\ell^2$ and the convergence of the function Ein (defined by equation 5.14) and noting that as $\ell \to \infty$, $\beta = \beta^*/2\pi\ell^2 \to 0$. Although an in-depth mathematical analysis was not undertaken, it is clear that for large values of ℓ, R_0 asymptotically approaches a limiting value.

to threshold $R_0 > 1$ predicted by 5.13.

5.4.1 Tree mortality versus R_0

A swift response during the early phases of an outbreak increases the chance of successful control [Cunniffe et al., 2015b]. Nevertheless, knowing whether or not an early-stage outbreak will snowball into a large-scale epidemic is not always clear—as was the case for chestnut blight in Europe [Hillman and Suzuki, 2004]. Thus, we desire accurate predictions of the final epidemic outcome from observations of the first few infections. For a well-mixed (non-spatial) population as in the standard SIR, we can conveniently employ the reproduction number to estimate the number of expected cases as a fraction of the population[4].

However, the relationship between R_0 and the final epidemic outcome is undoubtedly more complex when there is spatial structure, as in the NLM. Hence, this section investigates the relationship between R_0 and the total number of host removals, denoted by χ. Investigating R_0 and χ is crucial to establish the existence of the threshold at $R_0 = 1$. In addition, this analysis is compelled further by noting that some authors fail to demonstrate the threshold at $R_0 = 1$ and give improper definitions of R_0, as explained by [Li et al., 2011].

Primarily, the main epidemic parameters in the NLM consist of ρ and β^*, therefore varying ρ and β^* in Equation x result in different R_0 predictions that we can plot against the observed tree mortality. Consequently, Figure 5.5(a) shows ensemble-averaged tree mortality plotted as a function of predicted R_0 values for three density parameters, indicated by solid lines blue-green. A total of 250 replicate simulations formed the ensemble average, and simulations evolved until all infected trees became extinct or for 2500 time-steps elapsed, whichever occurred first. The ensemble-averaged mortality is overlaid by a scatter plot depicting a small sample of data points; the colour of each sample point

---

[4] In the standard SIR model, the number of expected cases as a fraction of the populations can be shown to follow: $R_\infty = S_0\Big[1 - \exp\big(R_0 R_\infty\big)\Big]$. Here, R_∞ is the number of cases as a fraction of the population size, the so-called 'final-sized epidemic', see [Chowell et al., 2009] for a more in-depth break down.

reflects the infectivity parameter β^*.

Figure 5.5(a) shows the relationship between tree mortality, the predicted R_0 from equation 5.13, and the two epidemic parameters ρ and β. A scatter plot of coloured data points (depicting a small sample of data) overlays a continuous line representing an ensemble average. The corresponding colour bars on the scatter plot and ensemble-average characterise infectivity and density, respectively. The Figure illustrates that tree mortality remains low when $R_0 < 1$ and β is small, as indicated by coloured scatter plots in blue. Conversely, the tree mortality increases with infectivity when $R_0 > 1$, shown by the yellow scatter points. Therefore, equation 5.13 demonstrates a threshold-like behaviour defined by $R_0 = 1$. Above the threshold $R_0 = 1$, increasing the tree density increases the epidemic scale because more susceptible hosts are available to infect, as indicated by the significant rise in tree mortality in the continuous green line.

Despite being above the threshold, the numerical simulations can still fail to produce an epidemic, illustrated by the small number of data points beyond $R_0 = 1$ that map to a zero-sized epidemic. For example, at $R_0 \approx 2$ a small number of points can far below each ensemble mean. Similarly, the ensemble mean is lowered by pathogen extinction—demonstrated by slight differences between the collection of points and the ensemble mean in the interval $R_0 \in [1, 3]$. These observations result from the fact that under the influence of early stochastic forces, the probability of epidemic extinction is higher [Heffernan et al., 2005, Tildesley and Keeling, 2009], which unfortunately marks a flaw in the concept of R_0 in general [Li et al., 2011].

Figure 5.5(b) presents the same essential information as Figure 5.5(a). Although, the ensemble mean is plotted against the mortality ratio (i.e. the total number of removed trees as a fraction of the total population), denoted by χ. Given that each ensemble mean converges to the same epidemic scale, the quantity χ demonstrates utility when accessing the epidemic impact between different tree-densities. Consequently, the threshold $R_0 = 1$ is easily observable in Figure 5.5(b), indicated in shaded grey. The threshold-like behaviour witnessed in Figures 5.5(a-b) demonstrate that equation 5.13 provides a sim-

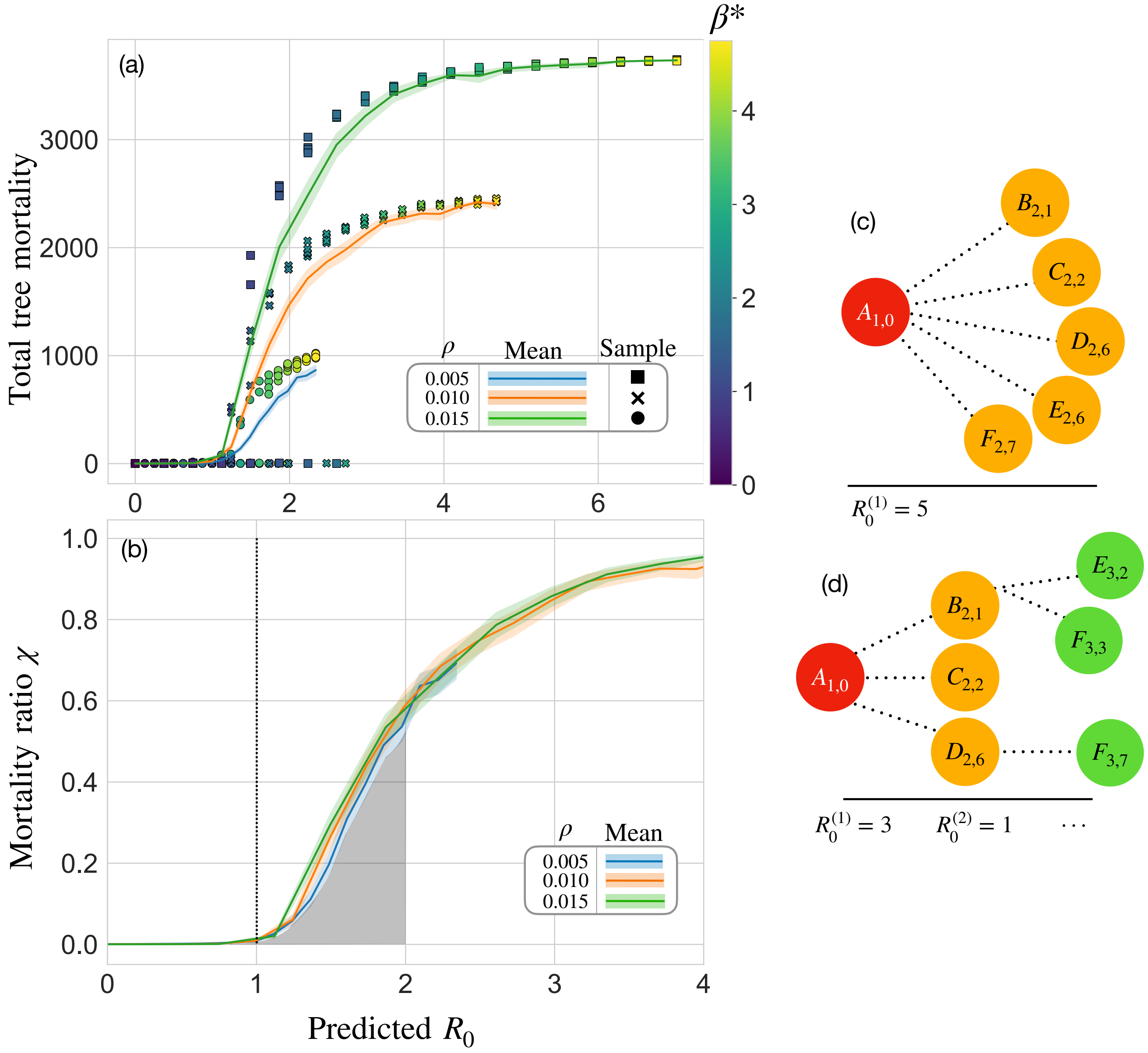


Figure 5.5: The relationship between the total tree mortality and the predicted R_0 value. (a) For three values of tree density and multiple infectivity values, the ensemble-averaged tree mortality (as a continuous solid line) is overlaid with a small sample of data points shown by the scatter plot. The shape and colour of each data point indicate the tree density and infectivity, respectively. (b) The fraction of removed trees as a fraction of the population is plotted against mortality the analytical value of R_0. Together panels (a) and (b) demonstrate a threshold-like behaviour at $R_0 = 1$. (c) A graphical representation of the idealised R_0 approximation, as per equation 5.13. (d) A more realistic network representing the epidemic branching process. In both panels, the arrow of time is left to right, and subscripts reflect the generation and time-step that trees become infected.

ple predictive framework for the NLM. That said, more complicated dynamics—such as exponential lifetimes, elaborate dispersal kernels or host aggregation—could significantly hinder the analytic solution proposed in section 5.3.2.

Another limitation to the equation 5.13 can be explained by Figure 5.5(c), which shows a typical NLM simulation used to access the analytic expression of R_0. Namely, a single 'fist-generation' infected tree at time $t = 0$ ($A_{1,0}$) that happens to infect a number of neighbouring trees ($B_{1,1}$ to $G_{1,7}$). According to the R_0 approximation, the local density reductions due to $B_{2,1} - G_{2,7}$ are neglected, and $A_{1,0}$ remains the only active source of infection. Neglecting the influence of other secondary infected hosts helped to keep analytical derivation simple. But for highly infectious outbreaks, equation 5.13 is likely to overestimate R_0. Figure 5.5(d) can be used to understand why overestimates of R_0 are likely. In Figure 5.5(d), non-trivial density reductions could be expected from the second and third generation of infected hosts, $B_{2,1}, C_{2,2}, D_{2,6}$ and $E_{3,2}, D_{3,5}, D_{3,6}$ respectively—here, the first subscript refers to time and the second subscript refers to the generation—leading to an environment where the primary infected host $A_{1,0}$ has less neighbours to infect. Given these limitations, a and more flexible method of calculating R_0 is investigated in the next section.

5.5 Contact-tracing secondary infections

In this section an alternate method of calculating R_0 is presented that incorporates the effect of secondary infections (up the n^{th} generations), analogous to contact-tracing emerging epidemics in human populations [Eames and Keeling, 2003]. By collecting individual tree-to-tree induced secondary infections, the entire history is captured, illustrated graphically in Figure 5.5(d). At $t = 0$, the first generation primary infected tree, denoted by A, produces three 2^{nd} generation infections B-D in orange that in turn produce third generation infections are shown in green. The contact-traced R_0 can be defined by:

Definition 5.5.1. *at $t = 0$, simulations begin with one or more infected hosts, and the entire epidemic history captures which host infects which others. The mean number of* *infections that result for each generation i is computed and denoted by* $R_0^{(i)}$.

In Figure 5.6(a), observations from the (ensemble-averaged) contact-traced R_0 is shown for 10 generations and four infectivity parameters; in all plots, density, dispersal and domain sizes remain fixed ($\rho = 0.01, \ell = 50, \mathcal{L} = 500$). For highly infectious outbreaks, the host population quickly decreases, and $R_0^{(i)}$ begins high and gradually decreases with each generation. In contrast, for infectivities just above the threshold $R_0^{(i)}$ remains approximately stable with each generation because the population of susceptible hosts remains high. For all boxes in Figure 5.6(a), the interquartile range decreases with the generation, suggesting that early stochastic forces increase the spread of $R_0^{(i)}$ values. The number of secondary-infected trees will therefore vary over time and mirror the host population. Had the re-growth of susceptible trees been considered, $R_0^{(i)}$ can be speculated to behave very differently for later generations.

Figure 5.6(b) compares the ensemble-averaged $R_0^{(i)}$ values, shown in Figure 5.6(a), to predictions from the analytic expression for R_0. Analytic R_0 values are plotted as horizontal dashed lines, and each $R_0^{(i)}$ ensemble average (shown by the solid lines) is surrounded by shaded bounds reflecting a 95% confidence interval. For the lowest infectiviy parameter around the threshold $R_0^{(i)} = 1$, shown in blue, the contact-traced value of R_0 compares well with equation 5.13. However, increasing the infectivity to $\beta^* = 2$ (in orange) equation 5.13 agrees well for early generations, but deviates for later generations, revealed by comparing the dashed horizontal and solid orange lines. Deviations only grow larger as β^* increases. That is to say, looking at the green and red lines in Figure 5.6(b), one can confirm that equation 5.13 does indeed overestimate pathogen transmissibility. Although significant disparities exist, Figure 5.6(b) implies that both analytic and contact-traced reproductive ratios agree on the threshold $R_0 = 1$.

Simplistic interactions between trees permit an alternative network representation of disease spread. In Figures 5.6(c-e), a directed network of disease spread is shown for three typical simulations with infectivity parameters $\beta^* \in \{1, 2, 3\}$. Nodes depict individual trees, while colour and size represent the generation infected and the number of induced

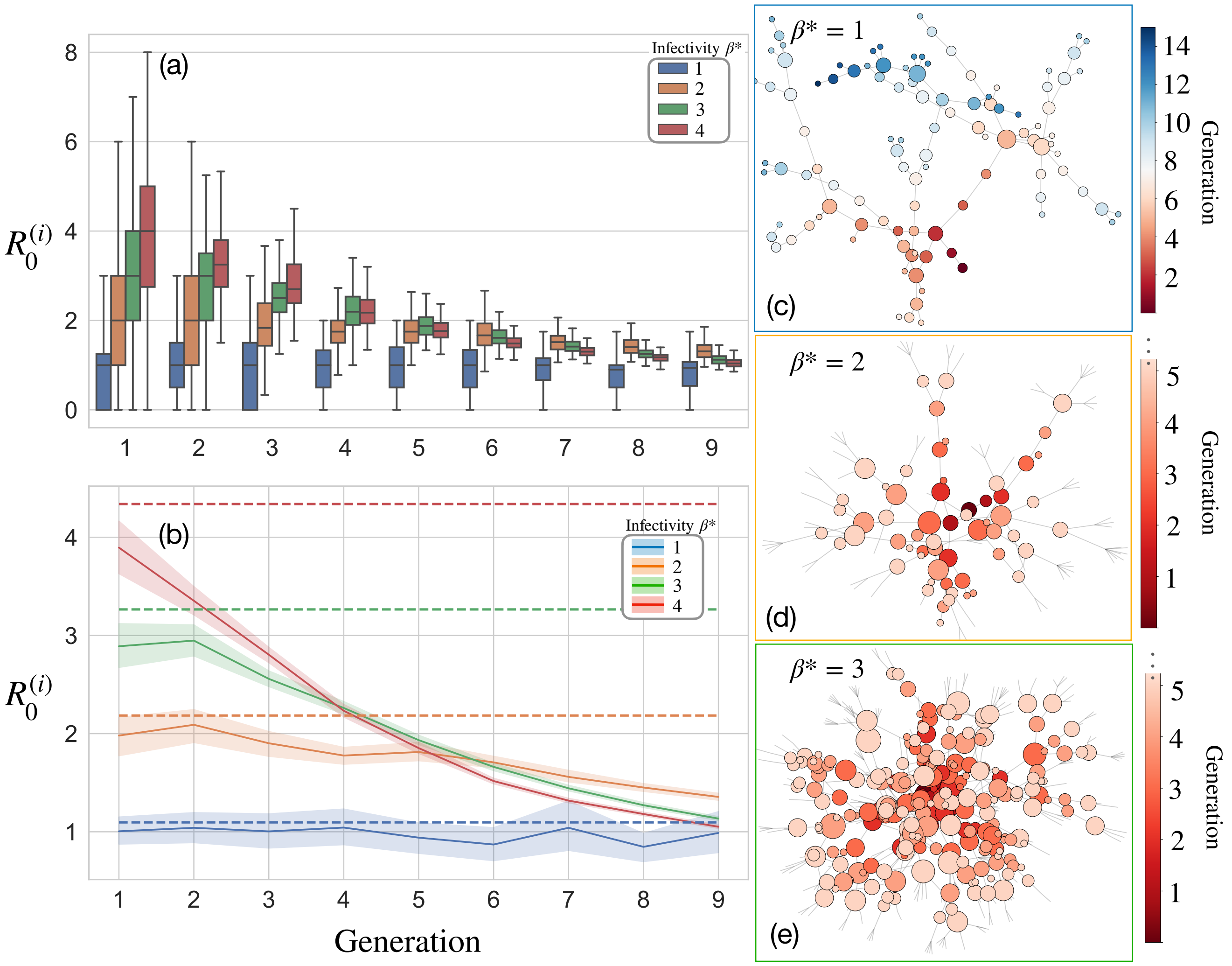


Figure 5.6: Contact-tracing the mean number of secondary infections for the i^{th} generation of infected trees, $R_0^{(i)}$. For all simulations, tree densities are fixed along with dispersal kernels, ($\rho = 0.01$) and $\ell = 50$ respectively, inside a domain of size $\mathcal{L} \times \mathcal{L} = 500 \times 500$. (a) A box and whisker plot showing the mean number of secondary infections plotted over 10 generations over an ensemble of size $N = 500$. Four different infectivity values are shown, from blue to red. (b) The ensemble-mean in (a) is compared against predictions from the analytic expression of R_0, plotted as horizontal dashed lines. For early generations, equation 5.13 agrees well with the contact-traced value of R_0 but overestimates the spread for higher infectivities. (c-e) A network representation of typical simulations for parameters $\beta^* \in \{1, 2, 3\}$. The nodes' colour and size reflect the generation and number of secondary infections, respectively. As β^* is increased, the network quickly explodes—thus reflecting the complexity of controlling highly infectious outbreaks.

secondary infections. Between two nodes, one edge connects infected generations n and generation $n+1$ (arrows showing the direction are omitted for visual clarity). When epidemic parameters are around the threshold, as in Figure 5.6(c), the network appears sparse and chaotic; whereas Figures 5.6(d-e) show that if infectivity is increased, the network proliferates rapidly. Indeed for $\beta^* = 2$ and $\beta^* = 2$ the network became large if plotted for all generations—consequently $R_0^{(5)}$ was truncated to permit visualisation.

From Figures 5.6(d-e), one can discern assumptions implicit within the NLM and visualise how targeted epidemic control might be optimised. In the NLM, tree-to-tree interactions are particle-like, meaning that infections spread unidirectionally between two trees at any time. In real life, interactions are more complex. For example, consider two infected trees (A, B) in the vicinity of one healthy susceptible neighbour C. Here, infection pressure on C is likely a continuous function of A and B. We can collectively represent these more complex interactions by multiple edges between A, B and C in a network diagram. However, the probability of C becoming infected by A, B (or $A \cap B$) occurs with statistical independence. Therefore, we employ the standard method of combining statistically independent events from the inclusion-exclusion formula—given in appendix B.3.

Despite the necessity of modelling the simultaneous infection pressure from multiple infected trees, it complicates the definition of R_0. Primarily, infections can originate from multiple sources, and we cannot tell from which tree(s) the infection initially spread. See appendix B.3 for a more in-depth discussion on combining probabilities. Nonetheless, another assumption relates to self-loops in the network. In the framework proposed here, trees transition into the R compartment and self-loops are negated. However, in reality, re-infection is possible (e.g. the yearly cycle of ash dieback [Gross et al., 2014a]) that could be supposed to cause a host's infectivity to increase with each re-infection.

The networks shown in Figures 5.6(d-e) present a simple, yet insightful, representation to gauge what connections epidemic control would need to disrupt. Well-known results suggest that the scale of control should reflect the spatio-temporal scale of disease spread [Gilligan et al., 2007a]. In Figure 5.6(d), even a minimal control effort might disrupt the network well below the threshold of spread, whereas Figures (d) and (e) would incur an increased effort to stop the spread. In the network representation, optimised control would involve targeting high priority links between nodes; this reflects various methods that aim to optimise control by identifying high-risk hosts.

5.5.1 Contact-traced R_0 and tree mortality

In Figure 5.7, the contact-traced value of R_0 is compared against the tree mortality, similar to the previous analysis in section 5.4.1. Figure 5.7 shows an ensemble-averaged reproductive ratio (up to $R_0^{(i=5)}$) plotted against the mean tree mortality, indicated by the dashed curves. The ensemble mean is then overlaid with a coloured scatter plot depicting a small sample of the data. Each simulation in Figure 5.7 was repeated 10^3 times for fixed density $\rho = 0.01$, dispersal $\ell = 50$ and $\mathcal{L} = 500$ over 2500 time-steps. As before, a threshold can be witnessed at $R_0^{(i)} = 1$, above which tree mortality rises steeply. For $R_0^{(1)}$, the threshold phenomena witnessed in Figure 5.7 (shown in dashed blue) appears similar to the previous analytic R_0 examination. However, each successive generation appears to define a sharper threshold, especially when β^* is high, suggested by the steeper coloured dashed curves.

The steeper thresholds witnessed in Figure 5.7 can be understood by considering initial stochasticity. If an outbreak (with high β^*) does not become extinct at early times, the epidemic is likely to continue to spread until no susceptible trees are left to infect. Hence, epidemic impact is high and the threshold is sharp for $R_0^{(i)}$, where $i > 1$. The reader is referred back to the network diagrams shown in Figure 5.6(d-e) to gain intuition behind this idea; if the pathogen survives beyond the initial outbreak to establish new centres of infection, the network quickly explodes, and pathogen extinction is unlikely.

The ensemble shown in Figure 5.7 was re-run with 10 centrally-located initial infections at $t = 0$ to test the initial stochasticity. Intuitively, increasing the number of infected trees reduced early extinction events and subsequently raised the mean tree mortality. In addition, raising the number of initially infected trees reduced stochasticity in the

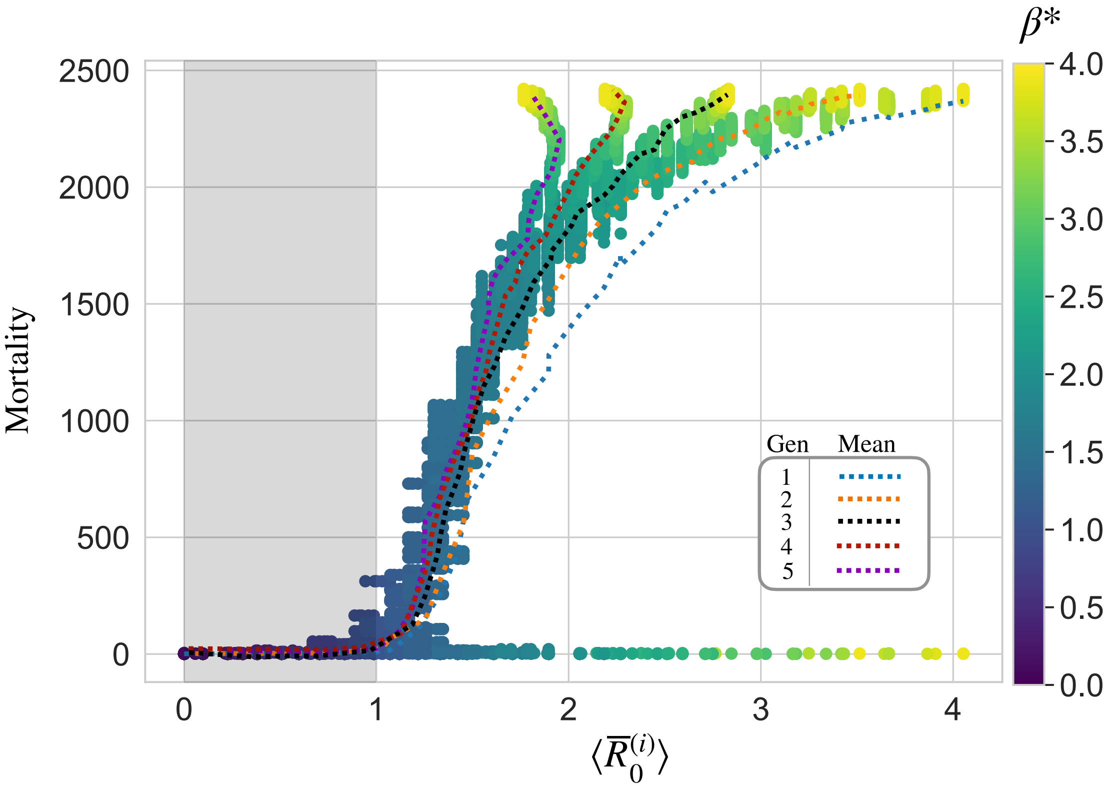


Figure 5.7: Comparing the contact-traced reproduction ratio and the total tree mortality. In each simulation, the tree density and dispersal kernel was fixed to $\rho = 0.01$ and $\ell = 50$, respectively. For each value of infectivity β^*, the ensemble-averaged value of $R_0^{(i)}$ is plotted against the mean tree mortality, shown by the dashed curves, for five generations, i.e. $R_0^{(5)}$. A coloured scatter plot overlays the ensemble-averaged line plots showing a sample of the data—and also reflecting the value of β^*. As before, a threshold arises around $R_0^{(i)} = 1$, although this time the threshold appears steeper.

ensemble and presented a more abrupt threshold in comparison to Figure 5.7—more information can be found in Appendix B.4.

5.6 Discussion and future work

The main aim of this Chapter was to construct a more realistic, dispersal-based model of tree disease. Hence, a non-local dispersal model (NLM) of tree disease was constructed with SIR compartments and a Gaussian kernel. Notwithstanding its generic construction, the NLM provides a foundation for the remaining chapters in this thesis. After describing the NLM, it was compared to the standard SIR framework for a number of dispersal length scales and domain sizes. Despite some differences, the NLM compared more favourable with the SIR model when the dispersal scale parameter was comparable to the domain size. Comparisons, therefore, provide compelling indications that spatially-explicit contact-mixing in the tree population may arise with sufficient dispersal. From this observation, we may question the utility of spatial models for systems where the dispersal length scale is comparable to the size of domain, e.g. when modelling the spread of disease in small fields or plantations. However, comparisons to the SIR model were simplified and limited to one parameter (i.e. the ratio β/γ). As such, the analysis constitutes a preliminary result, and a more sophisticated comparison method is required to glean further insight. For example, a better method might involve using inference (MCMC methods) to fit the NLM against the standard SIR model.

Two methods of calculating a reproduction ratio, one analytic (R_0) and one numerically contact-traced ($R_0^{(i)}$), were outlined to categorize the NLM. The analytic threshold predicted by equation 5.13 agreed well with the 'actual' contact-traced reproduction ratio computed through NLM simulations, caveat-ed by the observation that it tends to overestimate R_0 when epidemic severity is high. The overestimation of R_0 can be compared to well-known results by [Tildesley and Keeling, 2009, Keeling and Eames, 2005], who showed that the first-generation basic reproduction ratio for farms infected with foot-and-mouth overestimates the growth rate of infection. In addition, [Tildesley and Keeling, 2009] found that the second generation of infected farms gave a better predictor of the final-sized epidemic (conditional on the epidemic occurring), thus presenting a clear link to our observation that the tree mortality threshold defined by $R_0^{(i)} = 1$ was sharper for later generations. Crucially, each method of calculating the reproduction ratio defined a threshold around unity, beyond which epidemic impact becomes non-trivial.

An attractive feature of the contact-tracing method, as per definition 5.5.1, pertains to its flexibility in the face of more complex spatially explicit models. Analytical solutions of R_0 may become hard to determine for more elaborate life-cycles, dynamics and aggregated host distributions. Although contact-tracing provides an easy-to-implement method of calculating $R_0^{(i)}$, it should remain, first and foremost, an abstract modelling tool. For example, consider the immense difficulty of experimentally contact-tracing secondary infections when an epidemic spreads through a forest/landscape. Contact-tracing the reproduction ratio can therefore be presumed as unobservable in nature[5] However, given sufficient data, one might fit a value of β and reverse-engineer a value of $R_0^{(i)}$ from the model; which leads us to a discussion around the infectivity parameter β.

A probability of state-change represents infectivity in the NLM, which followed naturally from the percolation model outlined in section 3.3 [Orozco-Fuentes et al., 2019]. However, growth rates are usually employed to describe infectivity—going back to the original SIR framework [Kermack and McKendrick, 1927] and the logistic growth model of [Van der Plank, 1963]. The approach adopted in this Chapter is, therefore, a-typical of contemporary dispersal models based on rates, e.g. [Fabre et al., 2021, Bussell et al., 2019, White et al., 2017, Cunniffe et al., 2016]. Given that growth rates parameterize most diseases in the literature, the NLM might arguably require a modification towards a rate-based implementation; however, it must be remarked that sometimes this may not be needed, given that measuring growth rates—particularly for time-varying infectivities—is extremely difficult [Cunniffe et al., 2015a]. Although unconfirmed, we may suppose an equivalence between the infectivity β and an emergent growth rate. This assertion is supported by the similarities exhibited between the NLM and the rate-based standard SIR model in Figure 5.2 (in addition to appendix B.2.1). Nevertheless, presenting β as a probability describes an intuitive low-level (microscopic) perspective of disease spread which

---

[5] We may speculate about measuring a time-varying reproduction number based on the observed numbers of infected hosts, a well-known concept for characterizing epidemic transmission in human populations [Thompson et al., 2019].

would, in theory, be observable/measurable in reality, in contrast to the $R_0^{(i)}$. Hence, we may consider infectivity β as a fitting parameter.

Chapter 6

Constructing R_0-maps over Great Britain

In the last Chapter, we considered a generic SIR-based non-local dispersal model (NLM) that spread via a Gaussian dispersal. The NLM resolved the major problem witnessed in section 4.9, namely, the failure of the percolation-based model to spread on a realistic host density. However, Chapter 5 and the NLM lacked biological specificity. Therefore, the present Chapter presents a simplified SEIR model of ash dieback capturing only the essential dynamics of wind-borne dispersal. Ash dieback (ADB), caused by the fungus *Hymenoscyphus fraxineus*, poses a threat to the survival of European ash (*F. excelsior*)—the reader is referred to section 2.3 for a more in-depth discussion of ADB symptoms, life-cycle and management.

In section 6.1, a seasonal-based SEIR-like model is developed to simulate the natural wind-dispersal mechanism of ADB at local-spatial scales. Next, section 6.2 examines the general pattern of epidemic spread in the SEIR model before section 6.3 outlines a formal definition of an effective reproduction number to measure pathogen invasiveness. The effective reproduction number is spatially explicit and based on the same contact-traced method presented in Chapter 5. Critically, the wind-dispersal model of ADB rests on R_0 measured over an appropriate spatio-temporal scale.

Lastly, in section 6.4 a framework is developed to spatially-scale R_0 values over Great Britain (GB) using a map of predicted ash abundance [Hill et al., 2017]. In effect, this means embedding R_0 values from a small-scale epidemic model inside each pixel of an extensive host data set spanning GB, thus linking two models at different spatial scales.[1] Fundamentally, projecting R_0 values onto a map of ash densities permits the visualisation of epidemic impact over GB. The framework is generic and adaptable to any dispersal-based pathosystem, provided a sufficient host density data-set.

The local-scale epidemic model demonstrates that long-distance dispersal (LDD) and long-term pathogen survival can occur even below the threshold $R_0 = 1.0$. In addition, analysis at the landscape-level covering GB reveals that susceptible clusters of ash grow most rapidly over a narrow range of infectivity parameters, presenting behaviour akin to a global epidemic phase transition. Finally, the method developed in this Chapter demonstrates how the epidemic scale can vary significantly, albeit with minor epidemic parameter variations. Together, these observations support the call for a risk-based approach to modelling the spread of epidemics in tree populations.

6.1 A spatially-explicit seasonal SEIR model

H. fraxineus has two modes of reproduction, sexual and asexual (teleomorphic and anamorphic, respectively). Some research suggests that the asexual element has contributed to the European epidemic [Fones et al., 2016], although the sexual reproductive mode of ADB is widely acknowledged to be the dominant driver of disease-spread [Mansfield et al., 2018, Haňáčková et al., 2017, Gross et al., 2012, Timmermann et al., 2011]. Subsequently, only the sexual reproduction of *H. fraxineus* is reflected in the SEIR model.

The entire sexual reproductive cycle occurs on ash leaves, yet the fungus will spread

---

[1] In general, combining two models at different spatial scales has clear analogies to a sub-grid model [Herring, 1979], or more recently, a metapopulation-like model [Grenfell and Harwood, 1997].

through ash hosts once infected. As such, the precise sexual reproductive mode presents a peculiar modelling scenario where both leaves and trees can be argued to host the pathogen *H. fraxineus.* Typically, leaf dispersal emanating from a large deciduous tree species (such as *H. fraxineus*) remains close to the tree trunk and rarely exceeds 30m [Nickmans et al., 2019]. Hence, leaf litter is assumed to fall close to infected trees to treat both ash hosts and infected litterfall at the same lattice position in the model. A more intricate model of ADB could aim to relax this assumption and describe autumn leaf-shed with a suitable leaf-fall dispersal kernel[2].

6.1.1 Infection dynamics

The infection dynamic comprises four states: susceptible S, latently infected E, infectious I and removed R that transition through $S \to E \to I \to R$ without the possibility of recovery. Figure 6.1 shows the infection process during the n^{th} year: a previously infected tree (I_m, where $m \leq n$) infects a susceptible tree (S) that becomes latently infected (E_n). A newly exposed/latently infected ash tree then transitions into an infectious tree the following year (I_{n+1}). Notably, the same flat, randomly-distributed population (used previously in Chapters 3-5) describes the configuration of ash trees.

Here, each lattice point in the $L \times L$ domain is chosen to represent a 5m × 5m patch of land that approximates the canopy cover of an average ash tree—although young saplings usually assume less area. A domain resolution of 5m × 5m yields an upper bound of 400 ash trees per hectare of canopy cover, indicative of densely populated ash stands [Dobrowolska et al., 2011a, Thomas, 2016].

---

[2] Patterns of leaf fall are vital for ecosystem nutrient recycling [Staelens et al., 2003]. Leaf-fall dispersal kernels have been collected (and modelled) for oak, beech, hornbeam and birch [Nickmans et al., 2019], although despite a detailed search, no such dispersal kernel for ash could be found.

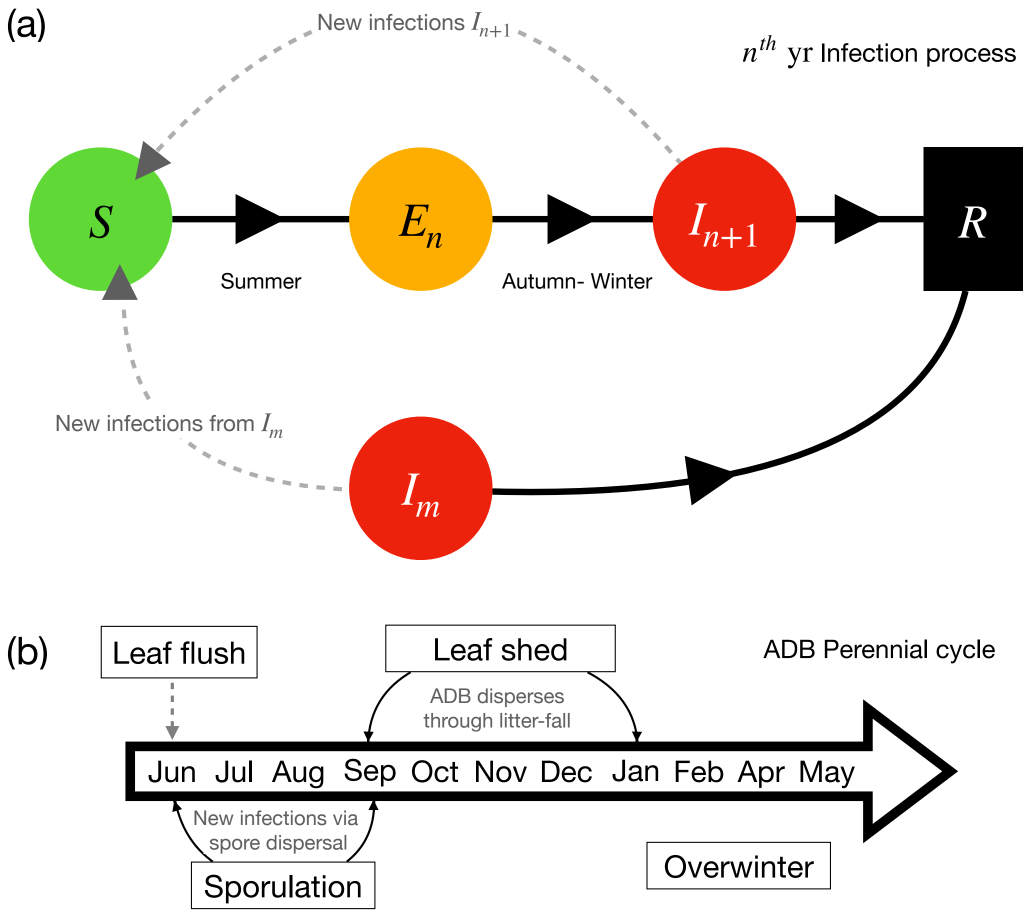


Figure 6.1: The seasonal SEIR model of ADB. (a) In year n of an outbreak, a previously infected tree I_m ($m \leq n$) may infect susceptible ash (in S), causing a transition ($S \rightarrow E_n$) during summer, as depicted by the bottom dashed grey arrow. A tree that becomes latently infected in the n^{th} year (E_n) becomes infectious in the following year (I_{n+1}). At that point, I_{n+1} can infect more susceptibles in the following summer, represented by the upper dashed grey arrow. All infected ash trees die without the possibility of recovery. (b) The yearly cycle of the pathosystem ADB. Leaf flush coincides with the sporulation season. Sporulating fruiting bodies release ascospores between June-September. Fungal fruiting bodies begin to grow on ash litter fall from late summer-early to winter. Overwinter, the fungus continues to develop on decaying leaves until leaf-flush and sporulation occurs in the following summer.

Transitions: $S \to E_n$

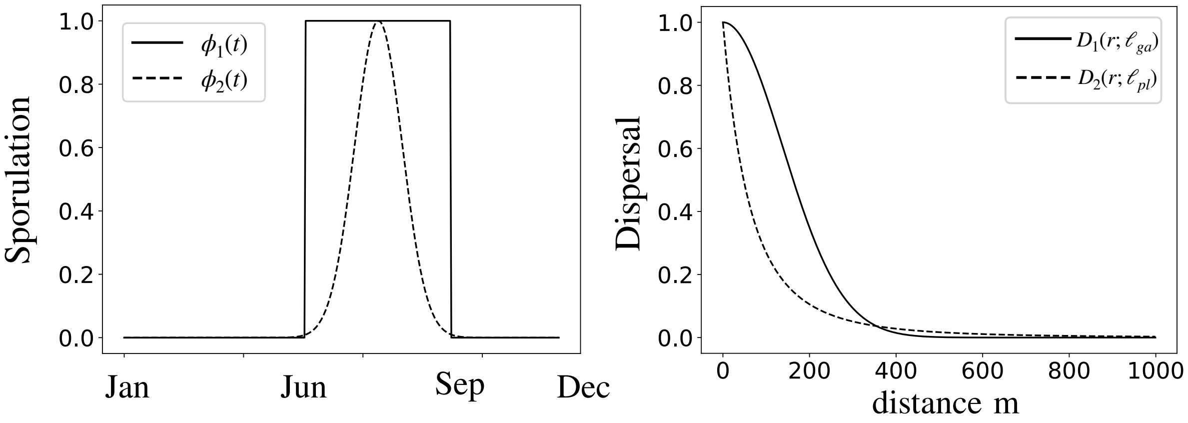
Individual tree-to-tree interactions are modelled as a system of particles (introduced previously in Chapter 5). Two component functions then control the infection probability, a time-varying infectivity $\beta\phi(t) \in [0,1]$ and dispersal function $D(r;\ell) \in [0,1]$. In particular, we will consider a thin-tailed Gaussian and a normalisable fat-tailed inverse power law to model dispersal. In year n, an infected tree located at x' can infect a susceptible tree located at x, represented mathematically as $S_x \to E_{x,n}; I_{x',m}$. The indices n and m are introduced for convenience to describe the year ash trees first become exposed, where $m \leq n$ denotes a previously infected tree. Thus, infected trees I_n and I_m occupy identical infectious states that do not affect the infection dynamics. If an infected tree remains infectious during the step $t \to t + \delta t$, the probability of transition follows:

$$Pr(S_x \to E_{x,n};\ I_{x',n}) = \beta\phi(t)D(r;\ell) \tag{6.1}$$

where r is the distance between x and x', and $\phi(t)$ is a time-dependant function reflecting the seasonal life-cycle of ADB, henceforth referred to as a *sporulation function.* Following the method outlined in Chapter 5, transition probabilities are calculated at each time-step, taken to be days, while trees are infectious.

The sporulation function ϕ dictates when infectious trees in I_n produce new secondary infections. For ADB, sexual reproduction repeats yearly from June to September during its *'sporulation season'*. As such, ϕ is non-zero during the sporulation period. Two sporulation functions are considered, ϕ_1 and ϕ_2, based on the step function and normal distribution, respectively. Sporulation is developed later in section 6.2.1, alongside the precise functional form of ϕ_1 and ϕ_2. Altogether there are four infection models, summarised in table 6.1.

Once in the E compartment, trees are infected but not infectious, i.e. latently infected. For example, ash infected with *H. fraxineus* take approximately two weeks to display symptoms [Cleary et al., 2013, Mansfield et al., 2018], and start to shed infectious



Model name	$Pr(S_x \to E_{x,n}; I_{x',m}) \in [0,1]$	β^* factor
1: ϕ_1-ga	$= \begin{cases} \beta \exp\Big[-\frac{r^2}{\ell_{ga}^2}\Big] & t \in [\text{June},\ \text{September}] \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{1}{T}\dfrac{1}{\pi\ell_{ga}^2}$
2: ϕ_1-pl	$= \begin{cases} \beta\big[1+\frac{r}{\ell_{pl}}\big]^{-a} & t \in [\text{June},\ \text{September}] \\ 0 & \text{otherwise} \end{cases}$	$\dfrac{1}{T}\dfrac{(b-1)(b-2)}{2\pi\ell_{pl}^2}$
3: ϕ_2-ga	$= \beta \exp\Big[-\dfrac{(t-T_{SP})^2}{2\sigma_{SP}^2}\Big] \exp\Big[-\dfrac{r^2}{\ell_{ga}^2}\Big]$	$\dfrac{1}{\sqrt{2\pi}\sigma_{SP}}\dfrac{1}{\pi\ell_{ga}^2}$
4: ϕ_2-pl	$= \beta \exp\Big[-\dfrac{(t-T_{SP})^2}{2\sigma_{SP}^2}\Big]\Big[1+\dfrac{r}{\ell_{pl}}\Big]^{-a}$	$\dfrac{1}{\sqrt{2\pi}\sigma_{SP}}\dfrac{(b-1)(b-2)}{2\pi\ell_{pl}^2}$

Table 6.1: Four infection models are considered, composed of either: Gaussian (ga) or inverse power law (pl) dispersal, and step function (ϕ_1) or peaked (ϕ_2) sporulation. All component functions yield values in the interval $[0,1]$ and are normalisable. The corresponding model parameters are shown below in Table 6.2. The right-hand column shows the normalisation factor used to construct β^*, i.e. used to fix the infectivity and epidemic-impact between models.

Model parameter	**Description**	**Value(s) taken**
ρ	Tree density	$0.00 - 0.10$
β	Infectivity	$0 - 10^{-3}$
ℓ_{ga}	Gaussian scale parameter $= \sqrt{2}\sigma_{ga}$	196m
ℓ_{pl}	inverse power law scale parameter	203m
a	Inverse power law exponent	3.3
t	Time-step	1 day
T	Days in Sporulation peak June - September	122
T_{SP}	Peak sporulation ϕ_2	end of July
T_{LS}	Peak ash leaf-shedding	mid-November
α	Lattice constant	5m
L	Lattice dimension (i.e. an $L \times L$ patch)	200 - 2000
A	Domain area (defined by $\alpha L \times \alpha L$)	1km^2 - 20km^2
λ	Mean infectious life-time	5 years
R_0	Mean reproduction number	$0 - 20$

Table 6.2: Parameters used in the SEIR model of ash dieback. The dispersal parameters are taken from [Grosdidier et al., 2018] and the typical tree densities of ash are informed from by [Hill et al., 2017].

leaves in following autumn after infection [Gross et al., 2014a]. The sub-compartments $(E_1, E_2, .., E_n)$ represent the same biological state of latently infected, the index n is merely included for convenience to highlight the fact that in this model, the mean latent period is one year and new secondary infections are produced during the $(n+1)^{th}$ sporulation peak—as elaborated in section 6.1.1 below.

Transitions: $E_n \to I_{n+1}$

As defined here, latently infected trees (in E_n) transition into the infectious compartment (I_{n+1}) during their first seasonal leaf shedding following infection, between autumn and winter. Then, infected ash trees disperse infectious leaf litter each year following infection until its eventual demise. This dynamic assumes that all ash are susceptible and *H. fraxineus* spreads to the xylem after infecting ash leaves. In reality, a variety of genetic and environmental factors determine the ability of ADB to invade an ash[3].

A Gaussian distribution represents the onset of ash leaf-shed, and therefore the transition $E_n \rightarrow I_{n+1}$. Leaf-shed repeats yearly, centred in mid-November (denoted by σ_{LS}) with a two-week standard deviation (σ_{LS}) and repeats yearly. Centring the normal distribution in mid-November, with a two-week standard deviation, ensures that the earliest possible onset of leaf-shed (described by the left-hand tail) begins in September. In the field, the timing of ash shedding their leaves is more complicated and depends on tree age. Observations by [Hietala et al., 2013] suggest that younger ash begin to defoliate in late August, while large dominant ash starts to shed in early October.

Despite a thorough search, research on ash litterfall distributions appears absent from the literature, so choosing a normal distribution centred in mid-November with a two-week standard deviation is ultimately ill-informed. Nevertheless, selecting a normal distribution is supported by observations of leaf litter-fall in deciduous forest [Zhang et al., 2014, Dixon, 1976], which typically follow peaked distributions that repeat yearly[4]. That said, within the SEIR model, a time delay exists between the transition $E_n \rightarrow I_{n+1}$ in autumn/winter and the sporulation function $\phi(t)$ becoming sufficiently large in summer; so we can afford some degree of flexibility in the exact time scale. From a mathematical standpoint, provided that latently infected ash transitions into the infected compartment before the sporulation season, the epidemic spread in this model will remain the same.

Transitions: $I_n \rightarrow R$

The last transition to consider is from infected to removed $I_n \rightarrow R$. Given a 95% mortality rate, ADB can be regarded as lethal [Hill et al., 2019]. Therefore, we assume an eventual transition to the R compartment once an ash tree becomes infected. As mentioned above,

---

[3] The pathogen *H. fraxineus* can infect ash leaves regardless of tolerance. However, only the minority of tolerant individuals can prevent inoculum from spreading to the xylem.

[4] For example, research conducted on wetland forests in South Carolina demonstrates that the temporal pattern of litter-fall peaks yearly between August and November, albeit with some variation [Shure and Gottschalk, 1985].

the picture is more complex in reality. For example, stress-free trees in urban settings can survive for long periods if pruned [Marciulyniene et al., 2017], and a low minority of healthy tolerant trees can fight off the infection. Additionally, other outlying edge-cases can contradict this assumption, including laboratory experiments that have shown leaf-shed can result before the infection has the chance to spread through to the petiole following a *'massive ascospore inoculation'* [Gross et al., 2014b].

Ash dieback affects trees of all ages, with younger ash being more susceptible while larger, mature ash appear more tolerant. As a first approximation, infected ash trees were chosen to have exponentially distributed lifetimes with a mean of five years, see table 6.2. Experimental observations of ash mortality after years of infection support this decision. In particular, reports of 5% mortality after two years of infection [Keßler et al., 2012], 75% mortality within five years [Langer et al., 2015] and no observations of infected ash surviving beyond 15 years [Wylder et al., 2018] provide some guidance towards an approximate time-scale. The precise probability distribution describing $I_n \rightarrow R$ is, to my knowledge, non-existent in the current literature.

6.1.2 Dispersal parameterisation

Dispersal was informed by data collected in France by [Grosdidier et al., 2018]. The study conducted by [Grosdidier et al., 2018] tracked ADB ascospores about known sources of infection—reviewed in Chapter 2. The authors considered a Gaussian and inverse power law kernel of the form:

$$Pr(a,r) = \frac{1}{\pi \ell_{ga}^2} \exp\big[-\frac{r^2}{\ell_{ga}^2}\big] \tag{6.2}$$

and

$$Pr(a,r) = \frac{(b-1)(b-2)}{2\pi \ell_{pl}^2}\big[1+\frac{r}{\ell_{pl}}\big]^{-b} \tag{6.3}$$

where parameters ℓ_{ga} and $(\ell_{pl},\ b)$ are given in table 6.2. Notably, the researchers measured dispersal parameters at local and regional spatial scales, over two orders of magnitude between $[0 \sim 1\text{km}]$ and $[10 \sim 100\text{km}]$ respectively. Since the SEIR model is small-scale, dispersal in the SEIR model only incorporates the local-scale values provided by [Grosdidier et al., 2018].

Dispersal data collected by [Grosdidier et al., 2018] relate to a French landscape and environmental conditions. Thus, we implicitly assume that ADB disperses comparably in France and GB when using these parameter values. However, we might expect differences in ADB dispersal over large spatial scales when long-distance dispersal and differences in French wind patterns become observable[5]. Notwithstanding, it is entirely reasonable to assume ADB dispersal across smaller scales $\leq$ 1km is indistinguishable in both landscapes, and we can comfortably use the local-scale parameters provided by [Grosdidier et al., 2018].

6.1.3 Normalising β between models

Undesirably, the scale of β varies between model variants, due to the difference in area under the curves of ϕ_1, ϕ_2, D_1 and D_2. Similarly, this behaviour was witnessed in Chapter 5 when the NLM was simulated with different dispersal length scales. To constrain each model on the same β-axis, the appropriate normalisation constant (shown in Table 6.1) is once again factored out of the infectivity probability β e.g. for ϕ_1-ga, this takes the form:

$$\beta = \frac{\beta^*}{T\pi\ell_{ga}^2} \in [0,1] \tag{6.4}$$

where β^* is an auxiliary parameter that does not depend on the form of sporulation or dispersal function (provided that both functions are normalisable). Thus, as before, β^* isolates infection pressure to a single parameter. On account of the sporulation function and longer range kernel, β^* in equation 6.4 assumes a larger value when compared to β^* in Chapter 5.

---

[5] In addition, the long-distance dispersal of ADB likely contains artefacts of human-mediated transport over these large regional scales, which goes beyond this Chapter's scope. However, the reader can see [Harwood et al., 2009] for an example of how to include LDD (i.e. by human-mediated trade) in a large-scale botanical disease model.

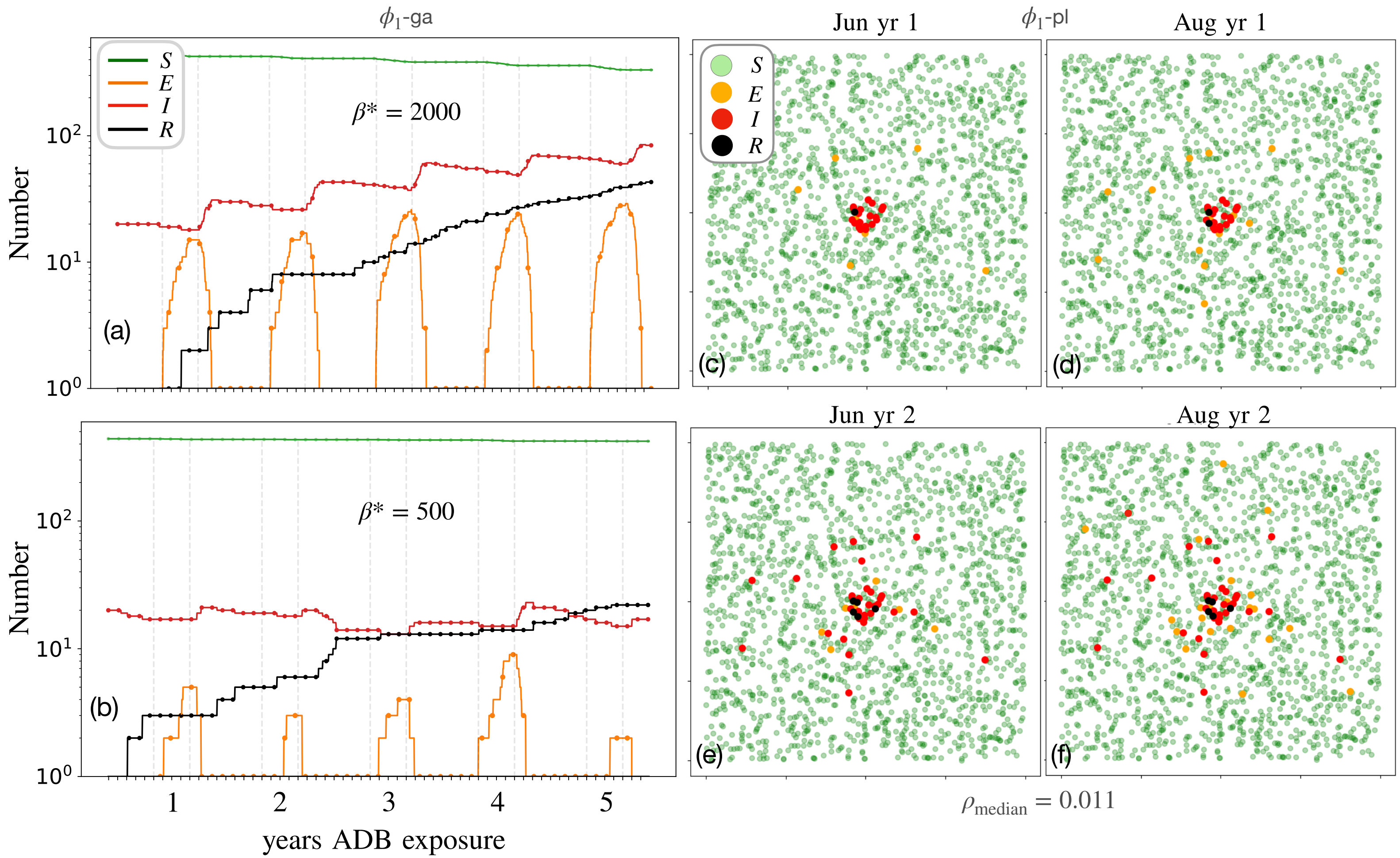


Figure 6.2: Simulating the SEIR model over the median ash density, i.e. $\rho_{\text{med}} = 0.011$. In all panels, a small clump of 20 infected hosts seeds the domain centre at time $t = 0$. (a-b) The number of ash in each SEIR compartment is plotted over five years for the model ϕ_1-ga. Panels (a) and (b) depict simulations above ($\beta^* = 2000$) and below ($\beta^* = 500$) the epidemic threshold, respectively. (c-f) The spatial progression of disease in a 2km × 2km domain over a full year for model ϕ_1-pl; infectivity is fixed to $\beta^* = 3000$, well above the threshold. From the first sporulation season (c)-(d) to the second (e)-(f), secondary infections scatter throughout the whole domain, in contrast to the wave-like behaviour witnessed for small dispersal length scales in Chapter 5.

6.2 Seasonal SEIR model behaviour

The essential SEIR model behaviour is shown in Figure 6.2, simulated with the median ash density in Great Britain—according to [Hill et al., 2017]. For each year after the outbreak, a rise in the number of latently infected ash in E can be seen during summertime sporulation, June-September, followed by a rise in the number of infectious ash in the autumn-winter. All model variants display the same pattern of seasonal behaviour.

The number of ash in the SEIR compartments is shown in Figure 6.2(a-b) for two infectivity parameters in a 1km × 1km domain. Figures 6.2(a-b) depict two scenarios above and below the epidemic threshold for model ϕ_1-ga; the compartments are plotted over 5 years with infectivity (β^*) parameters shown. In Figure 6.2(a), the number of ash in S decline, shown by the green line, and during seasonal sporulation, large spikes in the number of latently infected ash can be seen in orange. Then, latently infected ash transition into I during autumn and winter, as shown by the seasonal rise of I in green. For infectivity parameters below the epidemic threshold, Figure 6.2(b), S remains approximately constant as I slowly declines. Interestingly, Figure 6.2(b) demonstrates persistence-like behaviour in model of ADB; whereby, even if the epidemic parameters are below the threshold, the fungus may survive for long periods. In general, persistence in plant-based pathogens is one aspect that complicates epidemic control—also discussed in Chapters 3-5.

The spatial progression of ADB in the SEIR model is shown in Figures 6.2(c-f) over a full year. The simulation in Figure 6.2(c) begins with an initial condition of 20 infected ash centrally distributed in the host landscape during March (not shown). At this point in the year, fungal fruiting bodies on infected leaf litter will be preparing to release ascospores. During sporulation, secondary infections are produced, and latently infected ash spread throughout the domain, depicted by the orange dots in Figure 6.2(c-d). At any time step, the chance of removal to the R compartment is non-zero, demonstrated by the small number of black dots in Figures 6.2(c-d). After the sporulation season ends in September (not shown), latently infected ash begin to transition into the I compartment; eventually, all latently infect ash become infectious during the following season, reflected by the increased number of red dots in Figure 6.2(e). The cycle will continue when the next cohort of secondary infections are produced in the following summer when the sporulation function next becomes non-zero.

6.2.1 Sporulation: time-varying infectivity

Time varying infectivity rates are an important concept in epidemiology, this is true of epidemics in both human/animal [Svensson, 2007, Liu and Stechlinski, 2012] and botanical populations [Suffert and Thompson, 2018, Leclerc et al., 2014, Cunniffe et al., 2012]. In particular, ADB is known to have a seasonal life-cycle, and time-varying infectivity [Grosdidier et al., 2018, Hietala et al., 2013]. The peak of ADB infectivity occurs during summertime sporulation when fruiting bodies on shed litterfall release ascospores.

Two contrasting sporulation functions, $\phi_1(t)$ and $\phi_2(t)$, were used to model the time-dependent ADB ascospore production and demonstrate robustness in the approach. The choice of sporulation functions were inspired by the modelling work of [Cunniffe et al., 2012] and [Segarra et al., 2001] (reviewed in section 2.2). Sporulation functions are described by a step function and normal distribution located at the midpoint between June and September:

$$\phi_1(t) = \begin{cases} 1 & t \in [\text{June},\ \text{September}] \\ \\ 0 & \text{Otherwise} \end{cases} \tag{6.5}$$

and

$$\phi_2(t) = \exp\big[-\frac{(t - T_{SP})^2}{2\sigma_{SP}^2}\big] \tag{6.6}$$

where T_{SP} is taken to be the mid-point of June-September (i.e. late July/early August) and σ_{SP}^2 is a standard deviation of two weeks. As discussed above, the choice of sporulation peak T_{SP} and standard deviation σ_{SP} together inform the earliest transitions $S \to E$ for ϕ_2. Due to the seasonality and perennial nature of ADB, both $\phi_1(t)$ and $\phi_2(t)$ repeat yearly, becoming non-zero during the months of June and September[6]. For ϕ_2, the chance

---

[6] Variations in ADB sporulation have been noted between European countries [Gross et al., 2014b],

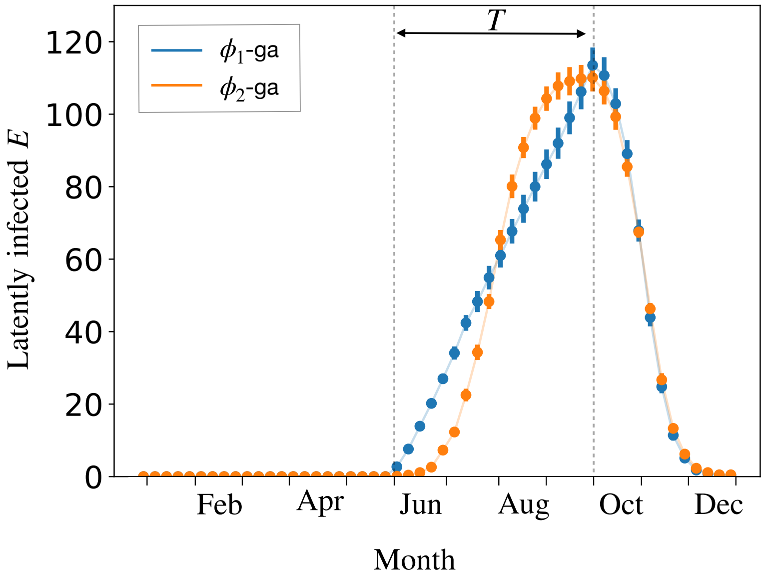


Figure 6.3: One year simulations contrasting the sporulation models $\phi_1(t)$-ga and $\phi_2(t)$-ga. For both sporulation models, the number of expected latently infected trees is plotted against time. Each plot depicts an ensemble of 10 replicate simulations and error bars outline the standard error. An auxiliary infectivity, of value $\beta^* = 2000$, effectively matches the epidemic impact between models—indicative of the area under each curve. However, each sporulation function admits a different shaped curve.

of new secondary infections outside of June-September is trivially small.

Figure 6.3 contrasts sporulation models by computing the number of ash transitioning into the E compartment for one season over an ensemble of size 10. More specifically, by simulating the spread of disease and counting the number of infections that result over one year and plotting against time. Simulations started with 100 infected ash distributed randomly throughout a 5km × 5km domain at the mean GB ash density in January (i.e. time $t = 0$). Both models show a rise in the number of infected ash during the sporulation season. The infectivity rate for function ϕ_2 can be seen to vary, in contrast ϕ_1 is uniform—both dispersal models ϕ_1-pl and ϕ_2-pl demonstrated the exact behaviour. Although infectivity rates differed, the normalised infectivity β^* ensured the same approximate area under each curve and, therefore, epidemic impact.

Both $\phi_1(t)$ and $\phi_2(t)$ aim to mirror the seasonal time-dependence of *H. fraxineus* sexual reproduction. In the seasonal SEIR model, a time-varying infectivity is achieved by

---

along with the potential for early-onset sporulation in the face of favourable environmental conditions. Although, the most generally agreed upon sporulation period is thought to be from June to September.

multiplying the infectivity parameter with a sporulation function $\beta\phi(t)$. In line with the parsimonious approach undertaken thus far, $\phi_1(t)$ and $\phi_2(t)$ are generic and aim to capture the essential dynamics of ADB spread. However, a more complex model could aim to fit $\phi(t)$ to spore-trapping data, which could resemble a Gaussian-k function (e.g. see Figure 2. in [Grosdidier et al., 2018]). Although different sporulation models lead to the same effective behaviour (as discussed more below in section 6.2.2), non-uniform infectivity rates for ADB have a more robust basis in the literature [Grosdidier et al., 2018, Cunniffe et al., 2012, Hietala et al., 2013, Segarra et al., 2001]. So, the function ϕ_2 can be considered more representative due to the non-uniformity and infectivity peak.

6.2.2 Tree mortality

Ash mortality due to ADB has been well-researched in different European countries, frequently over long-running 10-20 year experiments. In this section, tree mortality is studied over long time scales to give insight into the scale of the epidemic, pathogen invasiveness and domain sensitivity. However, host demography was neglected from the model, which is desirable for modelling the spread of pathogens targeting long-lived hosts [Swinton and Gilligan, 1996b]. Therefore, the long-running simulations presented in this section approximate the spread of ADB over considerable time scales to first-order.

Figures 6.4(a-b) display the proportion of ash in S, I and R over 30 years for the model ϕ_1-ga with infectivity parameters $\beta^* = 1000$ and $\beta^* = 1500$, respectively. The simulations take place on a 1km × 1km domain. Simulations reliably produced epidemics and spread through the entire domain for these parameter values. Thus both β values can be considered above the threshold for an epidemic. Initially, all simulations begin with a small number of infected ash at the domain centre. As expected, the higher infectivity in Figure 6.4(b) produces an outbreak that spreads through the domain much quicker. Although, all simulations above the epidemic threshold will eventually reach 100% tree mortality on average.

The seasonal SEIR model produced a characteristic sigmoidal s-shaped mortality curve for

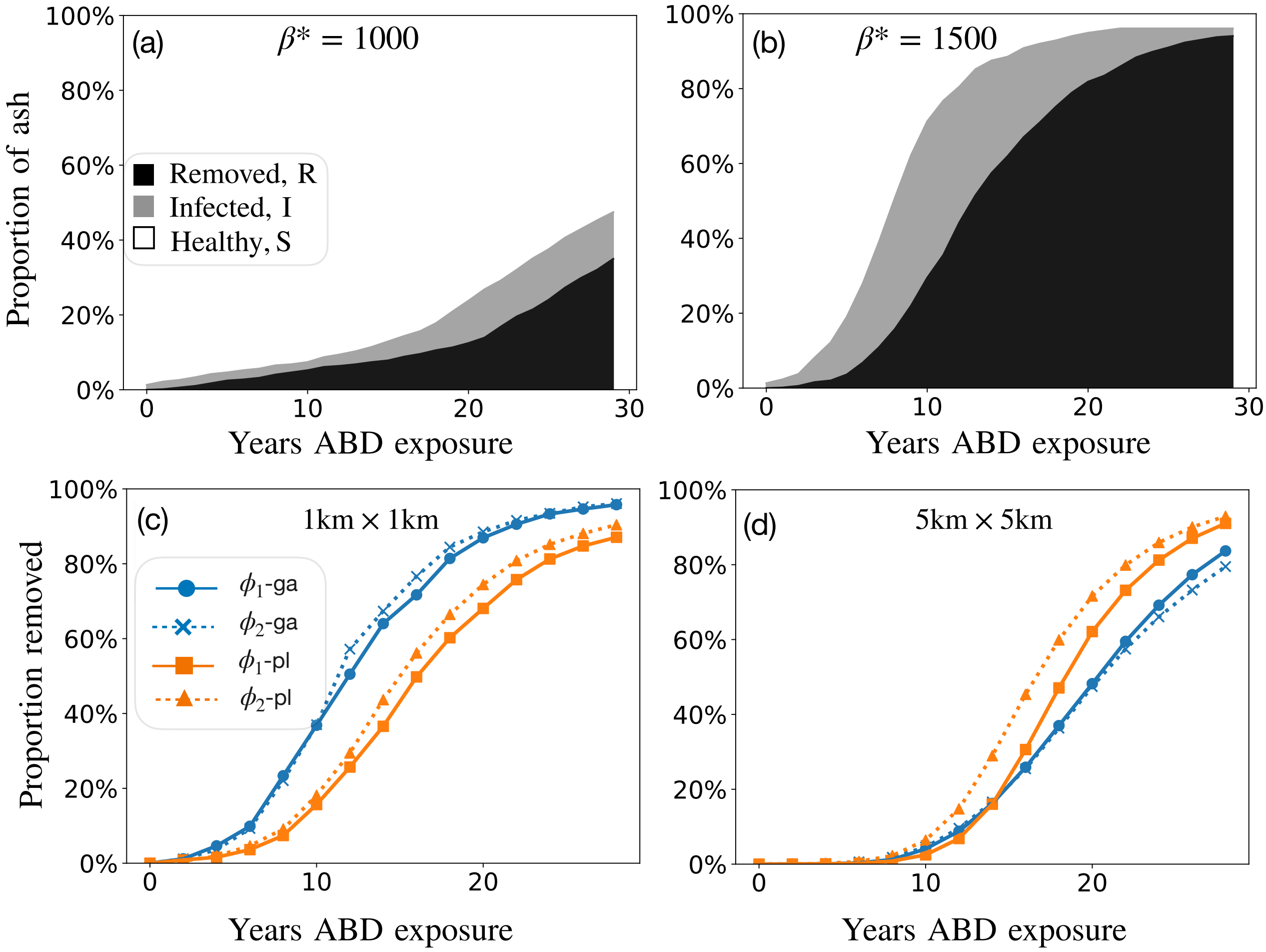


Figure 6.4: Epidemic-scale in the SEIR model, as measured by the proportion of ash in S, I and R (a-b) The proportion of ash in S, I and R is shown over 30 years of ADB exposure for ϕ_1-ga. Two different values of infectivity, $\beta^* = 1000$ and $\beta^* = 1500$, increase the proportion of trees in I and R. (c-d) The proportion of ash removed in all models after years of exposure with infectivity $\beta^* = 1000$. Differences between dispersal models are demonstrated by disparate epidemic-scales—indicated by the height between blue and orange curves.

high mortality proportions (asymptotically approaching 100%), shown in Figures 6.4(b), this contrasts with the logistic ADB mortality curves presented by [Coker et al., 2019] that, on average, saturate to $\sim$ 60%. Other authors have witnessed similar s-shaped mortality curves, e.g. [Lõhmus and Runnel, 2014]. Here, not including host demography precludes host-pathogen coexistence and prevents the SEIR model from saturating below 100% mortality—as explained in more depth by [Cunniffe et al., 2012] in their manuscript. Therefore, failing to reach a steady-state for long running simulations represents a limitation in the seasonal SEIR model of ADB.

Figures 6.4(c-d) show the mortality percentage on two different domain sizes for each model with a single infectivity parameter $\beta^* = 2000$, well beyond the epidemic threshold. For all models, the larger domain size effectively reduced the proportion of removed ash over the 30 year period because more trees populate the domain, shown in Figure 6.4(c). Figure 6.4(c) confirms that on a smaller 1km × 1km domain, the more localised Gaussian dispersal models give rise to a more significant epidemic impact. This behaviour can be understood by noting that for inverse power law spread, secondary infections are on average likely to be under-counted by virtue the of fat-tailed dispersal kernel extending beyond the domain boundary. Increasing the domain size to 5km × 5km brings both dispersal models into a closer agreement on the epidemic scale. Nevertheless, directly comparing the epidemic scale between models is difficult considering the relationship between domain size and dispersal.

Domain size is known to play an integral role in shaping the evolution in spatially explicit population growth models [Tang and Lin, 2011]. A related topic, 'plant-pathogen invasiveness and field-size', was investigated by [Mikaberidze et al., 2016] in the context of crop disease. In particular, [Mikaberidze et al., 2016] demonstrated that R_0 saturates to a maximum for a suitably large domain size, beyond which increasing the domain size had no effect on R_0. Therefore, choosing a suitable domain size for each dispersal mode remains critical to capture the epidemic scale accurately. In the next section, a definition of R_0 is outlined for the seasonal SEIR model, and we revisit the topic domain size.

6.3 Defining an R_0

Before properly investigating the SEIR model, the pathogens' ability to invade must be defined. Tree mortality, shown in Figure 6.4, gives insight into the final-size epidemic and the time-scale of spread. However, computing tree mortality requires long simulation run-times spanning years which is problematic given that host-demography is not included in the model. Moreover, as we look to scale up the SEIR model over large areas within GB, running tree mortality simulations over long periods becomes increasingly computationally expensive.

Consequently, the basic reproduction number will be employed to navigate these computational challenges. As we saw previously, through sections 5.3-5.5, the system can be characterised by a basic reproduction (R_0) number that corresponds to a threshold, above which epidemic severity drastically increases.

From this point on, unless otherwise stated, the method for calculating R_0 is based on the average number of secondary infections resulting from the first-generation of infected hosts, as per definition 5.5.1. However, we assume here that secondary infections characterise the transition $S \rightarrow E$, and the number of first-generation secondary infections ($R_0^{(1)}$) is measured over the mean lifetime of infected ash, i.e. five years. The first-generation contact-traced secondary infections $R_0^{(1)}$ will be denoted by R_0 for brevity.

6.3.1 Computing R_0: initial conditions

This section will investigate initial conditions for just ϕ_1-ga and ϕ_1-pl, as differences between sporulation models were negligible. Initial conditions play a role in shaping the final value of R_0 and produce distinct model behaviours. At $t = 0$, the following initial conditions can seed the domain: IC1) infected hosts occupy the domain centre IC2) infected hosts are scattered randomly throughout the domain. Figure 6.5 contrasts the contact-traced R_0 for first-generation infected ash over five years between the models ϕ_1-ga and ϕ_1-pl. Although the ensemble average $\overline{R}_0$ (i.e. the horizontal black line) compared

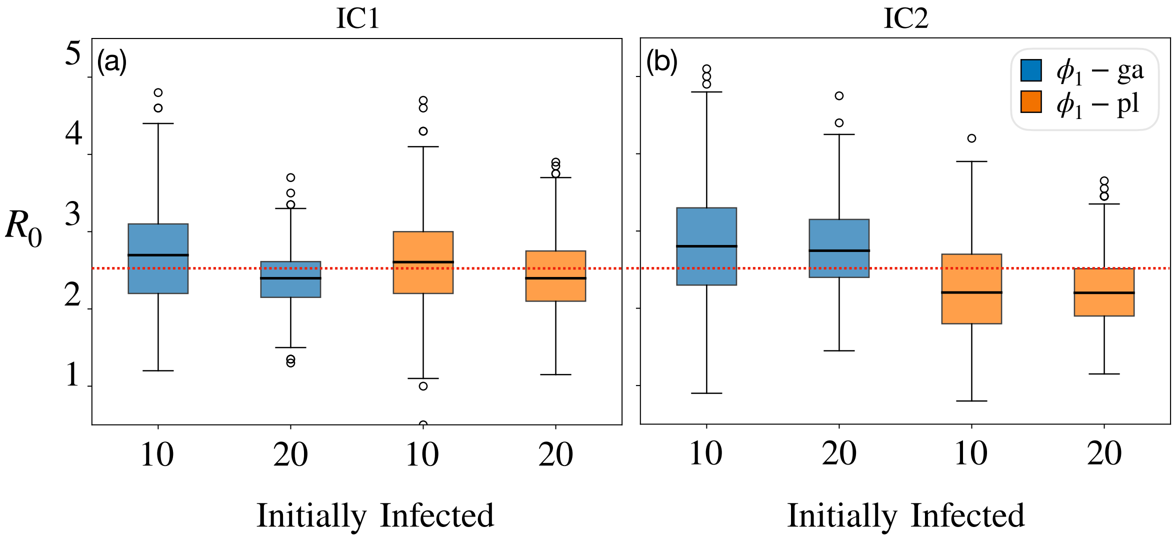


Figure 6.5: The effect of four different initial conditions on R_0 are compared for the Gaussian and inverse power law models, ϕ-ga (blue) and ϕ-pl (orange). Each box describes R_0 simulated over an ensemble of size $N = 500$. The mean R_0 for all simulations is depicted by the horizontal red line. (a) At $t = 0$, between 10 and 20 infected hosts are placed at the domain centre (denoted IC1) for ϕ-ga (in blue) and ϕ-pl (in orange). (b) At $t = 0$, between 10 and 20 infected hosts are randomly scattered throughout the domain (denoted IC2) for both models. The mean value of R_0 compares most similarly using IC1. Epidemic parameters for each ensemble: $\beta^* = 1500$, $\rho_{avg} = 0.017$, on a domain of size 2km × 2km.

similarly for each initial condition, differences can be seen in the statistics.

Figure 6.5 shows an ensemble of R_0 values for IC1 (a) and IC2 (b) with 10 and 20 initially infected ash with epidemic parameters well above the threshold[7]. A more localised dispersal kernel in ϕ_1-ga produces a value of R_0 that is generally higher for IC2 and lower for IC1. Adjacently infected hosts at $t = 0$ reduce the density of susceptible ash within a relatively small neighbourhood—up to a maximum radius of approximately $\approx 3\ell_{ga}$. When several infected trees are located nearby, as for IC1, R_0 is generally smaller because the surrounding neighbourhood quickly becomes saturated with infected trees, limiting the number of new secondary infections. The effect of R_0-saturation for model ϕ_1-ga is most apparent in Figure 6.5(a) when a lower value of $\overline{R}_0$ results by increasing the number of infected ash to 20.

In contrast, initial conditions had the opposite effect on inverse power law based epi-

---

[7] Following the findings of Chapter 5, a small number of initially infected trees seed the domain at $t = 0$; having more than one infected tree helps to reduce initial stochasticity, which thereby lowers the number of early pathogen extinction events.

demics. For inverse power law spread, Figure 6.5(a-b) demonstrate that R_0 is generally lower for IC2 and higher for IC1. Fat-tailed dispersal models show a higher sensitivity to the domain boundary. In particular, domain-sensitivity is demonstrated in 6.5(b), where more infected hosts are located closer to the domain edge. These observations can be explained by realising that fat-tailed dispersal kernels are more likely to extend beyond the boundary. Therefore, for the same normalised value of β^*, inverse power law models are more likely to under count R_0 locally if measured inside a smaller domain. Notwithstanding, secondary infections induced by ϕ_1-pl spread over a wider area, and the effect of R_0-saturation is far less than ϕ_1-ga.

The dashed red line depicts the aggregate mean R_0 over all initial conditions. Boundary effects and domain sensitivity were the least when using IC1, i.e. we can see that IC1 produced the least difference in R_0 in panel (a). Conversely, panel (b) outlines a more considerable discrepancy in R_0 between dispersal models for IC2. Therefore, going forward, simulations are chosen to evolve from IC1 because dispersal models agreed most closely.

6.3.2 Domain size L and R_0

Before the small-scale SEIR model of ADB can be spatially scaled up over GB, a suitable spatial and temporal scale must be chosen to measure R_0. The purpose of this section is to compute R_0 over different sized domains from IC1, i.e. a small number of infected trees located at the domain centre at $t = 0$. Choosing a sufficient domain length (L) is desirable to accurately capture pathogen invasiveness within each model. Following the arguments laid out in section 6.1.1, R_0 values are determined over the mean lifetime of infectious ash, i.e. $\lambda = 5$yr. According to their exponentially distributed lifetimes, infected ash can survive for more extended periods, although this accounts for a decreasingly small the host population.

As remarked earlier, studies have shown that R_0 for crop disease can depend on the field size [Mikaberidze et al., 2016] and saturates for a sufficiently large field. Here, we have a similar scenario: R_0 should be computed inside a sufficiently large domain, such that further domain size increases yield the same value of R_0 for repeated simulations. In contrast, running simulations inside a smaller domain will underestimate the reproductive ratio, and consequently, overall epidemic impact [Tildesley and Keeling, 2009, Cunniffe et al., 2012].

Figure 6.6 reveals how both dispersal models relate to the domain-size. By counting the number of secondary infections that result from the first-generation of infected ash (at $t = 0$), we can plot a distribution revealing how far away infections are likely to be produced. Figures 6.6(a-b) show the number of secondary infections induced a distance D away from each infected source over 500 repeated simulations. Distributions for three different domain sizes are shown in Figure 6.6(a-b).

Unsurprisingly, Figure 6.6(a) and (b) reflect the Gaussian and inverse-power law dispersal kernels. However, Figure 6.6(a-b) also show that secondary infections are unlikely to occur near first-generation infected ash. A low secondary infection count close to infectious sources reflect the average space between hosts, set by ρ. A higher density host distribution increases the relative proportion of induced secondary infections close to the source < 0.1km.

From Figure 6.6(a-b) becomes visually apparent why domain sensitivity varies between dispersal models. In Figure 6.6(a), the number of infections is reduced for Gaussian dispersal at $L = 100$, indicated by the smaller distribution tail in blue. Although, increasing the domain length to $L = 300$ produced no observable difference to $L = 200$. In all Gaussian simulations, no secondary infections were witnessed beyond 0.6km, in line with the results from [Grosdidier et al., 2018]. On the other hand, a small number of secondary infections in the model ϕ_1-pl can be seen up to the domain edge 1.5km away. Thus, if the domain size is below a minimum value of L, epidemics that observe inverse power law dispersal will be lowered due to an under-estimated total number of secondary infections. For the exact value of β^*, ϕ_1-pl can be seen to travel much further than ϕ_1-ga and echos the difficulty of controlling fat-tailed pathogen dispersal [Cunniffe et al., 2015b],

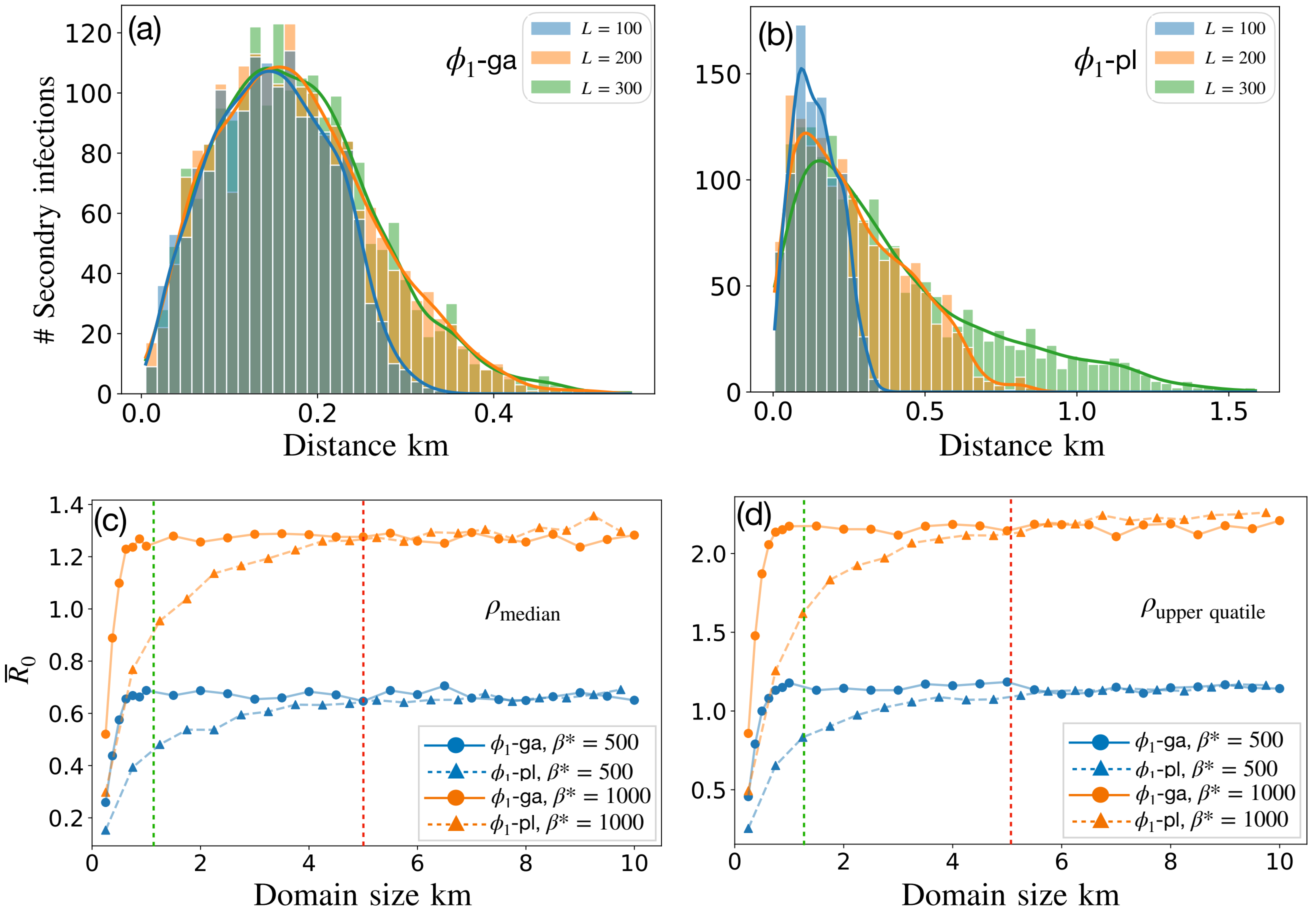


Figure 6.6: The relationship between R_0 and the domain length, L. For all panels, simulations run over five years or until the pathogen becomes extinct, whichever comes first. (a-b) Distributions show the number of secondary infections induced by distance over an ensemble of 500 repeats for Gaussian and inverse power law dispersal kernels. Secondary infections are computed inside three different domain lengths of $L \in [100, 200, 300]$ (i.e. 500m × 500m, 1km × 1km, 1.5km × 1.5km) for epidemic parameters above threshold, $\beta^* = 1000$ and $\rho_{avg} = 0.017$. Inverse power law spread demonstrates a higher domain sensitivity due to the fat-tailed kernel extending beyond the domain edge. (c-d) An ensemble-averaged reproductive ratio $\overline{R}_0$ is computed from 500 repeated simulations over different domain lengths, up to a maximum of $L = 2000$ or 10km × 10km. The value of $\overline{R}_0$ is gauged at the median and upper quartiles of ash tree densities, (c) and (d), respectively, and at two infectivities shown in blue and orange. The value of R_0 saturates at around 1km × 1km for Gaussian-based dispersal kernels and 5km × 5km for inverse power law dispersal kernels.

and more broadly, LDD.

Figures 6.6(c-d) show how the mean value of $\overline{R}_0$ for each dispersal model saturates for a critical value of L, for two domains at the median and upper quartile of ash density $\rho_{med} = 0.011$ and $\rho_{uq} = 0.019$ (a) and (b) respectively. Simulations for two infectivity parameters $\beta^* \in [500, 1000]$ are shown in blue and orange. The computed value of $\overline{R}_0$ shows the aforementioned characteristic increase, up to a maximum saturation value. Each infectivity and tree density combination produced a similar saturation point of $L \approx 1$km for ϕ_1-ga and $L \approx 5$km for ϕ_1-pl, the vertical green and red lines, respectively.

From Figures 6.6(c-d), a suitable domain length L can be ascertained for each dispersal model, found to be 1km × 1km for models ϕ_1-ga and ϕ_2-ga, and 5km × 5km for models ϕ_1-pl and ϕ_2-pl. Moving onward, a proper characterisation of the GB ash tree canopy cover data set produced by [Hill et al., 2017] will be conducted, before we determine R_0 as a function of host density in section 6.4.2.

6.4 Constructing R_0-maps over Great Britain

In this section, R_0 values of the small-scale SEIR model of ADB will be projected onto the host distribution of ash, given by [Hill et al., 2017], to create landscape-level R_0-maps over Great Britain. Doing so will permit the investigation of a novel control strategy in Chapter 7 and allow the local-scale epidemic severity, based on tree-to-tree interactions, to be efficiently scaled over large areas.

6.4.1 Ash host distribution

Ash densities were parameterised by ash the abundance data provided by [Hill et al., 2017]. Previously, the oak canopy cover dataset given by [Hill et al., 2017] was used in Figure 4.9 alongside a toy model of landscape-level tree disease. The canopy cover datasets produced by [Hill et al., 2017] combine several data sources that partly cover Great Britain, regression methods then extrapolate canopy cover over the whole of Great Britain. Conveniently, ash happened to be among the most accurate data sets given by [Hill et al., 2017].

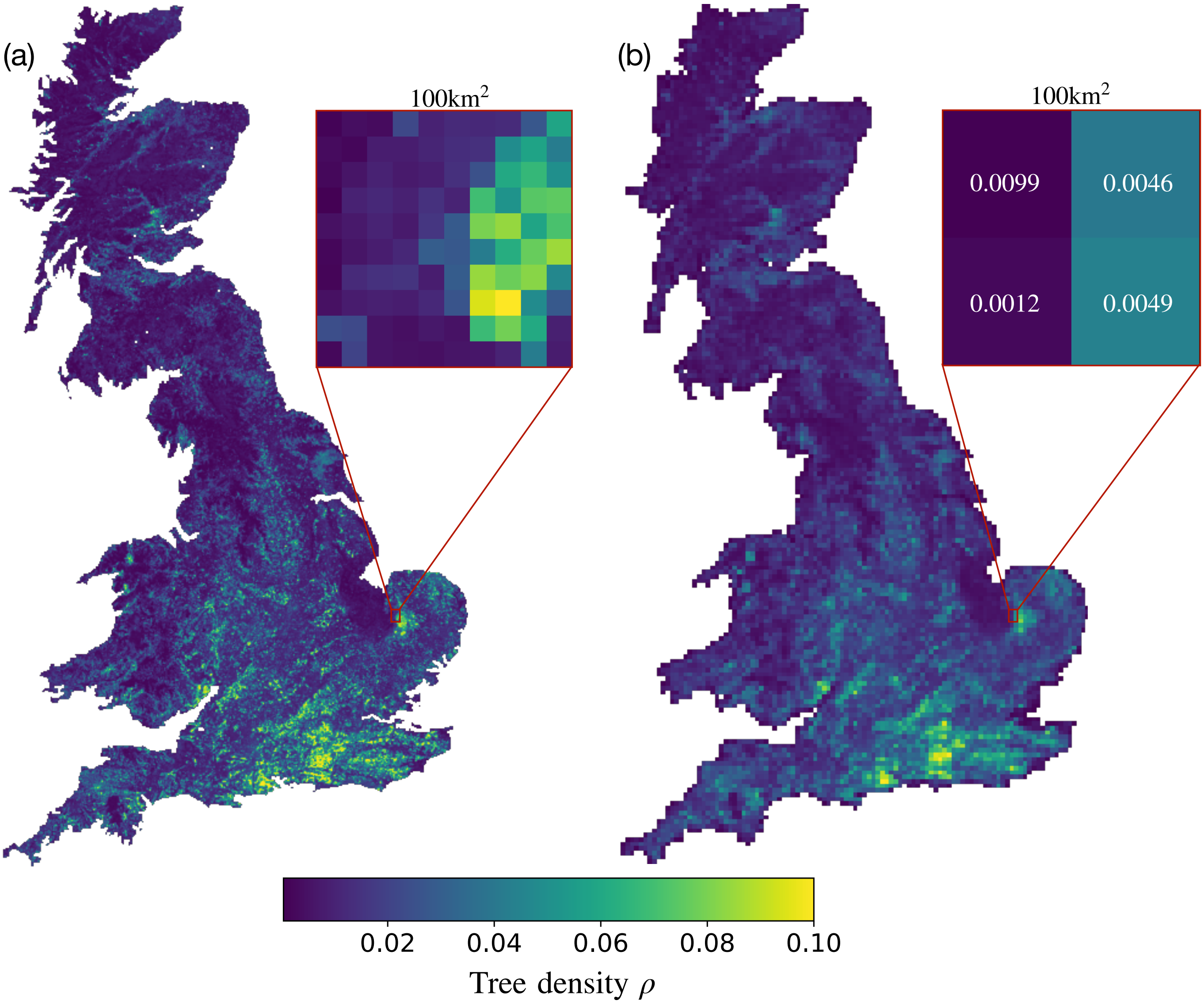


Figure 6.7: The ash canopy cover data, as modelled by [Hill et al., 2017], is converted into a map of tree density. (a) A map of ash densities at the original resolution of 1km × 1km, the inset consisting of 10 × 10 pixels (b) A coarse-grained map of ash densities at a resolution of 5km × 5km, the inset consists of 2 × 2 pixels. Both insets show the same 100km^2 area, and illustrates how coarse-graining the host distribution results in a loss of spatial structure. A small number of densities over 10% were excluded from the density-map.

For a more in-depth description of the methods used by [Hill et al., 2017] and a review of host data in general, see Chapter 4.2.

The modelled ash canopy cover data had to undergo minor modifications to complement the SEIR model. Firstly, the raw abundance values were re-scaled into a dimensionless tree density ρ. The exact process was outlined in Chapter 5, i.e. by converting the units ha/km^2 to kilometre-squared of ash cover per kilometre-squared of land. Secondly, the domain resolution has to be re-scaled to reflect the spatial scale of local wind-borne dispersal, as parameterised by [Grosdidier et al., 2018]. Lastly, the small number of patches with exceedingly high densities were capped to $\rho_{max} = 0.10$, thus forming a hard upper limit in the density map.

Figure 6.7(a) shows a density map of ash, at the original resolution of 1km × 1km, produced from abundance data given by [Hill et al., 2017]. The inset shows a block of 10 × 10 pixels, each of size 1km × 1km. In the small-scale Gaussian dispersal-based SEIR model, a domain size of 1km × 1km was shown sufficient to measure the average R_0 over a five-year period—demonstrated by Figures 6.6(c-d). However, the inverse power law models required a larger domain size of 5km × 5km to prevent underestimating epidemic severity. Figure 6.7(b) shows the host distribution coarse-grained to a resolution of 5km × 5km pixels. The insets of Figures 6.7(a-b) compare the same region. Although, in Figure 6.7(b), pixels are effectively averaged and coarse-grained to larger 5km × 5km patches. The resulting domain is smoother and therefore losses spatial structure.

From Figures 6.7(a-b), the south of England contains the highest concentration of high-density ash patches, and ash become progressively less abundant in Scotland and coastal locations, in western Wales, for example. A higher density of ash can be expected to yield a higher number of secondary infections, in line with the results of section 3.3 and Chapter 5.

6.4.2 Tree density and R_0

Figures 6.8(a-c) illustrates the relationship between R_0 and ash density. Figure 6.8(a) shows a PDF (probability density function) of ash densities ρ over the map of Great Britain—overlaid with a KDE shown in red. As noted above, few locations support densities of $\rho = 0.10$ and over[8]. Between the limits of $\rho \in [10^{-2}, 10^{-1}]$, the PDF follows a power law of the form $\sim \rho^{-k}$, as evident from linearity in the logarithmic inset axes. The distribution had a fitted exponent of $k = 1.90$, shown by the black line.

Figures 6.8(b-c) contrast the behaviour between R_0 and tree density ρ for linear and peaked sporulation functions, ϕ_1 and ϕ_2 respectively. Values of R_0 were ensemble-averaged over 100 repeated simulations. For the two values of infectivity, $\beta^* \in [500, 1000]$ shown, all model variants display the same linearity between R_0 and ρ —also explored in Chapter 5. The dashed orange and blue lines show how high values of β and ρ result in a considerable value of R_0. For high R_0 values, the Gaussian dispersal models ϕ_1-ga and ϕ_2-ga begin to deviate from linearity as density is increased, due to the phenomena of R_0-saturation witnessed in section 6.3.1. Figure 6.8(b) reveals a larger area of shaded grey, suggesting that the sole difference between sporulation models is that ϕ_2-ga deviates from linearity more than ϕ_1-ga.

In Figures 6.8(b-c), the regime of pathogen extinction is indicated by the horizontal red lines, which correspond to a critical 'density threshold' denoted by ρ_c (equivalent to the threshold $R_0 = 1$). When used in conjunction with the data-set from [Hill et al., 2017], Figures 6.8(b-c) represent an appropriate projection of R_0 over the map of Great Britain. Below the threshold $R_0 = 1$, infected ash can still survive and reproduce (i.e. persist below the threshold, as we saw in Figure 6.2(c)), albeit at slower rates. Negating below-threshold patches is a vital assumption—discussed more below.

---

[8] The original 1km × 1km map resolution contained 2.2×10^4 1km^2 data points, with some outlier pixels having densities in the interval $\rho \in [0.10, 0.30]$ which were excluded from the analysis due to A) the increased computational run-time required to simulate the SEIR model of ADB and B) densities beyond 0.10 account for a negligible portion of the overall population.

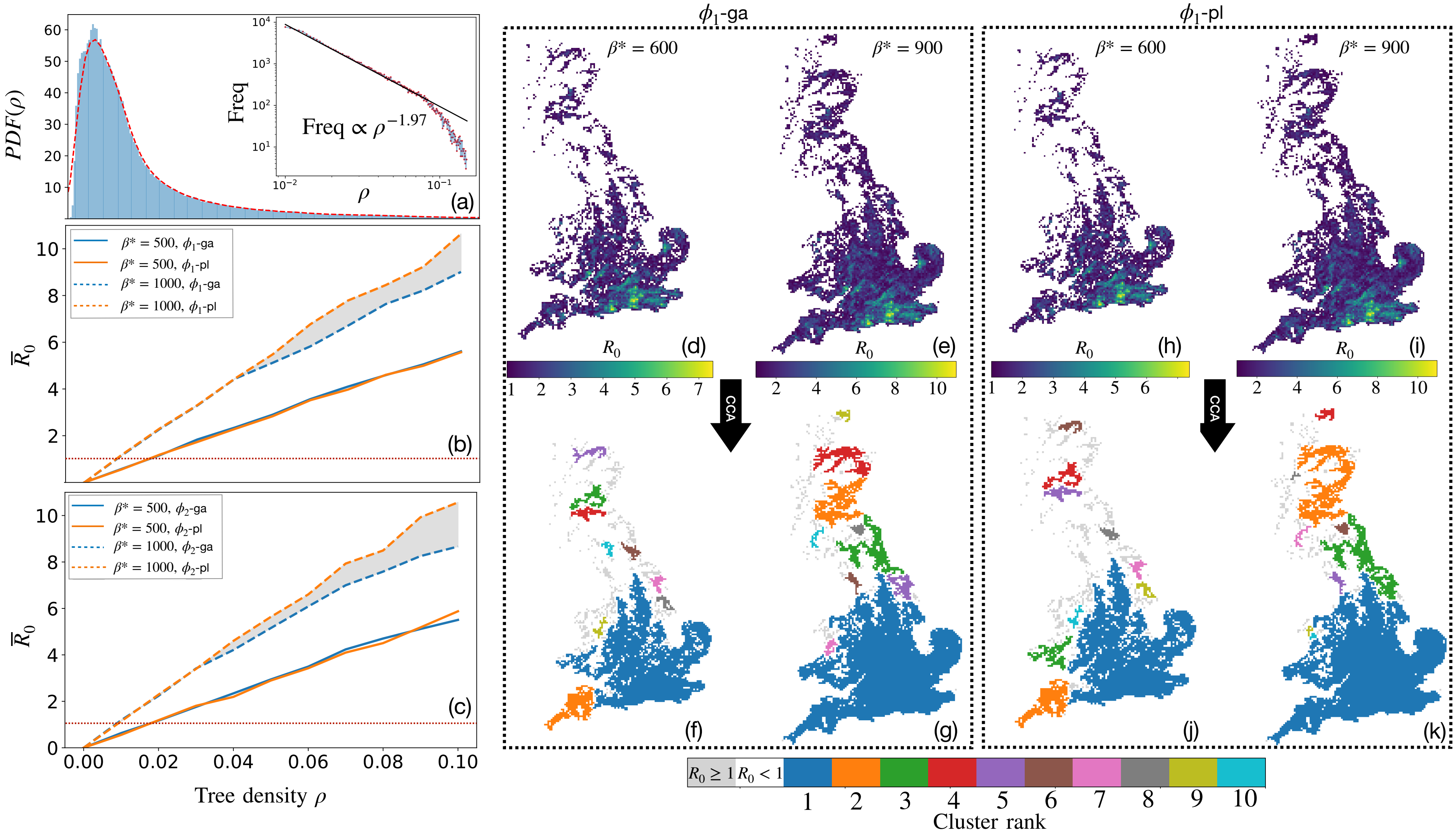


Figure 6.8: The process of generating R_0 maps over Great Britain. (a) The probability density function of ash abundance, as given by [Hill et al., 2017]. (b-c) Ensemble-averaged results of R_0 plotted against tree density for sporulation functions ϕ_1 and ϕ_2, respectively. (d-g) Projecting R_0 values onto the re-scaled 5km × 5km abundance maps for the Gaussian dispersal model. (h-k) The same process of projecting R_0 values onto the distribution for inverse power law dispersal models. Two values of β_{pl} are shown for both Gaussian and power-law. White space depicts patches below $R_0 = 1$. The bottom plots show the top 10 largest connected clusters where $R_0 \geq 1$.

Figures 6.8(d-e) and (h-i) present R_0-values of the small-scale ϕ_1-ga and ϕ_1-pl SEIR models projected onto the host distribution of ash. Both sets of R_0-maps are projected onto the same coarse-grained 5km × 5km re-scaled host distribution. For all variants of the SEIR model, R_0-maps compared similarly—shown for the two values of infectivity, $\beta^* \in [600, 900]$. All locations below the transmission threshold $R_0 = 1$ were given numerical values of zero and are depicted by inland white space. Under the influence of a more infectious pathogen, larger areas of the ash population become susceptible by supporting the growth and reproduction of the pathogen, illustrated by the difference in patch density in Figures 6.8(b-c). Importantly, each R_0-valued pixel portrays the local-scale epidemic impact experienced at that location, predicted from a five-year ensemble-averaged value of R_0.

Suppose that a fitted value of infectivity β^* is found to fall within a standard error of $SE = 250$, i.e. $\beta^* = 750 \pm 250$. In this scenario, epidemic uncertainty is captured by the divergence between dashed and solid lines in Figures 6.8(b-c) and conveniently visualised by differences in the resulting R_0-maps. Although landscape-level heterogeneity and regional susceptibility can be loosely identified from R_0-maps in Figures 6.8(d-i), visualising differences between model variants is untenable due to their similarity. Furthermore, determining which pixels connect to form susceptible clusters is non-trivial. To this aim, the next section will present a means to identify clustering in R_0-maps.

6.4.3 Clustering in the R_0 map

An image processing technique called 'connected component analysis' (CCA) was used to identify and label susceptible clusters and simplify the R_0-map [Samet and Tamminen, 1988, He et al., 2017] . The Python-SciPy package 'ndimage' [Virtanen et al., 2020] was used to implement CCA via the function 'label'. Doing so labelled all susceptible neighbours as connected members of the same cluster, according to a structuring element [Liang et al., 1989]. That is, if two susceptible patches of ash lie within the same neighbourhood, defined by the structuring element, they are connected members of the same cluster[9]. Additional information on structuring elements and CCA can be found in Appendix C.1.

Moore and Von-Neumann neighbourhoods were chosen as structuring elements to classify connected components—a comparative look exploring the differences between Von-Neumann and Moore structuring elements is resumed below in sections 6.4.5-6.4.6. Figures 6.8(f-k) correspond to the top ten largest connected R_0-clusters (by area km^2) present in the R_0-maps of Figures 6.8(d-i) according to the Moore neighbourhood. For both dispersal models, increasing the infectivity parameter β leads to a more susceptible and connected R_0-map, as demonstrated by the larger dominating cluster in Figures 6.8(g) and (k).

There are little to no differences between how Gaussian and inverse power law models spatially scale, according to the raw R_0-maps of Figures 6.8(d-i). However, CCA reveals some subtle differences in the projection of ϕ_1-ga and ϕ_1-pl under the same normalised infectivity parameter β^*. Although, given similar gradients in the graph of R_0 versus ρ for lower infectivity values (demonstrated by the convergence of solid orange and blue lines in Figures 6.8(b-c)), Gaussian and inverse power law models should produce the exact R_0-map. In contrast, for higher infectivity parameters, we might expect the inverse power law model to yield a more susceptible map, as per Figures 6.8(b-c) when the Gaussian model deviate below linearity.

For the infectivity parameter $\beta^* = 600$, the model ϕ_1-ga shows that the largest 'dominating' cluster (in blue) extends over a slightly larger area than for ϕ_1-pl. In Figure 6.8(j), the third-largest cluster (in green, located in Wales) remains disconnected from the largest dominating cluster (shown in blue), whereas in Figure 6.8(f) the corresponding clusters are connected. In a similar vein, Figure 6.8(k) shows a larger dominating cluster than Figure 6.8(g). However, differences are seen most visibly over Scotland and Northern England (Humber-Northumbria), where the second and third largest-ranked clusters ex-

---

[9] Structuring elements have their roots in shape, and image analysis [Shih and Mitchell, 1989], where they define how distinct binary shapes connect to form images [Liang et al., 1989, Nachtegael and Kerre, 2001]

tend over more significant areas. Nevertheless, differences in how dispersal models scale over GB are primarily trivial, and stochasticity in ensemble-averaged R_0-values could likely explain any differences.

6.4.4 Interpreting susceptible R_0-clusters

In a simplified interpretation, the R_0-clusters present two distinct types of dynamics: within-cluster and between-cluster spread. A continuum of tightly coupled susceptible patches within an R_0-cluster outlines a medium through which disease can spread. Therefore, due to the connectivity within an extensive R_0-cluster, the disease may disperse over large areas without directly (or indirectly) jumping between nearest neighbours.

Since disease gradients can extend over $10-1000$km, long-distance spread between clusters will invariably occur, albeit at low frequency. Moreover, alongside long-distance spread directly between clusters, the pathogen may infect trees inside below-threshold patches, thus spreading indirectly between clusters. This mechanism has been conceptually described for pathogens jumping between crop fields [Gilligan, 2002] [Wingen et al., 2013]. Next, we take a closer look at R_0-clustering.

6.4.5 Inverse power law R_0-map clustering

In the inverse power law SEIR model, we required a 5km × 5km domain to accurately gauge R_0 over the mean infectious lifetime of ADB—demonstrated previously in Figure 6.6. Consequently, when mapping the inverse power law over Great Britain, each pixel assumes a 5km × 5km patch. Hence, a larger patch size L outlines a more course-grained host distribution than the Gaussian spread model. Lower map resolutions could be resolved (e.g. 10km × 10km), however, spatial structure is lost, as Figure 6.7(b) eludes to. Therefore, the inverse power law models ϕ_1-pl and ϕ_2-pl were explored over a single landscape-level resolution of 5km × 5km.

For each value of infectivity, an R_0-map will contain a variety of differently sized R_0-clusters. Figure 6.9 presents R_0-clustering behaviour for the inverse power law model. A proper understanding of clustering can be achieved by exploring how the distribution of cluster sizes change in response to infectivity. Figures 6.9(a) and (c) show the frequency distribution of cluster sizes, for models ϕ_1-pl and ϕ_2-pl respectively. The distributions depict cluster sizes for three infectivity parameters $\beta^* \in [250, 500, 1000]$ according to the Moore neighbourhood. In both frequency distributions, a small proportion of large dominating clusters can be seen alongside more numerous smaller R_0-clusters.

Unsurprisingly, a lower infectivity parameter reduces the size of the highest-ranked cluster and lowers the overall coverage of susceptible R_0-clusters—shown in blue. Increasing the infectivity produced a larger dominating cluster and reduced the mean size of lower-ranked R_0-clusters, as demonstrated by contrasting the orange and green frequency distributions. The inset plots of Figure 6.9(a) and (c) depict the same behaviour in a rank-ordered list of the top 25 largest R_0-clusters by size. Both sporulation models display the same fundamental relationship.

In the SEIR model, a value of β has not been determined experimentally or statistical fitted. Therefore, we are compelled to study behaviour over the entire space of β. Figures 6.9(b) and (d) describe how R_0-clusters scale over different values of infectivity and structuring elements for models ϕ_1-pl and ϕ_2-pl respectively. As infectivity is increased to a maximum of $\beta^* = 1000$, the largest R_0-cluster (shown in blue) rises most rapidly over a narrow range of infectivity values in the interval $\beta^* \in [100, 400]$. Patches are more likely to become susceptible as β^* increases; this causes rank-1 R_0-clusters to amalgamate surrounding neighbouring clusters as channels open up. Thus, fluctuations arise in the second and third largest clusters as they form an ever-larger dominating cluster. Below $\beta^* = 100$, no cluster-growth is displayed as epidemic parameters are too low.

Up to this point, we have considered one structuring element, namely the Moore neighbourhood. The type of structuring element we choose to gauge regional-susceptibility in the R_0-map bears no influence on how an epidemic might progress in reality. Therefore, it is desirable for clustering in the model to remain invariant over the structuring element we choose. In general, visible, albeit small, differences can be seen due to artefacts of the

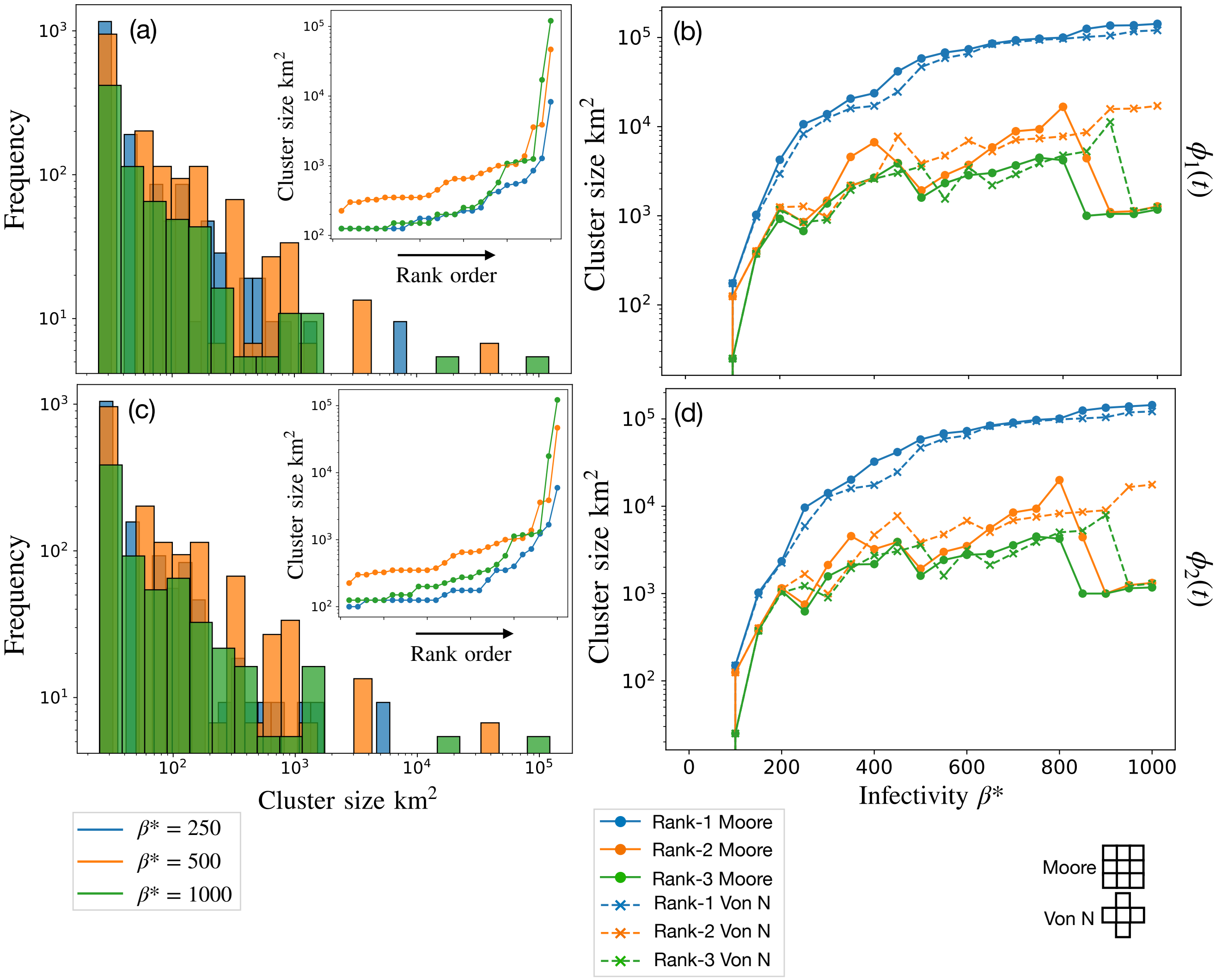


Figure 6.9: Susceptible R_0-clustering for the inverse power law model (a) For model ϕ_1-pl, the cluster-size frequency distribution is shown for three infectivity values utilising the Moore neighbourhood. The inset shows a rank-ordered graph of the top 25 most significant clusters by area km^2. (b) The top three ranked clusters, by size, for model ϕ_1-pl is shown over the infectivity parameter-space for both Moore and Von-Neumann neighbourhoods. (c) The frequency distribution of cluster size is shown for the peaked sporulation model ϕ_2-pl for three infectivity parameters (d) Cluster-size behaviour for model ϕ_2-pl shown over infectivity parameter-space for Moore and Von-Neumann neighbourhoods.

structuring elements.

Moore neighbourhoods can be seen to produce slightly larger rank-1 R_0-clusters, which in turn reduced the size of second and third-ranked clusters. The disparity caused between structuring elements can be understood by noting the two additional corner/diagonal pieces in the Moore neighbourhood. Although structuring elements account for a comparatively small deviation between cluster sizes, the rank-1 clusters compared similarly for some values for β^*. Observing minor artefacts due to structuring elements hint toward limitations in the definition of connectivity, as defined within the landscape-level component of the model.

6.4.6 Gaussian R_0-map clustering: varying map resolution

The essential dynamic of R_0 clustering holds when spatially scaling the Gaussian-based models ϕ_1-ga and ϕ_2-ga. Notwithstanding some negligible cluster size differences, as demonstrated above in Figure 6.8. However, as explored in section 6.3.2, R_0-values in the more localised Gaussian dispersal models can be discerned inside a smaller domain size of 1km × 1km. Given that R_0 can be captured inside a smaller domain, we may consider resolving R_0-maps to a finer landscape-level resolution in comparison to the inverse power law-based dispersal models. Resolving R_0-maps to the highest resolution possible is desirable to apprehend the nature of epidemic severity and spatial heterogeneity. Otherwise, relevant epidemic information is lost when coarse-graining the host distribution.

Figure 6.10 shows how CCA changes clustering in the R_0-map for different landscape-level resolutions. Figures 6.10(a-c) show the frequency distributions of cluster size for $\beta^* = 500$ and three different patch sizes, of lengths 5km, 3km, 1km respectively. As discussed in section 6.4.5, the same general trend remains, i.e. a small number of large clusters alongside more numerous small clusters. Although, decreasing the patch size produces a more fragmented domain, indicated by the greater variety of cluster sizes. The largest ranked cluster toward the far-right of each distribution also shrinks for successively smaller patch sizes. However, decreasing the patch size brings frequency distributions between structuring elements into closer alignment, seen by a closer agreement of blue and orange distributions in Figure 6.10(c).

As before, cluster size for the largest ranked cluster exhibits the most significant growth over the first few values of infectivity, $\beta^* \in [100, 400]$. However, performing CCA at different landscape-level resolutions produces a change in the size of R_0-clusters. At the most extreme point of $\beta^* = 500$, more significant deviations can be seen between the size of rank-1 susceptible R_0-clusters, illustrated by the vertical arrow in Figure 6.10(d).

Figures 6.11(a-c) depict visual differences of CCA at the domain resolutions: 5km × 5km, 3km × 3km, and 1km × 1km. We witness similar landscape-level features for each domain resolution in the largest ranked R_0-cluster (in blue), though disparities exist in the extent and size of clusters. As the R_0-map is resolved to small spatial-scales, connectivity within an R_0-cluster can depend more critically on a small number of above-threshold patches, identified by the red circles in the inset in Figure 6.11(c).

In the coarse-grained 5km × 5km resolution domain of Figure 6.11(a), the critically-connecting patches circled in Figure 6.11(c) are smoothed in the process of averaging the small number of surroundings patches. Similarly, some above-threshold patches can be smoothed to below-threshold patches, thus increasing the cluster size, shown in Figure 6.11(b) circled in red. Ultimately, differences in cluster size are caused by the simplistic notion of connectivity within CCA that fails to scale with the domain resolution. Undesirably, witnessing a difference in CCA-identified clusters outlines an inherent limitation of the framework. Even though clusters appear different, these observations emphasise that connectivity over a region can depend on a small number of patches, which sets the scene for Chapter 7.

6.5 Comparison to observational data

In response to the ADB epidemic, the UK Government commissioned presence-only surveys of ADB, conducted by the Forestry Commission (FC). To date, FC surveys of ADB are the most extensive throughout the UK, composed yearly over a large number of

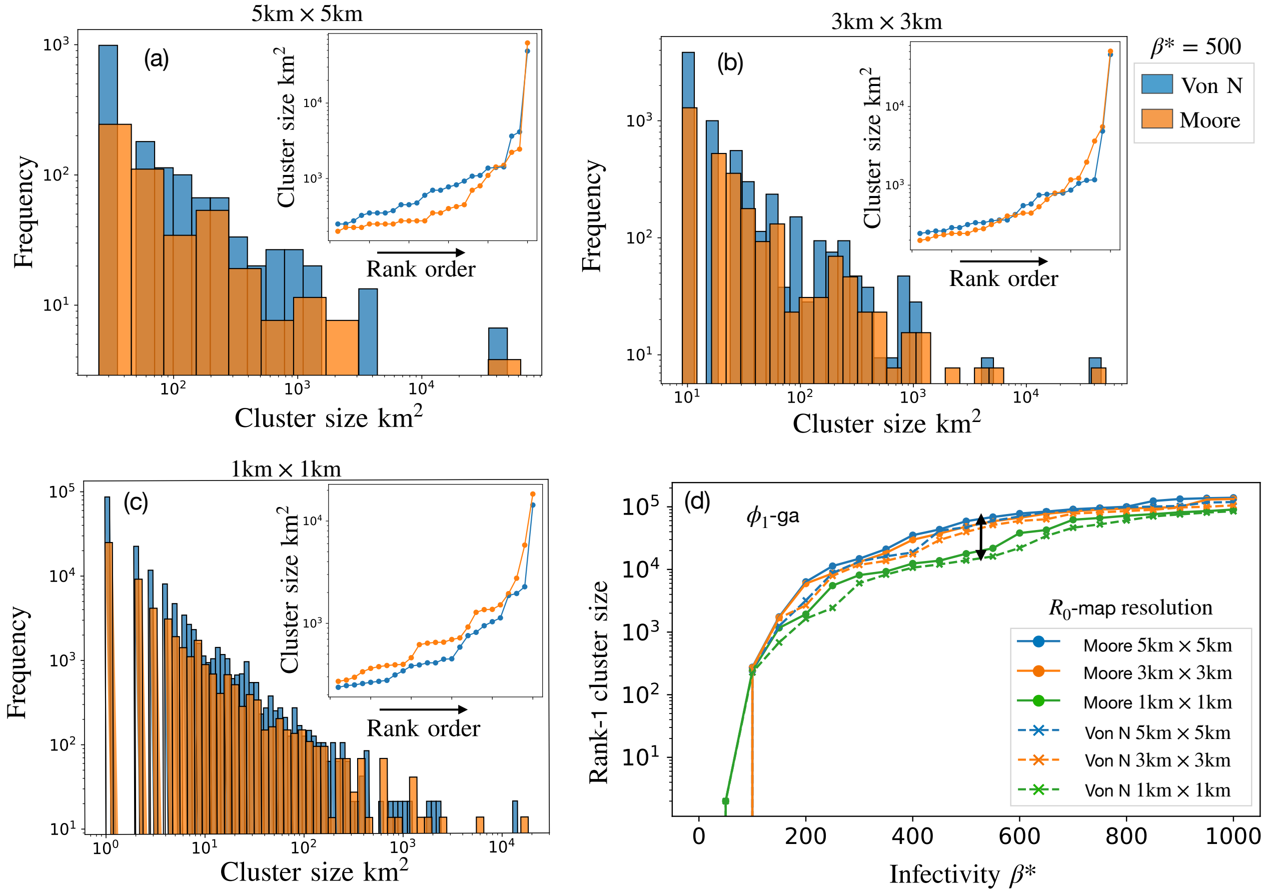


Figure 6.10: Multi-scale CCA performed over different landscape-level resolutions, namely from 1km × 1km to 5km × 5km patch-sizes. (a-c) Cluster-size distributions are shown for three landscape-level domain resolutions and both Moore and Von-Neumann structuring elements. As domain resolution is increased to 1km × 1km, clusters can be resolved to a finer scale and yield a set of clusters over more length-scales (d) Cluster sizes over infectivity β^* and structuring elements, comparing behaviour at two pixel-sizes. At the most extreme point, large disparities in cluster sizes become apparent, indicated by the vertical arrow at $\beta^* = 500$.

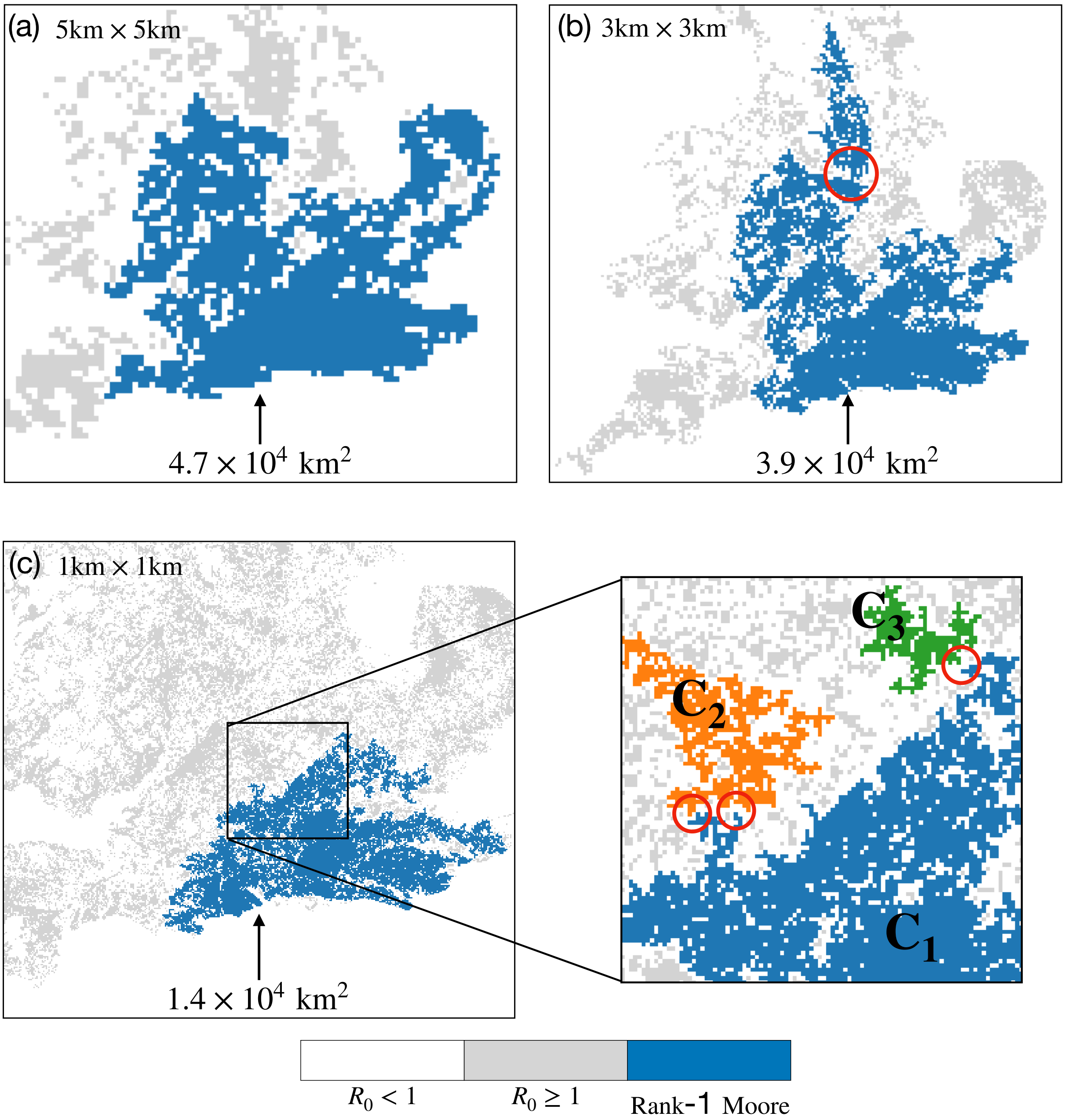


Figure 6.11: Spatial interpretations of Figures 6.10(a-c), according to the Moore neighbourhood, are shown for $\beta^* = 500$ and different patch resolutions. For each domain resolution considered, the largest ranked cluster appears different, pointing to an inherent limitation in the framework (a) The largest ranked R_0-cluster is shown in blue for the most coarse domain at resolution 5km × 5km (b) Resolving R_0-valued patches to a 3km × 3km resolution can lead to extra patches opening up (i.e. becoming susceptible $R_0 > 1$) to form connections between other, distinct R_0-clusters. Here, patches circled in red form a bridge to a previously disconnected cluster (c) Resolving R_0-valued patches to the highest resolution of 1km × 1km causes a more fragmented domain. A small number of below-threshold patches can fragment the R_0-cluster, highlighted as red in the inset.

10km × 10km grids. Figures 6.12(a-d) contrasts the FC data, collected from 2012 (green) to 2020 (purple), against three example R_0-maps from low to high infectivity parameters. The R_0-maps are based on the the model ϕ_1-ga, subsequently all 1km × 1km susceptible patches were grouped into clusters.

The low-valued infectivity map ($\beta^* = 250$) shown in Figure 6.12(b) looks significantly different from the ADB observations in panel (a). Figure 6.12(b) is more fragmented and sparsely-populated with susceptible patches. As such, reported cases of ADB contradict large below-threshold regions of Figure 6.12(b). Visual deviations become particularly apparent in Scotland, where according to Figure 6.12(b), a trivially small distribution of patches could support an epidemic.

Increasing the infectivity to $\beta^* = 650$ in Figure 6.12(c) produces map with more similarities to the FC survey data. The similarity is demonstrated by visual assessments, which reveal a comparable degree of patchiness/fragmentation between maps. Notably, the susceptible regions in Scotland (circled 1) begin to reflect where FC surveys report ADB infections, in stark contrast to Figure 6.12(b). However, a large region in the East of England (circled 2) also appears insusceptible, despite positive ADB reports shown in panel Figure 6.12(a). In addition, Northern Wales (circled 3) appears significantly more patchy and insusceptible compared to the FC data.

The highest infectivity $\beta^* = 1000$ gives rise to a densely populated map with few regions below the threshold. Now clusters of ash in Scotland (circled 1), particularly in the North, appear susceptible whilst the FC survey data in Figure 6.12(a) show no reports. Interestingly, the same insusceptible region (circled 2) in the East deviate from the FC survey dataa, however this the area looks smaller.

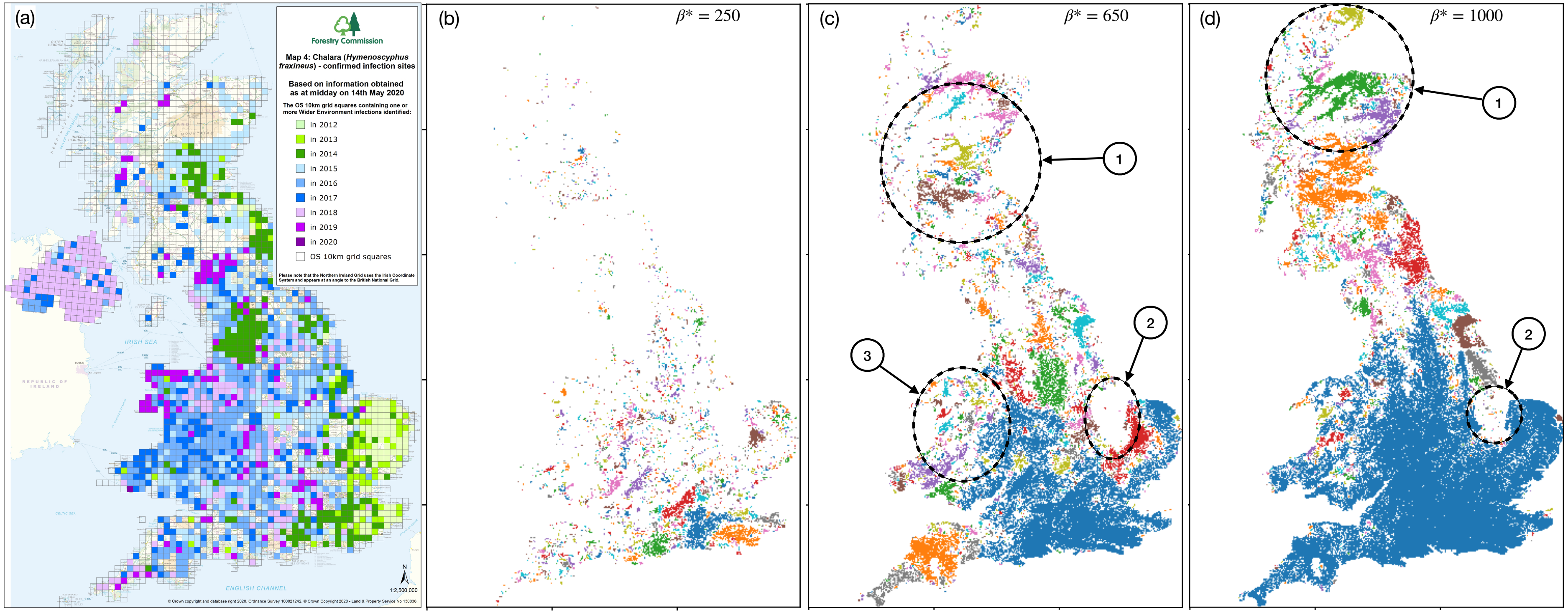


Figure 6.12: Visual comparisons between the Forestry Commissions ADB survey data and three example R_0-maps, from low to high infectivities. For the model ϕ_1-ga, all susceptible 1km × 1km patches were grouped into clusters indicated by colour. (a) Forestry Commission survey data on the spread of ADB collected yearly between 2012-2020, green-purple respectively. (b) A low value of $\beta^* = 250$ produces a highly fragmented R_0 map that looks significantly different from the FC survey data. (c) For $\beta^* = 650$, some landscape features begin to resemble the FC survey data, particularly in Scotland. (d) A high value of $\beta^* = 1000$ appears less fragmented and more densely populated with susceptible regions than indicated by (a).

Together, Figure 6.12 indicates that when infectivity is low, the resulting R_0-map is unlikely to reflect the spread ADB. Additionally, at a certain value, increasing the infectiviy produces an R_0-map with similar features to observation data. On the other hand, if infectiviy is high, the R_0-map is likely too overestimate susceptibility. Ultimately, validating the R_0-map against the FC survey is challenging due to the limited presence-only surveys collected over large 10×10 girds.

6.6 Discussion and future work

The work presented in this Chapter describes how to project a spatially explicit, small-scale compartmentalised model onto a host distribution. The method to 'spatially scale' a small-scale epidemic model is computationally inexpensive, flexible and generalisable. Here, results are focused on helping policymakers implement informed decisions about *where* to control the spread of disease, based on spatial arguments. In the face of low budgets, limited resources and short decision windows, an efficient method to prioritise targeted control could aid both policymakers and (tree health) stakeholders.

Given the sessile nature of trees, the spatial distribution of hosts remains of paramount importance for epidemic modelling. As such, a modelled distribution of ash covering Great Britain was used to parameterise host densities. This choice was motivated by the non-existence of freely available, high-quality ash abundance data that spans Great Britain. Consequently, combining our small-scale SEIR model with data given by [Hill et al., 2017] demonstrates a novel use of predicted abundance data—as remarked when summarising the toy SLM model of Chapter 4.

Due to the complexity of ADB, some important assumptions had to be employed in the SEIR model. In particular, leaf-shed was assumed to fall close to ash hosts; this permitted fungal ascospore dispersal to originate from the same location as infected ash. Still, this assumption is supported by the dispersal parameterisation given by [Grosdidier et al., 2018] that measured dispersal directly between a source of infected trees and spore-traps, thereby omitting the dynamics of intermediary leaf-shed dispersal. One possible improvement to the SEIR model might include host demography, as host-pathogen coexistence is not supported.

Performing CCA proved a beneficial yet problematic exercise. Multi-scale CCA demonstrated how failing to include relevant epidemiological information, in the form of coarse-grained host data, can lead to significant disparities in cluster size—sentiments echoed in the third challenge posed by [Cunniffe et al., 2015a]. Furthermore, resolving Gaussian R_0-maps at different landscape resolutions emphasises an implicit assumption when defining connectivity based solely on nearest-neighbour interactions between patches. That is, Moore and Von Neumann structuring elements do not scale with the dispersal kernel under a change of map resolution:

Suppose secondary infections can be induced up to a maximum distance of $D_{max} = 2$km*. Resolving the map to* 2km × 2km *(or lower) is permissible because patches interact locally i.e. infections between non-nearest neighbour patches are unlikely (neglecting atmospheric of human-mediated LDD). In contrast, suppose the map is now resolved to* 1km × 1km*. In this scenario, patches have the possibility of interacting non-locally because* D_{max} *exceeds intermediary patches; nearest-neighbour structuring elements therefore fail to describe connectivity accurately.* Given all the above, an improved understanding of the connectivity between patches is required to progress the framework.

Moreover, nearest neighbour CCA proved difficult when constructing R_0-maps for systems described by fat-tailed dispersal kernels. That is, R_0 cannot reliably define pathogen invasiveness inside a small domain when dispersal is longer-ranged. Thus, inverse power law dispersal required a larger patch size to measure R_0 from a single focus. In turn, a larger patch-level domain size produced a coarse-grained host distribution that omitted important finer-scale epidemic information. However, future research could examine non-local interactions between R_0-valued patches, which could permit the construction of a higher resolution R_0-map composed of ‘within’ and ‘between’ patch interactions—echoing the short and long ranged interactions inside a metapopulation [Benincà et al., 2020].

Despite the limitations of CCA, clusters grew up to four orders of magnitude over the initial interval of infectivity, $\beta^* \in [100, 400]$. Rapid cluster growth occurred between $\beta^* \in [100, 400]$ for all SEIR model-variants and domain resolutions. Therefore, landscape-level spread within our framework demonstrates a threshold-like behaviour, below which ($\beta^* \lessapprox 100$), no models would be able to invade Great Britain.

Given a set of data containing the *number* of confirmed ADB cases in each grid location, we could correlate R_0-values in the map to the observed number of cases. Nonetheless, correlating R_0 values with case numbers was unattainable because of the non-existent data sources. The only available option was visually comparing the R_0-maps to nationwide FC ADB survey data. However, visual comparisons still proved challenging due to the limited FC survey data that only reported the presence of confirmed ADB cases. Therefore, an improved method to corroborate the model against observational data is required to progress the framework.

Fitting β to ADB data remains a crucial next step for progressing the framework. Numerous studies have collated spatial-temporal ADB data in Europe [Grosdidier et al., 2020, Coker et al., 2019, Stocks et al., 2017, Lõhmus and Runnel, 2014], most of which tracks either tree mortality or the degree of symptom expression over time. Given sufficient spatial-temporal data on the spread of ADB, we could estimate epidemic speed and life-cycle parameters using Bayesian Markov-chain-Monte-Carlo methods. Similar statistical methods have proved noteworthy and effective for modelling the sudden oak death in California [Filipe et al., 2012] and the spread of citrus canker in Florida [Neri et al., 2014]. However, collecting the necessary data often span years and requires considerable resources to monitor disease progression, as demonstrated by the underlying citrus canker data sets [Gottwald et al., 2002]. Unfortunately, comprehensive spatial-temporal data appears absent in the current literature, so fitting the model to spatial-temporal data sources remains questionable. However, more work is required to rule out the approach.

A more straightforward (albeit less comprehensive) method to fit β to ADB data could be achieved using mortality curves. Different authors have fit logistic growth models to data on the spread of ADB—see [Alonso Chavez et al., 2015, Coker et al., 2019, Lõhmus and Runnel, 2014]. Ultimately, mortality curves and logistic growth are non-spatial, in contrast to the work presented in this Chapter. Nevertheless, Figure 6.4 demonstrated an s-shaped logistic growth-like curve that forms the basis for an expanded study focused on model fitting. Fitting the spatial dispersal model to a logistic growth equation could be easily achieved. Section 5.2.2 provides a helpful example of fitting a spatial dispersal model to a non-spatial equation.

The next Chapter builds on the R_0-map and outlines steps toward a novel landscape-level control strategy.

Chapter 7

Towards landscape-level control

This final science chapter shows how spatial heterogeneity leads to natural ‘pinch points’ in the host distribution. Focusing epidemic control through population pinch points may outline an effective strategy to slow disease spread between regions. Subsequently, we develop an algorithm to identify these locations and formulate ’regional containment’ as an approach to impede dispersal.

Some authors have expressed parallel notions to impede the spread of crop and livestock diseases [Papaïx et al., 2014, Gilioli et al., 2013, Gilligan, 2002]. However, no rigorous analysis has taken place, nor has any strategy been outlined for tree populations. Conversely, the method could help identify epidemic risk in potential new afforestation sites. In particular, identifying any epidemically-significant areas will become necessary as the UK government continues to increase afforestation and carbon sequestration, as set out by the carbon plan 2050 [Tol, 2021, Konadu et al., 2015].

A great deal of research has been carried out to understand the optimisation of control in crop and tree-based epidemics—as reviewed in section 2.2. Spatial structure is an essential factor when considering how to manage an outbreak [Picard et al., 2019, Parnell et al., 2009b]. The mainstream paradigm of control typically considers infected tree removals over a relatively small spatial scale, near infected hosts [Cunniffe et al., 2015b], or, more broadly, ahead of the wavefront [Cunniffe et al., 2016]. However, landscape-level epidemic control, based solely on the structure of large-scale spatial distribution of hosts incorporating topography, has yet to be explored in-depth.

Section 7.1 presents a novel method to disrupt epidemic connectivity in the host distribution. Following on, the fragmentation method is analysed in section 7.2. However, the analysis constitutes unfinished work and sets the scene for future research questions. The work conducted in this Chapter rests on the $SEIR$ model of ADB, although the outlook is less specific to ADB and more general-purpose, suitable to any class of wind-dispersed tree pathogens with comparable dispersal parameters.

7.1 Method: cluster fragmentation

In Chapter 6, R_0-clusters were introduced. Clusters of susceptible patches of land in the R_0-map revealed which regions in Great Britain are likely to be the most severely devastated by the pathogen. Whereas patches below the threshold present a natural barrier to the spread of disease; in this simplified interpretation, below-threshold patches can be presumed to not support high levels of infected biomass and infectious spore production. Moving forward, we assume that the pathogen will remain confined within isolated R_0-clusters and not spread through patches below $R_0 = 1$. Undoubtedly, several assumptions underpin this argument, including the omission of LDD, both wind-borne and human-mediated, alongside stochastic outbreaks below the threshold.

An epidemic propagating through the R_0-map describes a '*percolation-like*' problem (with the caveat of clustering that does not feature in classic percolation). All susceptible patches in the cluster risk infection if one patch becomes infected, which motivates a control strategy based on cluster fragmentation to accomplish regional containment. An uncalibrated strategy might involve frantic efforts to reduce the epidemic spread by random targeting of infected areas. However, by concentrating epidemic control on the smaller number of identified patches, connectivity within the cluster can be efficiently disrupted with minimal effort.

Following this epidemic control, cluster fragments define disconnected ‘sub-clusters’ (e.g. $\mathbf{C_1}$ and $\mathbf{C_2}$) that eliminates risk for trees inside one sub-cluster. At the very least, host patches inside the confining cluster risk infection by the pathogen but are more likely to face removal/death. In comparison, hosts inside the other remaining sub-cluster are saved and remain susceptible. Moreover, we achieve efficient epidemic control when the effort is small compared to the reduction of epidemic impact. Moreover, we achieve efficient epidemic control when small efforts and low resource expenditure reduce epidemic impact significantly.

7.1.1 Targeting epidemic control

This section aims to develop a method to answer the question, *“where in the host distribution should we undertake control initiatives to reduce the spread of disease maximally?”*. Consequently, we develop and employ an algorithm to identify a set of optimal epidemic control locations. Specifically, the algorithm finds which sites would lead to cluster fragmentation if R_0 is reduced below the threshold. Notably, when epidemic control is concentrated in these locations, we lower epidemic connectivity between regions, which could slow the spread. Henceforth, the set of patches that fragment a cluster, if controlled below $R_0 = 1$, are defined as “high-priority” patches.

Given a CCA-identified cluster in the R_0-map, $\mathbf{C}$, we can define a binary-valued threshold function $\Phi(\xi)$ by:

$$\Phi(\xi) = \begin{cases} 1 & R_0(i,j) \geq \xi \\ \\ 0 & R_0(i,j) < \xi \end{cases} \tag{7.1}$$

where $R(i,j)$ is the patch-level reproduction ratio at spatial coordinates (i,j) in $\mathbf{C}$. The parameter ξ takes values in interval $\big[1, \max\big(R_0(i,j)\big)\big]$, where $\max\big(R_0(i,j)\big)$ is the highest value of R_0 in **C**—henceforth denoted by R_{max}.

Figure 7.1 shows how the function $\Phi(\xi)$ allows the identification of high priority spatial locations. In Figure 7.1(a) we begin with a target-cluster, **C**, shown in blue; **C** is the largest cluster detected in an arbitrary R_0-map for the model ϕ_1-ga and infectivity $\beta^* = 450$ (as described previously in Chapter 6). Applying the threshold function $\Phi(\xi = 1)$ recovers **C** entirely, as all patches in **C** are over the threshold. In contrast, larger values of ξ result in a lower-density map with sparsely distributed clusters, demonstrated by the small number of labelled clusters in Figure 7.1(b) at $\Phi(\xi = 4)$. Similarly, only one (or at most a handful) of patches populate the domain at the limiting value $\Phi(\xi = R_{max})$.

Connected component analysis (CCA) is performed at each step $\xi \in [1, R_{max}]$ in order to identify and label sub-clusters. At particular steps $\xi \to \xi - \delta\xi$ (i.e. back-stepping), **C** will begin to form when distinct sub-clusters[1] (e.g. $\mathbf{C_1}$ and $\mathbf{C_2}$) suddenly connect when certain critical links become non-zero—analogous to the formation of a spanning cluster opening up in a percolation. Figure 7.1(d) depicts a scenario with a number of disconnected sub-clusters at $\xi = 1.19$ that merge together at $\xi = 1.15$. The 'binary dilation' operator [Liang et al., 1989, Shih and Mitchell, 1989, Nachtegael and Kerre, 2001] was used to detect all the patches which become non-zero and bridge sub-clusters, as elaborated in Appendix C.2. Henceforth, 'connecting patches' refer to patches that connect sub-clusters in a discontinuity step.

When $\mathbf{C_1}$ and $\mathbf{C_2}$ abruptly merge to form the basis of **C**, a significant discontinuous jump in cluster size is detected. All spatial locations (i, j) which bridge the gap between $\mathbf{C_1}$ and $\mathbf{C_2}$ are then identified and removed by taking R_0 below the threshold. Figure 7.1(e) shows two singular patches in blue (and highlighted in red) that if taken below $R_0 = 1$, would fragment **C** into two sub-clusters $\mathbf{C_1}$ and $\mathbf{C_2}$. Successive steps through ξ continue until $\xi = 1$. For each discontinuity event, connecting patches are detected and removed from the system, thus preventing $\mathbf{C_1}$ and $\mathbf{C_2}$ from merging over $\xi \in [1, R_{max}]$. In this manner, the target-cluster **C** is fragmented into two separate sub-clusters. As we can see,

---

[1] It is possible that three or more sub-clusters suddenly connect to form **C** in the same step; these complexities are taken into account by the algorithm.

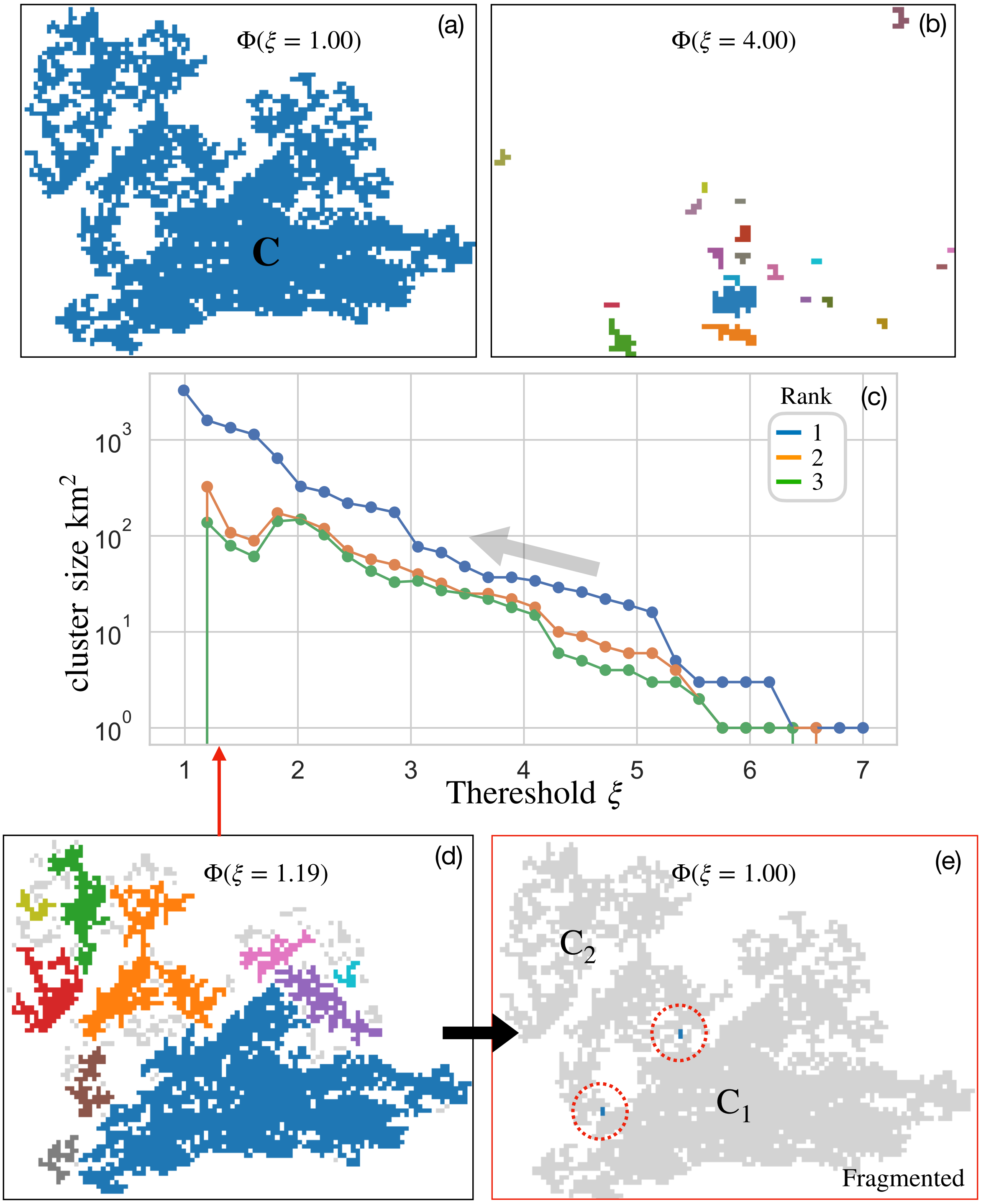


Figure 7.1: A graphical illustration of the algorithm developed to fragment R_0-clusters. (a) The largest cluster, denoted by **C**, is shown inside an arbitrary domain at resolution 3km × 3km and infectiviy $\beta^* = 450$ for the model ϕ_1-ga. Applying the threshold function Φ at $\xi = 1$ recovers the **C** exactly because all patches are above the threshold $R_0 = 1$. (b) Applying the threshold function Φ at $\xi = 4$ yields a low-density map with sparsely distributed clusters, as few patches surpass $R_0 \geq 4$. (c) The top three largest clusters, by area km^2, are shown as a function of ξ. (d) At specific values of ξ, some sub-clusters join to form larger clusters—here, the blue and orange clusters proceed to join at $\xi = 1.15$. (e) Connecting patches are identified when large discontinuities arise when back-stepping $\xi \to \xi - \delta\xi$, shown here by the blue pixels; removing these patches fragment the cluster **C** in $\mathbf{C_1}$ and $\mathbf{C_2}$.

connectedness within the domain can depend on a small number of patches.

7.1.2 Iterative cluster fragmentation

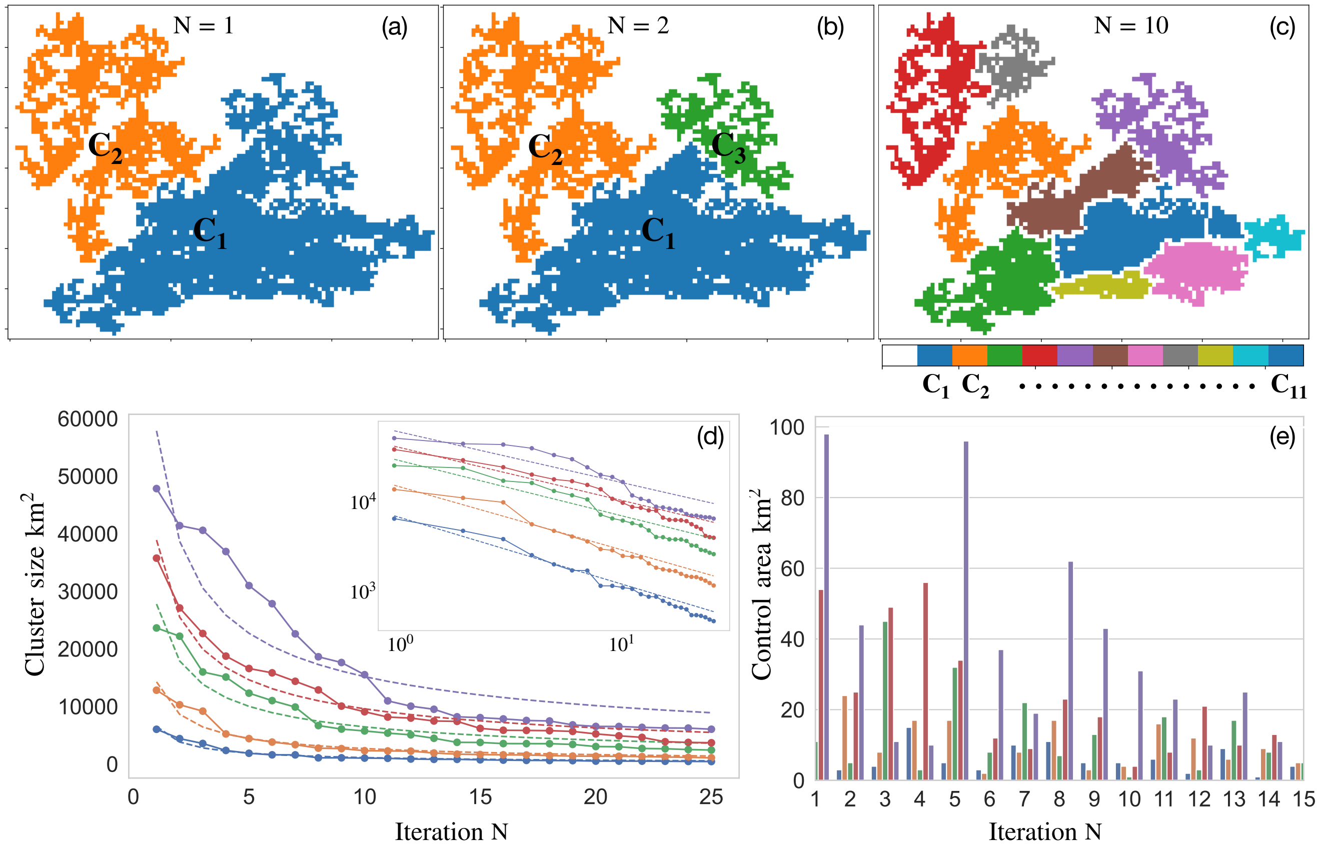
Each R_0-cluster can be iteratively fragmented N times to produce a set of disconnected sub-clusters, where, on average, each fragmentation iteration produces $N+1$ sub-divided clusters. Figure 7.2 demonstrates the iterative process for the example cluster $\mathbf{C}$ from Figure 7.1(a). During first iteration the target cluster $\mathbf{C}$ is fragmented into $\mathbf{C_1}$ and $\mathbf{C_2}$, shown respectively in Figure 7.2(a) as orange and blue. After each iteration, sub-clusters are ranked according to the area they cover, the next iteration proceeds by targeting the largest sub-cluster; this is illustrated in Figure 7.2(b) when $\mathbf{C_1}$ is targeted during the next iteration at $N = 2$, producing a third disconnected sub-cluster ($\mathbf{C_3}$) shown in green. The process is then repeated $N = 10$ times to produce 11 disconnected fragments in Figure 7.2(c).

Figure 7.2(d) shows the sub-cluster size reductions for $N = 25$ iterations and a number of different infectivity parameters. For all infectivity parameters considered, the largest sub-cluster continually decreased for each iteration of fragmentation. Moreover, cluster size reductions occurred more rapidly at first and slowed down as $N \rightarrow 25$. Sub-cluster size reductions were observed to approximately follow an inverse power law of the form $f(x) = ax^{-k}$. Hence, an inverse power function was in Figure 7.2(d), indicated by the coloured dashed lines. Fitted parameter values of a and k reflect the initial cluster size and rate of decrease, respectively. Higher infectivity parameters fit a larger constant a and a smaller exponent k, indicated in the legend. Therefore, the fragmentation process becomes progressively inefficient as β^* increases, demonstrated most clearly by comparing the gradient of the purple and blue lines in the logarithmic inset axes[2].

Lastly, Figure 7.2(e) shows the corresponding number of connecting patches, or ‘control area’, identified over each β^* value and iteration. The number of removed patches tends to decrease with iteration—most likely due to the smaller areas involved—and increase with

---

[2] Additionally, fragmentation was tested alongside the 2^{nd} and 3^{rd} largest R_0-clusters (not shown); for each value of β^*, a and k compared similarly to the 2^{nd} and 3^{rd} largest ranked clusters.



	Fitted params - - -	
Infectivity	*a*	*k*
$\beta^* = 250$	6.4×10^3	0.76
$\beta^* = 400$	1.4×10^4	0.72
$\beta^* = 500$	2.8×10^4	0.63
$\beta^* = 650$	3.9×10^4	0.60
$\beta^* = 850$	5.8×10^4	0.58

Figure 7.2: The fragmentation algorithm is shown as an iterative process. (a) The example cluster (**C**) is fragmented into two sub-clusters ($\mathbf{C_1}$ and $\mathbf{C_2}$) during the first iteration of the algorithm. (b) During the next iteration, the largest sub-cluster $\mathbf{C_1}$ is targeted to produce an additional cluster fragment, denoted here by $\mathbf{C_3}$ in green. (c) The process is iteratively repeated $N = 10$ times to produce 11 sub-dived clusters. (d) The sub-cluster size reductions are plotted for 25 iterations of the algorithm over a range of infectivity parameters. It was observed that size-reduction approximately follow an inverse power law, as indicated by the corresponding fitted dashed lines. Lower infectivity parameters correlate to an efficient fragmentation in contrast to higher β^* values—suggested by the logarithmic inset plot. (e) The area of connecting patches, or 'control area', is plotted against the iteration—truncated to 15. Generally, the number of connecting patches increases with infectivity and decrease with iteration.

infectivity. For example, when $\beta^* = 850$, Figure 7.2(e) shows a control area on the order of 100km^2, arguably, treating a spatial extent of this magnitude would be exceedingly challenging in practice. Thus, when β^* becomes high, it is clear to see a limitation in the framework. In the next section, we outline a potential framework for using cluster fragmentation as a means to achieve 'regional containment'.

7.2 Efficient epidemic control: scenario testing

Regional containment can be tested as an epidemic control strategy by considering hypothetical outbreaks from various epicentres. Figure 7.3 demonstrates containment for a single epicenter marked by the black cross—located inside the same target cluster shown in Figures 7.1 and 7.2. We can achieve epidemic containment in several ways, as alluded to by Figure 7.3. The connecting patches (identified over $N = 25$ fragmentation iterations) can be combined in many ways to define different boundaries around the epicentre. For example, Figure 7.3(a) defines a boundary by considering connecting patches determined in the $1^{st}, 3^{rd}, 7^{th}$ and 8^{th} iterations, shown in red. The boundary then defines a confining sub-cluster around the epicentre; in theory, light-grey patches remain disconnected and susceptible whilst the dark grey patches are removed/at-risk.

The ratio $N_S/(N_R \times N_C)$ provides a simple notion of how epidemic control efficiency; where N_S, N_R and N_C are the number of patches that remain susceptible, become removed and are targeted for control, respectively. We have efficient containment when the number of 'saved' patches is high and the number of patches removed and controlled is low. Henceforth, the term 'control payoff' denotes the ratio $N_S/(N_R \times N_C)$. That is to say, we have a large *control payoff* when $N_S/(N_R \times N_C)$ is large. Additionally, we keep the control method generic, though it usually involves either culling or the biological treatment of infected hosts.

Resources to control the pathogen are finite and depend on governmental budgets. However, with more work, we may also be able to compute a limit on N_C and perform a more sophisticated analysis. Similarly, N_R paints the simple picture of patches becoming removed/at-risk; in reality, the number of patches at risk of removal would be more complicated and subject to LDD and stochastic below-threshold outbreaks. However, we continue with a theoretic investigation with no expressed limit on N_C and a simple notion of N_S.

Lastly, it is worthwhile to describe some edge cases and complexities that arise when determining containment scenarios. Suppose that containment is detected about an epicentre by combining the connecting patches identified in two arbitrary iterations N_i and N_j. In this scenario, some patches from N_i and N_j may be located away from the containing sub-clusters boundary, i.e. in distant (non-bounding) locations. Thus, irrelevant (non-connecting) patches were identified by employing the binary dilation operator; in this way, only patches neighbouring the confining sub-clusters perimeter contribute to N_C. Moreover, combinations of connecting-patch iterations that failed to define a confining sub-cluster about the epicentre must be ruled out from the analysis.

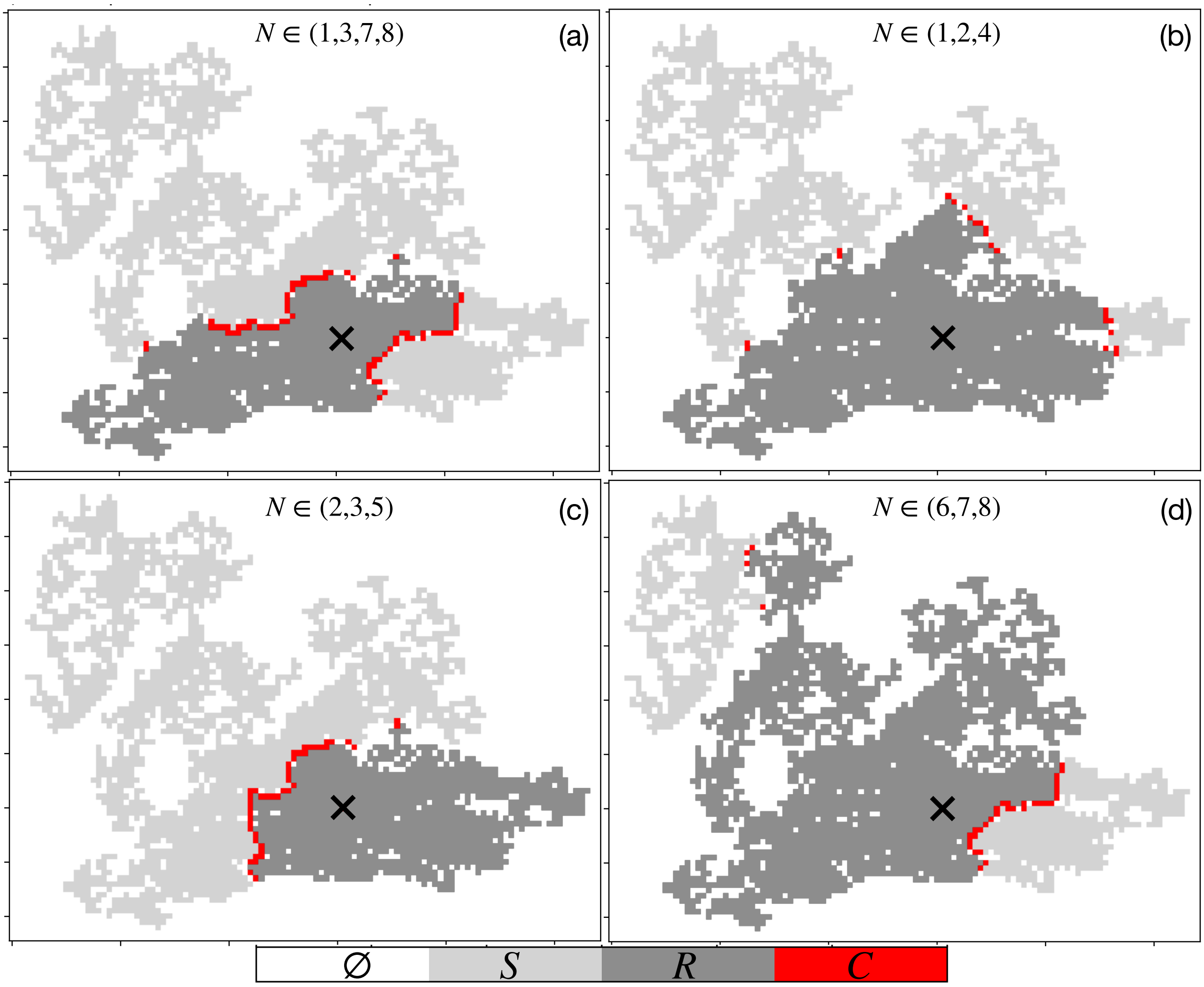


Figure 7.3: A variety of different control choices are possible for each epicentre, based on the landscape-level host aggregation. Here, the algorithm recursively fragments the target cluster **C** through $N = 25$ iterations, then different combinations of connecting patches can be used to contain the outbreak in a variety of ways. Panels (a-d) represent a small sample of control scenarios for an arbitrary epicentre marked by the central black cross. Red patches indicate where landscape-level control C should be targeted to contain the epidemic. Light grey patches remain susceptible (S) whilst dark grey parches are assumed removed/at-risk, denoted by R. In practice, every possible control scenario is assessed against every possible epicentre.

7.2.1 Results: control payoff

Containment scenario payoffs were assessed against all possible epicentres in Great Britain. For each scenario, the sub-cluster centre of mass (COM) defines an epicentre, i.e. we compute the COM for each sub-cluster produced from $N = 20$ iterations of target-cluster fragmentation. As before, ‘target cluster’ refers to the largest dominating cluster detected in the R_0 map for each β-valued domain. Containment scenarios can then be determined for each epicenter—as described above. The payoff ratios were then ranked according to the payoff ratio $N_S/(N_R \times N_C)$, as shown in Figure 7.4—all panels consider the model ϕ_1-ga resolved to 5km × 5km sized patches.

Figure 7.4(a) presents the top 25 epidemic containment scenarios over a range of infectiviy parameters in the interval $\beta^* \in [0, 1000]$. The reader is referred back to Chapter 6 and Figures 6.9-6.10 for background information on cluster size with infectiviy β^*. Beyond $\beta^* = 200$, each value of infectivity involved a large number of containment scenarios, i.e. typically between $10^3 - 10^4$ containment scenarios per β^* parameter. The payoff ratio starts small with low infectivity values and begins to peak before dropping off. The control payoff is low for low values of β^* since the number of trees saved is generally lower on account of small susceptible clusters. Figure 7.4(a) therefore indicates that this strategy of control is not desirable for pathogens with low infectivities.

For exceptionally high values of β^*, a small number of control scenarios record high-valued payoff ratios, plotted in the shaded region Figure 7.4(a); this arises on account of efficient edge-location control that saves vast swathes of the host population with a low number of host removals. Arguably, this represents a limitation of the payoff ratio, as we’ve defined in this Chapter, because in reality edge-location control is highly idealistic and unlikely to be realised.

Interestingly, Figure 7.4(a) indicates that regional epidemic containment is most efficient over a specific parameter regime, since the highest payoff scenarios occur in the interval $\beta^* \in [400, 600]$. Nonetheless, these are preliminary indications. As section 6.4.6 demonstrated, nearest-neighbour structuring elements do not scale with changes landscape-level resolution. Given that Figure 7.4(a) is based on ϕ_1-ga at the resolution 5km$\times$5km, the results presented by Figure 7.4(a) are likely to change if containment scenarios are computed with smaller patch sizes. Although Figure 7.4(a) demonstrates an important observation, more work is needed to be confident that regional containment is most effective over the interval $\beta^* \in [400, 600]$.

Figure 7.4(b) illustrates the complete set of scenario tests for $\beta^* = 500$ (i.e. the β^* parameter that registered the highest control payoff). The number of saved patches that remain in S (i.e, N_S) is plotted against $N_R \times N_C$. The upper and right-hand marginal plots show the corresponding probability density functions for N_S and $N_R \times N_C$. Each PDF shows a skewed distribution, with most scenarios involving a smaller value of N_S and $N_R \times N_C$. Of the 3850 containment scenarios, a small number of high performing tests populate the bottom right-hand quadrant, when N_S is large and $N_R \times N_C$ is low. This region in the plot represents scenarios where containment is most efficient i.e. the right and left-hand distribution tails when N_S is large and $N_R \times N_C$ is small.

Figures 7.4(c-e) show three scenarios of interest, orange and blue clusters depict removed/at-risk and susceptible/saved patches, respectively. Red crosses indicate where the control should be focused to achieve containment. Figure 7.4(c), depicts the likely scenario with a central focus of infection. Most of GB remains at risk, yet some Northern and Southern regions surrounding the target cluster remain protected from the pathogen. If the disease has not yet reached the connecting patches (identified by the red crosses), Figure 7.4(c) indicates that regional containment might be attempted alongside targeting a disease wave-front [Cunniffe et al., 2016], or more local-scale pathogen eradication [Cunniffe et al., 2015b]. Although, ultimately, more work is needed to assess the efficacy and possibility of halting/slowing the spread.

The targeted control patches in Figure 7.4(d) resemble those identified previously in Figure 7.1(e), yet this time for a higher β^*-valued domain at a lower landscape-level resolution. Desirably, observing connecting patches in approximately the same location for different infectivity values suggests that some patches may be important for control

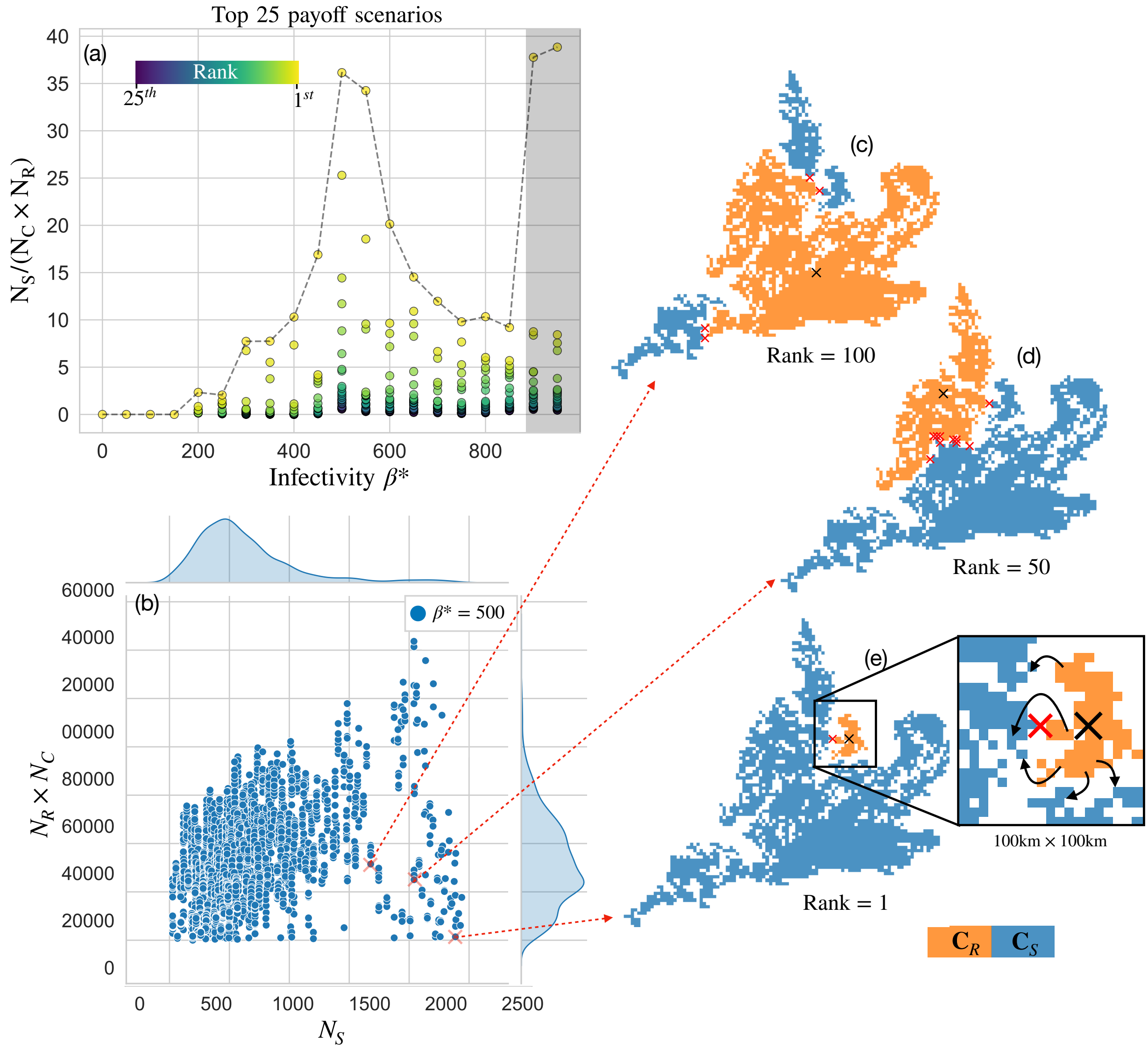


Figure 7.4: The control 'payoff' is accessed by comparing the number of patches that remain susceptible (N_S) against the number of patches removed (N_R) and controlled (N_C). (a) The payoff ratio $N_S/(N_R \times N_C)$ is plotted against the infectiviy parameter β^* for the top 25 highest payoff scenarios. (b) The complete list of (3850) different scenario tests are plotted for the highest payoff infectivity parameter $\beta^* = 500$; the lower right-hand quadrant defines the most efficient control scenarios. (c-e) Spatial plots that show three hypothetical scenarios from panel (b), with payoffs ranked 1, 50, 100. Blue and orange clusters outline patches that remain susceptible and become removed, respectively.

at various epidemic scales. Moreover, Figure 7.4(d) hints toward there being similarities at and different resolutions—the concern mentioned above.

Figure 7.4(e) reveals the highest payoff result for $\beta^* = 500$. In this scenario, the epicentre lies close to the Eastern coastline (Skegness), and the control area is located slightly inland (approximately between Lincoln and Sheffield). The zoomed inset highlights a single 5km × 5km patch, and control saves the vast majority of GB from infection. This scenario is no doubt idealistic and unlikely to be realised in a real-life outbreak. In a real-life outbreak, epicentres are unlikely to be so conveniently located about the coastline. Furthermore, detecting and controlling the pathogen remains difficult before it spreads to distant (and more centralised) locations. Nevertheless, Figure 7.4(e) reinforces the intuitive notion that epicentres around edge positions are contained efficiently. The inset highlights where pathogen dispersal might jump between clusters, indicated by the curved arrows. Although wind-dispersed secondary infections are likely rare over these spatial scales (5-20km), they are nevertheless thought possible for fungal spores [Grosdidier et al., 2018]; the analysis presented here neglect these complexities, which requires further investigation.

7.3 Discussion and future research

This Chapter outlined a conceptual strategy of landscape-level control, targeting a pathogens local wind-dispersal mechanism. The strategy aims to preferentially control the pathogen through identified patches in the host population. In an idealistic scenario, cluster fragmentation can be achieved along with 'regional containment' if the identified patches are taken below the epidemic threshold. Nonetheless, fungal spores are thought to be able to disperse over large distances, e.g. from mainland Europe to Great Britain [Wylder et al., 2018, Freer-Smith and Webber, 2017]. Given that the effects of LDD (beyond 5km) were neglected, the utility, efficiency and practicality of real-life implementation remain unproven. However, the work undertaken in this Chapter sets the scene for future investigations.

Various large-scale models have examined spread over contiguous cells, or patches, e.g. [Gaydos et al., 2019, Cunniffe et al., 2016, Meentemeyer et al., 2011]. However, to our knowledge no large-scale studies involve the spatial arguments based on population heterogeneity presented in this Chapter. Although the idea of density reductions unpinned control in this work (consistent with more traditional methods of eradication), the possibility of other biological control methods remains an attractive option, e.g. fungicide treatments [Hauptman et al., 2015].

Targeted epidemic control goes hand-in-hand with an on-the-ground understanding of which regions become infected. Subsequently, the method presented to identify high-risk areas might find applications in enhanced monitoring and surveillance [Parnell et al., 2017], discussed previously in section 1.2. Consequently, future work could access the method of critical/connecting patch detection for enhanced landscape-level monitoring.

The algorithm constructed in section 7.1 presented a means to identify 'connecting patches', which if taken below $R_0 = 1$ would fragment the cluster. Related concepts of network fragmentation exist in telecommunications [Albert et al., 2000], ecological space modeling [Luo et al., 2021] and human epidemiology [Chami et al., 2017]. However, no

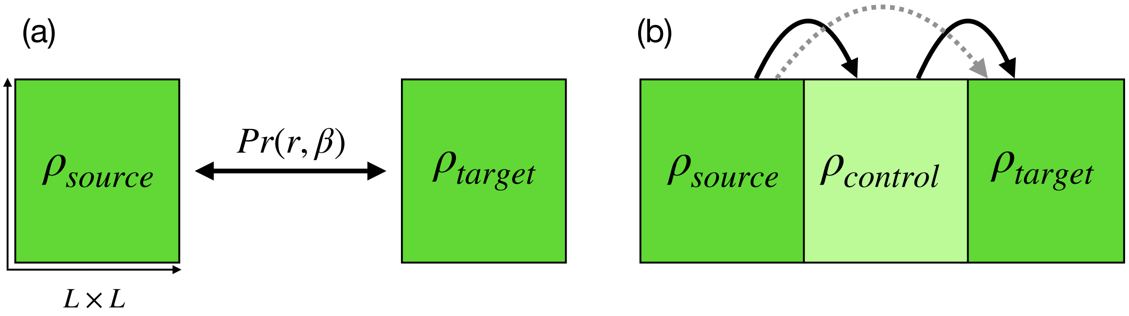


Figure 7.5: A diagram illustrating potential research avenues to improve the framework and better evaluate the control strategy. (a) Examining the transmission probability between a source and target patch as a function of distance r and infectivity β. (b) Understanding the degree to which control of an intermediary patch ($\rho_{control}$) can disrupt dispersal between a source and target patch.

sources could be found relating to spatially-explicit epidemic spread over a landscape of trees. The algorithm described in 7.1 treated patches below the threshold as negligible, without this assumption CCA would cease to work. Given the possibility of stochastic below-threshold spread, future work should move towards a generalised (risk-based) scheme that factors in the presence of patches below $R_0 = 1$; in such a scheme, we may ultimately question the choice of CCA due to its binary nature and inability to describe non-local coupling. Subsequently, routes to progress and improve the framework and control strategy are outlined below.

7.3.1 Future research questions

Throughout Chapters 6 and 7, nearest-neighbour connectivity underpinned R_0-map clustering. Ultimately, this followed from defining R_0-maps based solely on 'within-patch' interactions. Moving forward, we require an enhanced understanding of 'between-patch' interactions. Figure 7.5 illustrates two future research questions that could help develop the framework and assess the utility of landscape-level control.

Patch-to-patch transmission

A modelling setup could define two patches, a source patch (ρ_{source}) that hosts the infection at $t = 0$, and an infection-free target patch (ρ_{target}). Ideally, patches should mirror the highest possible resolution—given by 1km^2 for the maps of predicted ash abundance given by [Hill et al., 2017]. One could then assess the transmission probability between ρ_{source} and ρ_{target} as a function of distance r and infectivity β, reflected in Figure 7.5(a). A similar on-the-ground experiment was conducted by [Grosdidier et al., 2020], who accessed the prevalence of collar canker and rachis symptoms in neighbouring ash. The authors found that the influence of ADB decayed exponentially up to $200 - 300$m away from a high-density source.

Although inverse power laws are more likely to describe the spread of ADB, contrasting the transmission probability between Gaussian and power-law dispersal models for different levels of β could prove insightful. In particular, this examination could provide two length-scales, above which, patch-to-patch transmission is unlikely. Such insights could help understand the disease spread between infected ash stands and help bolster the knowledge of forest managers, especially when constructing new stands.

If the source patch reflects on-the-ground observations of infected areas, we can speculate about constructing a ‘transmission probability’ map in GB. In theory, an ensemble average calculating the transmission probability as a function of distance r, infectivity β and target/source densities could be projected onto the map of predicted ash abundance. Then, we could aim to construct a probability map depicting which neighbouring regions are likely to become infected. An accurate and reliable probability map could provide immense value to rapid-response control initiatives by helping direct the allocation of resources in the early phases of an invasion. Although, this notion undoubtedly requires effective monitoring and surveillance to detect disease inside infected patches before the disease has had time to propagate over large areas.

Between-patch control

Lastly, targeted control rests on artificially reducing R_0 below unity—outlined in section 7.2. Subsequently, non-existent epidemic spread was assumed for patches $R_0 = 1$. More research is required to understand this assumption and further assess the efficiency of control. In addition, resource constraints were neglected. In reality, control initiatives are limited by governmental budgets and the funding made available to stakeholders.

Therefore, any robust control strategy should make efforts to incorporate the influence of government budgets in large-scale control [Cunniffe et al., 2016].

Figure 7.5(b) illustrates a potential setup to progress the understanding of landscape-level control by considering a system of three coupled patches. The coupled system shown in Figure 7.5(b) is similar to Figure 7.5(a), although this time an intermediary 'control' patch separates the target and source patches. In this system, dispersal can either jump between or directly over the control patch, shown by the solid black and dashed grey arrows, respectively. In $\rho_{control}$, control can be achieved by reducing tree density through sanitation felling or by reducing β (mirroring biological control, e.g. fungicide treatments).

Studying the effect of different control measures in $\rho_{control}$ could provide crucial insight into the control strategy proposed in this Chapter. In particular, by studying the number of infected trees that arise in ρ_{target} conditional on different control measures. Intuitively, we could expect efficient control when propagation occurs through a thin-tailed Gaussian kernel. However, for fat-tailed dispersal kernels, the efficacy of between-patch control remains far less obvious and subject to more scrutiny. In summary, three hypothetical outcomes can be expected from studying the coupled system:

1. Between-patch control is insufficient because large numbers of secondary infections arise in ρ_{target} due to infected trees in ρ_{source}, regardless of epidemic control.
2. The development of an outbreak in ρ_{target} is slowed down by control measures in $\rho_{control}$, conditioned on the degree of epidemic and control.
3. Complete epidemic containment between patches can be achieved, thereby preventing the pathogen from propagating to ρ_{target}.

For inverse power law dispersal, complete epidemic containment, as described in (3), is unlikely due to its fat-tail. However, future research could assess the degree to which control slows down the spread between patches, conditioned on epidemic and control parameters. Although, the answer to these questions remains open to investigation.

Chapter 8

Discussion

In this thesis, a simple model of forest epidemics was incrementally extended into a more elaborate framework. Each incremental improvement led to a novel, general-purpose framework to visualise epidemic severity across GB. The framework is computationally efficient and adaptable to any wind-dispersed tree pathogen, provided a sufficient host density distribution is available. Conclusively, more research is required to progress and test the model against observational disease incidence data. However, several exciting research avenues emerge from the work conducted in this thesis—discussed more below.

After setting the scene with a simple lattice model (SLM) of tree disease in Chapter 3, Chapter 4 linked the SLM with a map of predicted oak abundance given by [Hill et al., 2017] to produce a large-scale 'toy' model. A key result emerged from Chapter 4. Namely, that NN interactions were fundamentally insufficient to describe the spread of disease across lower, more realistic landscape tree densities. Fortunately, Chapter 5 resolved the issue by constructing a non-local dispersal model. Nonetheless, Chapter 5 also demonstrated that a small dispersal length scale can still gives rise to an unnatural, wave/percolation-like epidemic. At the very least, these results bolster the growing body of research highlighting the importance of dispersal and help guide modellers to construct more representative models in botanical epidemiology.

Additionally, Chapter 5 examined two methods of calculating R_0 for a non-local dispersal model (NLM) over a range of epidemic parameters. Results from Chapter 5 indicate that when the scale of dispersal is comparable to the domain, the model approximates a mass-action well-mixed system described by the standard SIR model. However, comparisons to the SIR model were simplified and limited to one parameter (i.e. the ratio β/γ). Subsequently, the analysis constitutes a preliminary result, and a more sophisticated comparison method is required to glean further insight.

Most intriguingly, the spatially-explicit derivation of R_0 led to an 'entire function' in Equation 5.14, a well-known function in complex analysis [Abramowitz and Stegun, 1948]. Entire functions have been examined in several theoretical settings (e.g. [Littmann, 2005, Hryniv and Mykytyuk, 2009, Sixsmith, 2011]), including a spatially-structured population model with dispersal [Zhang, 2017]. However, despite a detailed search, no theoretical studies expand or develop upon Equation 5.14 in the context of epidemics. As such, Equation 5.14 marks a different approach for determining R_0 that considers the spatial structure, host density, dispersal mechanism and infectivity. Further theoretical studies could explore a more rigorous analysis of Equation 5.14 and perhaps examine an alternate derivation incorporating inverse power law dispersal with exponentially-distributed removal lifetime dynamics. However, the mathematical derivation might become too challenging in the face of more complicated epidemic models

Then in Chapter 5, the analytic expression of R_0 was compared against the 'actual' contact-traced reproduction ratio. Both methods used to determine the reproduction ratio demonstrated a clear epidemic threshold at $R_0 = 1$, marking an important finding. Nevertheless, comparing both methods revealed that the analytic expression overestimates R_0 for progressively higher values ($R_0 \sim 10$), highlighting a significant limitation in the approach. Had the analytic expression been applied to a dense forest or highly aggregated distribution (where $\rho > 10\%$ and R_0 is likely higher), the overestimation would only increase. The utility of Equation 5.14 therefore diminishes when used to describe highly infectious regimes but serves as an accurate approximation around the threshold $R_0 \approx 1$. Notwithstanding, tree-based diseases are unlikely to reach such exceptionally high epidemic values (where $R_0 > 10$), as supported by estimates of $1 \lessapprox R_0 \lessapprox 6$ for the Dutch elm disease epidemic in GB [Swinton and Gilligan, 1996a] and $1 \lessapprox R_0 \lessapprox 4$ for oak processionary moth epidemics in London [Wadkin et al., 2022].

In contrast, the contact-tracing method computes the mean number of infections per infected tree over different infected generations. Computing the mean-generational reproductive ratio proved convenient and led to a sharper epidemic threshold for later generations. These observations relate nicely to [Tildesley and Keeling, 2009], who studied a similar method to calculate R_0 for foot and mouth disease. Despite the convenience of contact-tracing secondary infections in the model, it would be challenging to measure in natural systems experimentally. As such, the method remains applicable to abstract modelling work alone and not in-the-field experiments.

The central value of Chapter 6 results from outlining a novel framework to link a wind-dispersed epidemic model with species abundance data. The framework focused on the fungus ADB, which is well-established and already spread throughout Europe and the UK. Thus, unfortunately, attempting eradication at this stage of pathogen development is untenable [Hill et al., 2019]. Still, Chapter 6 marks a step towards a general framework that can help policymakers reach informed decisions about *where* to focus epidemic control. In particular, the framework could provide value as an approach to threat assessment and rapid response modelling during the early phase of an epidemic.

In addition, Chapter 6 developed a simplified spatially-explicit SEIR model that described the seasonal spread of ash dieback (ADB) over local spatial scales. Subsequently, we coupled the ADB model to a map of predicted ash abundance given by [Hill et al., 2017] to produce an R_0-map that covered GB. Under the threat of ADB, the UK government commissioned some early 'rapid response' modelling work as part of the Chalara management plan[1] [Defra, 2013]. Although, the modelling work undertaken by Chalara Cambridge plant sciences remains unpublished. Regardless, no compartmentalised model of ADB could be found in the literature. Therefore, to my knowledge, the ADB model

---

[1] As set by the government, readers can find more information by reading the Chalara Management Plan: https://www.gov.uk/government/publications/chalara-management-plan

presented in this thesis is the first mechanistic (compartmentalised) attempt to model ADB's epidemic spread.

Treating infectivity as a free parameter in Chapter 6 led to several notable observations. When infectivity is low, the R_0-maps become sparsely populated with susceptible patches above the epidemic threshold. Consequently, the R_0-maps indicate that ADB would be unlikely to invade GB below a hypothetical minimum infectivity value. Similarly, clusters in the R_0-map proliferated over four orders of magnitude between a narrow range of infectivity parameters. Altogether, the model alludes to behaviour akin to a global epidemic phase transition across GB, though numerous assumptions limit the framework. Keeping infectivity as a free parameter enabled R_0-map analysis over a spectrum of parameters. Nevertheless, arbitrarily defining infectivity underpinned a significant limitation in work, and future research should aim to address this by fitting a value of β to data.

Several published datasets record the spread of ADB throughout different European countries, e.g. [Grosdidier et al., 2020, Stocks et al., 2017]. Such datasets could form the basis of fitting β using Bayesian Markov-chain-Monte-Marlo methods to estimate tree pathogens' epidemic speed and lifecycle parameters. Similar approaches have been adopted to infer the spread of SOD in California [Filipe et al., 2012], and citrus canker in Florida [Neri et al., 2014]. However, no such work has been undertaken for ADB. Further research is ultimately required to confirm the suitability of statistically fitting data to the spread of ADB. At the very least, comparing mortality in the simulated ADB model to mortality curves reported in the literature [Coker et al., 2019, Alonso Chavez et al., 2015, Lõhmus and Runnel, 2014] could form a more straightforward approach to fitting infectivity.

ADB mortality rates depend heavily on the landscape, composed of either rural, woodland or urban settings. The type of environment is one particular consideration relevant for future epidemic parameter inference studies. For simplicity, Chapter 6 neglected environment types and employed a one-to-one mapping between infectivity and each R_0-map. In a more sophisticated model, each gird in the R_0-map would depend on a

β parameter dependent on the type of environment. Thus, statistical inference based on the environment type goes hand-in-hand with improving accuracy in the R_0-map. The national forest inventory could conveniently aid these future improvements, as it holds relevant information on which regions are woodland, forest or urban—discussed previously in section 4.2.1.

Chapter 7 outlined the first steps toward epidemic control predicated on the host spatial structure. That is, identifying and targeting positions in the host distribution that may disrupt epidemic dispersal between regions. Initial results reveal that epidemic connectivity can depend on a small number of 'connecting' positions in the host distribution. Nevertheless, more research is required to assess the utility and efficiency of the control strategy. Chapter 7 outlined a potential research direction to examine and test the control strategy by considering a set of coupled patches in section 7.3.1. From the coupled system, the transmission probability and effect of control between host patches can be assessed. Until this work is undertaken, the strategy remains speculative. Following the research direction posed in section 7.3.1, future work should develop the definition of connectivity inside the R_0-map. Each pixel within an R_0-map reflects only isolated within-patch interactions and not between-patch LDD. As such, there is no non-local connectivity between pixels in the map, a limitation clearly revealed when analysing clusters at different landscape resolutions in section 6.4.6.

Unsurprisingly, insufficient host data underpins a significant limitation in this thesis. Although the predicted ash abundance map captured the overarching large-scale distribution of ash in GB, Hill et al. reported a RMSE of 5ha. Therefore, in reality, regions below the threshold might be susceptible—and vice-versa for above threshold regions. Unfortunately, such errors mean that the host distribution is not accurate enough to inform the hypothetical (fine-scale) management scenarios presented in Chapter 7. Until species abundance data captures the host distribution more reliably, country-wide applications remain extraordinarily ambitious. In response to this limitation, future research could concentrate on smaller-scale areas where abundance data is known. For example, by examining well-surveyed areas inside the UKCEH Countryside Survey data.

The regional containment strategy of Chapter 7 could help enhance the government's contingency and preparedness planning for new invasions. Presently, DEFRA lists contingency measures for numerous botanical diseases on the plant health portal[2]. Generally, contingency measures describe 'demarcated zones', consisting of an 'infected area' and a 'buffer zone'[3]. Infected areas outline ($r \lessapprox 100$m) regions directly surrounding verified infections where the destruction of susceptible plant material is recommended. Then, a buffer zone ($1\text{km} \lessapprox r \lessapprox 2.5\text{km}$) is established around the infected zone. Buffer zones are subject to continuous surveillance and monitoring to detect new infections. However, no initiative exists to coordinate epidemic control between different infected zones. Instead, contingency plans aim to manage infected/buffer zones independently, which contrasts with the regional control strategy proposed in Chapter 7.

In a large-scale outbreak with several confirmed infected areas, the method illustrated in Chapter 7 could help prioritise which infected sites undergo epidemic control, with the added benefit of effectively slowing the spread between regions. For example, regional containment would be advantageous when an emergent infectious epidemic threatens to invade an uninfected yet high-risk neighbouring area. Another use-case pertains to a scenario where outbreaks established close to the coastline threaten high-risk regions situated more inland, as illustrated by Figure 7.4(e). In any case, slowing the spread between regions gives tree populations added time to recover and offers policymakers and stakeholders vital time to respond.

In conclusion, the research narrative developed in this thesis aims to help inform policymakers about where to focus epidemic control. The approach constitutes an epidemic mapping framework for tree disease with parallels to the emerging field of Infectious Disease Cartography in human epidemiology. The framework is computationally efficient, flexible, and adaptable to other pathosystems. Several theoretical insights were ascertained from deriving a spatially-explicit expression for R_0 and comparing it against a

---

[2] DEFRA's current set of contingency plans can be found at the following address: https://planthealthportal.defra.gov.uk/pests-and-diseases/contingency-planning/.

[3] Relevant contingency measures include DEFRA's plans to manage *Xylella fastidiosa*, oak wilt, oak processionary moth and emerald ash borer

stochastic non-local dispersal model. Lastly, a novel epidemic control strategy was outlined, though more work is needed to progress the framework and rigorously validate results.

Appendix A

Simple lattice Model

A.1 Propagation algorithm

In Chapter 3, a percolation-based epidemic model was outlined following previous work [Orozco-Fuentes et al., 2019]. Starting from simplicity, the model assumes local transmission between nearest neighbours. Local structure within the network is described by the von Neumann neighbourhood [Toffoli and Margolus, 1987]. Consider a small 3×3 matrix, denoted by $\mathbb{S}$, representing a small patch of forest. In the matrix, values of 1 and 0 represent susceptible (S) and insusceptible ($\emptyset$) states, respectively. Similarly, infection and removal matrices, $\mathbb{I}$ and $\mathbb{R}$, track infected and removed tree states. A hypothetical system at time $t = 0$ is outlined below:

$$\mathbb{S} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad \mathbb{I} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \mathbb{R} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{A.1}$$

A.1.1 Transition probabilities

An infected tree $\mathbb{I}_{1,1} = 2$ occupies $\mathbb{I}$. Whenever $\mathbb{I}_{i,j} \geq 2$, $\mathbb{S}_{i,j} = 0$, i.e. susceptible and infected states cannot occupy the same lattice location. The algorithm locates nearest neighbour (NN) susceptible trees about each infected tree, and then generates random numbers for each NN. In the matrix equations outlined in system A.1, we generated $\{R_{0,1}, R_{1,0}, R_{3,2}\}$. Hence, a 'potential' infection matrix $\mathbb{I}^{'}$ is formed:

$$\mathbb{I}^{'} = \begin{pmatrix} 0 & R_{0,1} & 0 \\ R_{1,0} & 2 & 0 \\ 0 & R_{2,1} & 0 \end{pmatrix} \tag{A.2}$$

Each random number is generated between $[0, 1]$ according to a continuous uniform distribution. For a system with infectivity β, a transition into the infected compartment occurs if $R_{i,j} \leqslant \beta$, following a Bernoulli trial. For example, if $R_{0,1} \leqslant \beta$ and $R_{1,0}, R_{2,1} \geq \beta$, then at time $t = 1$ the matrices are updated:

$$\mathbb{S} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \mathbb{I} = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \mathbb{R} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{A.3}$$

Note, for each time-step, the numerical values of infected trees increase by one, i.e $\mathbb{I}_{i,j} += 1$. Finally, when the number of time-steps reaches the infectious lifetime (T), transitions into the removal matrix occur. That is, when $\mathbb{I}_{i,j} = T + 1 \rightarrow 0$ and $\mathbb{R}_{i,j} = 1$.

Throughout the course of a simulation, the steps listed above are iteratively repeated over a set time-horizon, typically $t = 3000$. Eventually, one of the three boundary conditions

(BCs) is met and the simulation ends:

- No infected trees are registered in the domain
- Percolation to either lattice edges are detected
- The number of time-steps reach the time-horizon

Moreover, at $t = 0$, initial conditions (ICs) in the model can vary as:

- A centrally located 'focal' source of infected trees
- A randomly distributed set of infected trees
- N number of infected trees

Listing A.1: An algorithm written in Python to compute matrix equations and simulate disease spread. A GitHub repository contains all the computer code used to generate Chapters 3 and 4: `https://github.com/John-Holden/percolation_tree_model.git`

```python
def run(S, I, R, beta, T=10, L=500):
    """
    Run algorithm
    :param S: array-like, susceptible matrix
    :param I: array-like, infected matrix
    :param R: array-like, removed matrix
    :param beta: float, transmission probability
    :param T: int, infectious life-time of a tree
    :param L: int, lattice dimension
    :return:
    """
    # - Begin - #
    for t in range(3000):
            # nn : nearest neighbours,
            # - single out vertical and horizontal nn respectively
            nn = np.roll(I, 1, axis=0) + np.roll(I, -1, axis=0)
            nn = nn + np.roll(I, 1, axis=1) + np.roll(I, -1, axis=1)
            nn = (nn * S) > 0  # sigle out susceptible trees only
            # inf_dyn : infection dynamics (a probability)
            inf_dyn = np.array(np.random.uniform(size=[L, L]) < beta)
            # add 1 to exitsting ifectes
            # combine neaibourhood to infection status
            I = I + (I > 0) + 2 * nn * inf_dyn.
            S = S * np.where(I > 0, 0, 1) # take away infecteds from S
            R = R + np.where(I == T, 1, 0)  # transition I to R
            R = np.array(R_tree > 0).astype(int). # Hold R as binary
            I = I * np.where(R > 0, 0, 1) # Remove infected status
            # continue...
```

A.2 Towards a continuum model

Here, we move away from the discrete stochastic percolation model outlined previously and examine alternate modelling paradigms. A set of field equations describing the evolution of the probability fields S, I and R for the simple lattice model are described.

From the percolation model in section 3.3, three states were defined for each lattice site i: susceptible, infected, removed S_i, I_i, R_i. The evolution of these lattice sites were dictated by Mote-Carlo steps where each infected lattice site has a chance to infect a nearest neighbour moving a susceptible tree in state S to infectious state I before finally transitioning into the R compartment in $T = 10$ time steps. Each simulation can be seen as an individual realisation of the physical process, however, a great many other potential realisations could have occurred. This is easily demonstrated around percolation threshold where successive iterations lead to differing results, some simulations will percolate and be considered an epidemic while some will not. Even if well above or below the threshold of transmission, infected trees will propagate differently and trace out unique pathways through the domain owing to a very small probability of tracing out exactly the same pathway. Therefore, this individuum paradigm is noisy and only useful for understanding average behavioural quantities when ensemble averaged which is limited in scope by computer memory.

An alternate method, requiring no ensemble-averaging, would be to formulate a set of differential equations which describe the probability-evolution of lattice sites i in the domain being in either states S, I, R. This method would yield the benefit of not needing to ensemble-average results in order to determine average behaviours as a single iteration of the simulation by construction gives us the *mean* field evolution. Furthermore, simulations when animated spatially would give *average* travelling wave behaviour equivalent to running many stochastic simulations and combining frames (which could require large amounts of data). As before, the initial population of trees in the domain is seeded with probability p, with exception of a small number of initially infected trees at the center of the domain. The empty lattice sites at $t = 0$ and the removed lattice sites at $t \geq T$ behave exactly the same remaining non-infectious to susceptible trees therefore both empty and removed lattice sites can be described by state R

$$S_i = p; \quad I_i = 0; \quad R_i = (1 - p) \tag{A.4}$$

Where p is the probability of being occupied by a susceptible tree (*equivalent to tree density*). Now a set of dynamical equations which govern the evolution of fields S, I, R can be outlined by considering the dynamics of the simple lattice model; over a single time-step an infected tree at position i has probability β of infecting a healthy nearest neighbour, this is given by $\beta I(t)_i$ i.e. the probability of transmission multiplied by the probability being infected. Considering a healthy tree at position i surrounded by 4 nearest neighbours j, where j denotes $j = \pm\Delta x$ or $j = \pm\Delta y$, we can describe the probability of S_i remaining healthy by:

$$\prod_j (1 - \beta I(t)_{i+j}) \tag{A.5}$$

where the repeated product gives us the chance of *not being* infected iterated over all nearest neighbours [1]. The reduction of probability in a tree at site i remaining susceptible is then given by:

$$S_i(t + \Delta t) - S_i(t) = -S_i(t) \prod_j \left[1 - \beta I(t)_{i+j}\right] \tag{A.6}$$

From this, the field S can be seen to monotonically decrease. From this point it is easiest to consider the evolution of field R. In the simple percolation model transition times where set to n time-steps before an infected tree transitions into the removed compartment, therefore a tree infected at time-step $t - n\Delta t$ will during time-step $t \to t + \Delta t$. In original simulations in chapter 3.3 the value was held constant at $n = 10$. The change in field R

---

[1] In reality the event of a healthy tree being infected by its neighbours are statistically independent events and therefore calculating the total probability of being infected constitutes a combinatorics problem of combining the union of n events. It is much simpler to consider the single event of *remaining healthy*.

can then be given in terms of field S by:

$$\begin{aligned} R_i(t+\Delta t) - R_i(t) = -\Big[S_i(t - n\Delta t) - S_i(t-(n+1)\Delta t\Big] = \\ S(t-(n+1)\Delta t)\Big[1 - \prod_j \big(1 - \beta I_{i+j}(t-(n+1)\Delta t)\big)\Big] \end{aligned} \tag{A.7}$$

Where the right hand side of the top line is substituted with the right hand side of equation A.6 at time $t-(n+1)\Delta t$. Lastly, noting that all probabilities add to unity, $S_i(t) + I_i(t) + R_i(t) = 1$, the equations which govern the evolution of infectious field I can be written as:

$$\begin{aligned} I_i(t+\Delta t) - I_i(t) = S_i(t)\Big[1 - \prod_j \big(1 - \beta I_{i+j}\big)\Big] - \\ S(t-(n+1)\Delta t)\Big[1 - \prod_j \big(1 - \beta I_{i+j}(t-(n+1)\Delta t)\big)\Big] \end{aligned} \tag{A.8}$$

equations A.6 - A.8 are finite-difference equations and thus describe the evolution of probabilities at lattice site i given a set of initial conditions. These finite difference equations can be iterated over a set of time-steps to model average behaviour of an ideal system away from the critical regime. At criticallity there will be large fluctuations in behaviour, as the system is in a highly chaotic state and does not belong to any one power-law distribution, therefore simulations would deviate quite considerably and fail to show any fractal-like critical structure. Furthermore, equations A.6 - A.8 only describe systems of sufficiently large domain size due to domain sensitivity effects.

A.2.1 Alternative toy landscape SLM metrics

Below in Figure A.1, a distance-based metric is projected onto the example distribution of modelled oak data. In generally, a variety of metrics could be captured and displayed in the heterogeneous landscape SLM; although here, observations of the pathogens maximum distance is recorded over many simulations and ensemble-averaged. In comparison to the mortality ratio (as discussed in section 4.3.4), a similar pattern is witnessed, i.e.

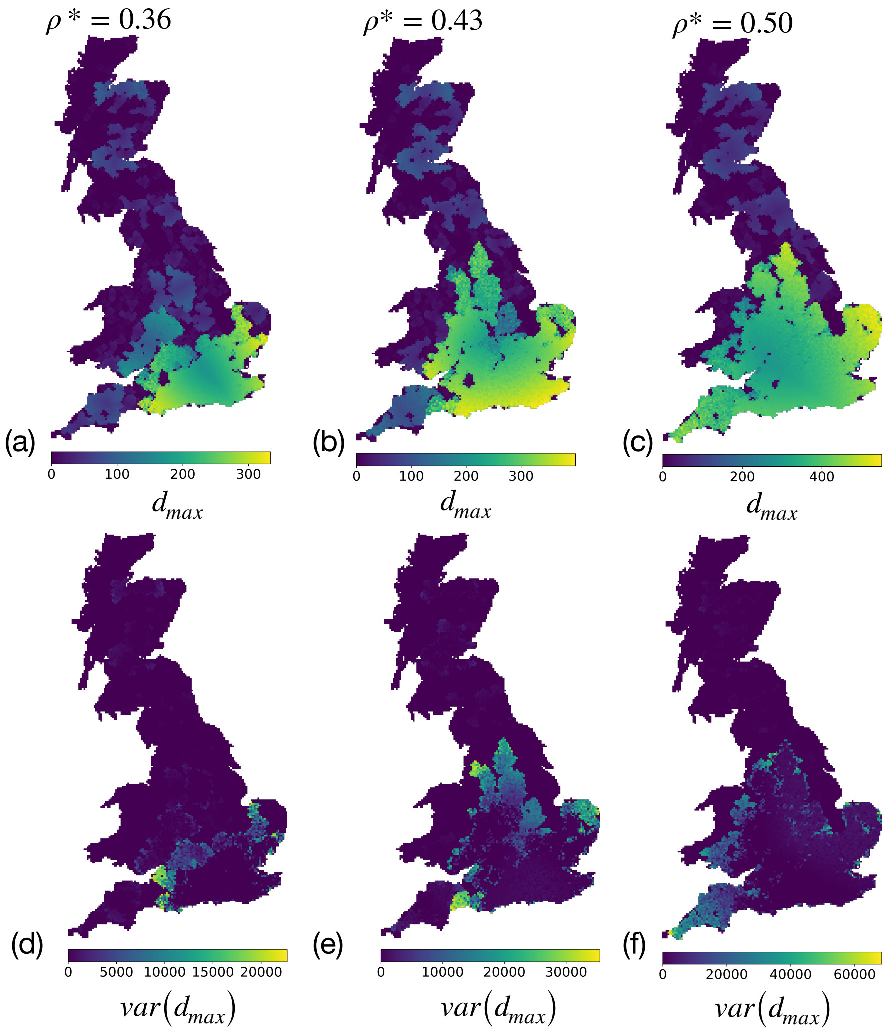


Figure A.1: Spatial plots showing the pathogens ensemble-averaged maximum distance d_{max} for infectiviy $\beta = 0.25$ and density parameters shown.

southerly regions are most effected and maximum distance increases with the effective density.

Nevertheless, Figure A.1 paints a clearer picture of the extent of disease progression. Intuitively, epidemics beginning from effected edge locations in Figures A.1(a-c) can travel further than centrally located epicentres, verified by the more yellow cluster edges. Moreover, the same high-variance regions are witnessed through Figure A.1(d-f). Together, the pathogens maximum distance and mortality ratio complement one another, and as we see from their similarities, a clear correlation exists between them.

Appendix B

The non-local dispersal model

B.1 An alternate R_0 derivation

Starting from an un-normalised Gaussian kernel $g(p,q;\ell) = \exp(\frac{p^2-q^2}{2\ell^2})$ and infectivity constant β, we may define the probability of position q being infected due to an infected tree at p as $Pr(q;p) = \beta g(p,q;\ell)$. The domain has tree density ρ_0 at time $t = 0$ and trees transition through states: $S \rightarrow I \rightarrow R$, with I lasting for T time-steps. Considering the probability of point $q = (x,y)$ becoming infected on account of an infected tree located at the origin during the first time-step:

$$Pr(x,y,t=0) = \beta\rho_0 \exp(-\frac{x^2+y^2}{2\ell^2}) \tag{B.1}$$

Integrating this over an infinite domain gives $R_0(t=0)$ expected infections, given by:

$$R_0(t=0) = \beta\rho_0 \int_{-\infty}^{\infty} \exp(-\frac{x^2+y^2}{2\ell^2})dxdy = 2\pi\beta\rho_0\ell^2 \tag{B.2}$$

At time-step $t+1$ there are less trees to infect. Therefore tree density ρ should also be considered as a monotonically decreasing function of time $\rho(t)$, and the number of expected infections should be given by:

$$R_0(t) = 2\pi\beta\ell^2\rho(t) \tag{B.3}$$

Considering density as a function of time[1] in a discrete domain of size L, the average decrease in tree density over one time-step is given by:

$$\begin{aligned}\rho(t+1) &= \rho(t) - \frac{R_0(t)}{L^2} \\ &= \rho(t)\Big(1 - 2\pi\beta\frac{\ell^2}{L^2}\Big)\end{aligned} \tag{B.4}$$

at $\rho(t=0) = \rho_0$, therefore, equation (B.4) forms a series from which we may expand to give a continuous equation of ρ:

$$\rho(t) = \rho_0\big(1 - 2\pi\beta\frac{\ell^2}{L^2}\big)^t \tag{B.5}$$

upon substitution back into equation (B.3) we have an approximation for how the number of expected infections from one infected tree is expected to change over time:

$$R_0(t) = 2\pi\beta\ell^2\rho_0\big(1 - 2\pi\beta\frac{\ell^2}{L^2}\big)^t \tag{B.6}$$

This expression is compared against numerical simulations in Fig 5.1(c). Then integrating over the infectious life-time $t = T$ gives an approximation to an effective reproductive number denoted by R_0:

$$\begin{aligned}R_0 &= 2\pi\beta\ell^2\rho_0\int_0^T \big(1 - 2\pi\beta\frac{\ell^2}{L^2}\big)^t dt \\ &= 2\pi\beta\ell^2\rho_0\frac{(1 - 2\pi\beta\frac{\ell^2}{L^2})^T - 1}{\ln(1 - 2\pi\beta\frac{\ell^2}{L^2})}\end{aligned} \tag{B.7}$$

---

[1] Density also varies with space as trees are removed quicker for regions closer around the primary infection. However, negating this lead to an easily solvable expression valid for lower-value regimes.

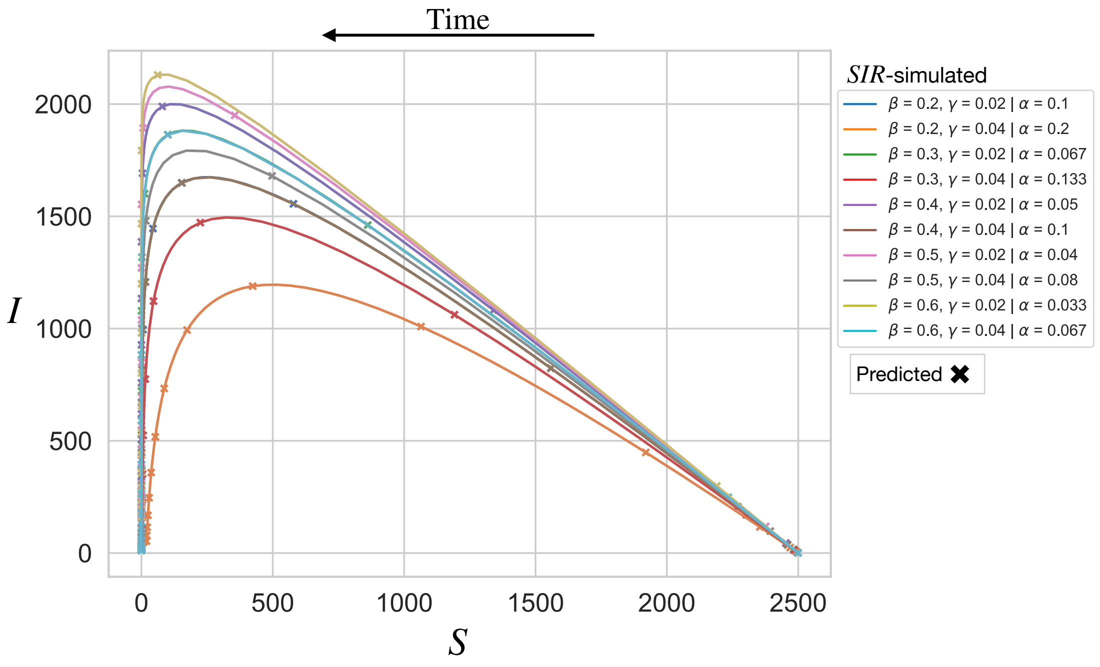


Figure B.1: Infectives are plotted as a function of susceptibles, according to the SIR model for various ratios of $\alpha = \gamma/\beta$. Numerical simulations of the SIR model are plotted against predictions from equation 5.4, shown as crosses. In all simulations, initial conditions began from one infected host and 2500 susceptible hosts—the same number of hosts in a 500 × 500 domain at tree density 0.01.

(from $\int c^t dt = \frac{c^t}{\ln(c)}$). The expression for R_0 can be simplified by noting the pathogen is unlikely to infect trees beyond a distance of 3ℓ, therefore, we can replace the area of the domain with the area over three standard deviations (i.e. $9\pi\ell^2$), thus leading to the approximation:

$$R_0 = 2\pi\beta\rho_0\ell^2 \frac{(1-2/9\beta)^T - 1}{\ln(1-2/9\beta)}$$

B.2 SIR fitting

The standard SIR model has no well-known analytic solution, which complicates model fitting. As such, a simplified scheme that reduced the SIR model to one parameter was used as a comparative tool to access NLM simulations. Previously in chapter 5, details were omitted about the behaviour of equation 5.4, i.e.

$$I(S) = -S + N\Big(1 + \alpha \ln(S/S_0)\Big)$$

In Figure B.1, numerical simulations of the SIR model are compared against analytic predictions from equation 5.4. Specifically, the SIR model was simulated—using the Euler method—for various combinations of infectivity rate β and removal rate γ beginning from one initially infected and 2500 susceptible hosts. Numerical SIR simulations and predictions of infected hosts I (from equation 5.4) are shown as solid lines and crosses respectively. As the ratio $\alpha = \gamma/\beta$ decreases, a sharper rise in the infections field I results, indicating a more infectious outbreak. In contrast, a larger value of α defines a smoother curve which attains a lower peak. Although equation 5.4 has clear limitations and descriptive power, it allows a simple one-parameter model to fit the NLM against.

B.2.1 Exponentially distributed times

In Chapter 5, the NLM was constructed with uniform transitions into the removed compartment. Here, 'uniform' refers to a transition into the R compartment exactly T time-steps after a host becomes infected. Arguably, uniform life-time transitions are simple and unrealistic. As such, the NLM constructed in Chapter 5 was re-run with exponentially-distributed life-times. Figure B.2 shows the SIR fitting procedure against the exponentially-distributed variant of the NLM. Model behaviour in Figure B.2 looks much the same, although fitting the exponentially-distributed NLM to the SIR model resulted in a closer fit for all panels except (b). Exponential life-time are implicit within the standard SIR framework—discussed previously in section 2.1.3. Therefore, it is unsurprising that Figure B.2 generally agrees more with the SIR model, as per equation 5.4. What is surprising is Figure B.2(b), showing a much larger disparity. An explanation of the disparity can be put forward by noting that simulations with fixed, uniform transitions with T lifetimes are typically more infectious that exponentially-distributed lifetimes with mean T. Figure B.2(b) therefore depict simulations with lower epidemic severity than the equivalent panel shown in Chapter 5, Figure 5.2(b). As such, the lower-valued epidemic severity compounds the slowly-spreading wave-like regime (small ℓ, large L) shown in Figure 5.2(b). Consequently, the fitted SIR model shown in red predicts a considerably higher rate of spread.

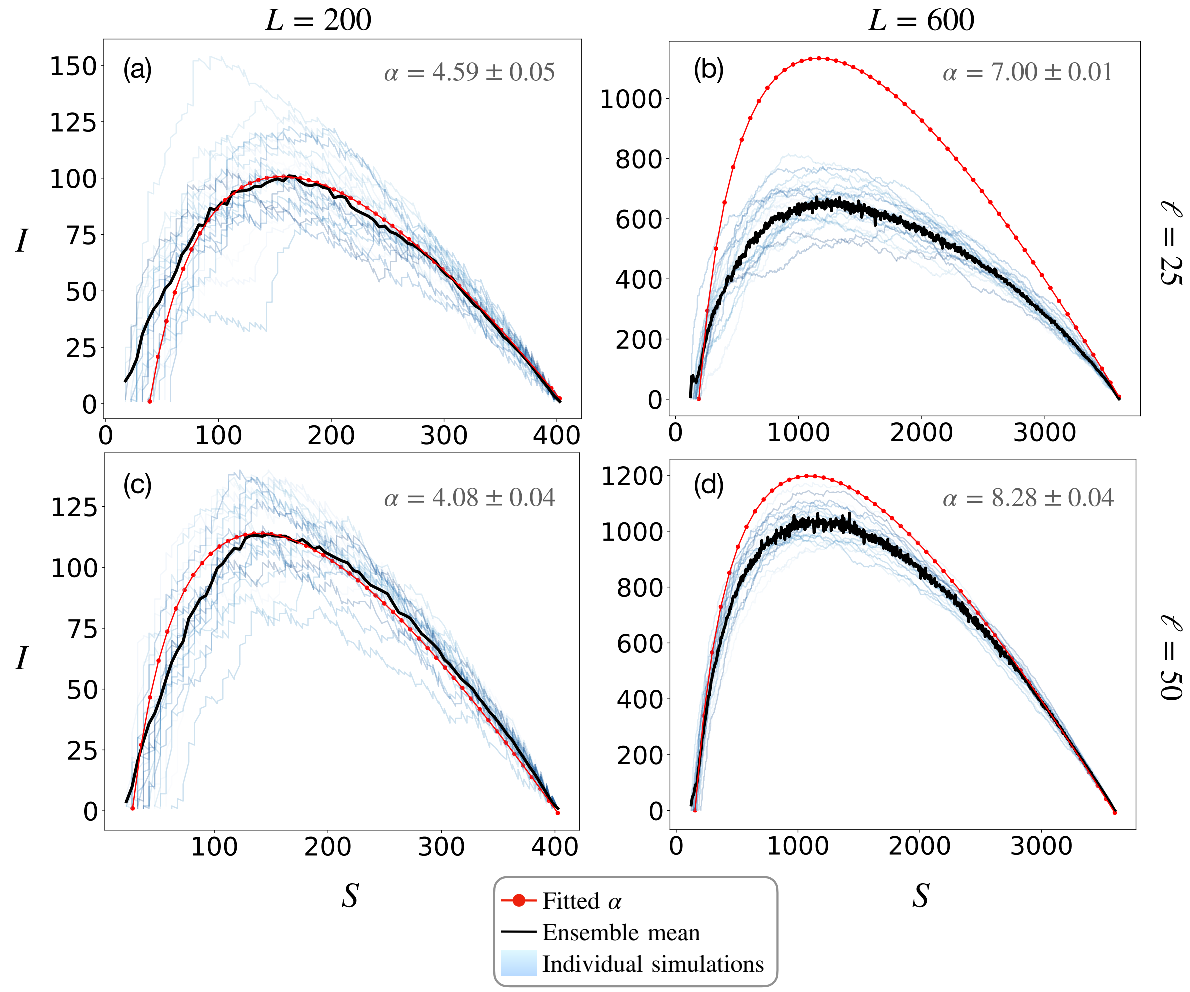


Figure B.2: Fitting the NLM with exponentially-distributed infectious lifetimes to equation 5.4. The equivalent Figure with ‘uniform’ infectious lifetimes was shown in Figure 5.2.

B.3 NLM: probabilistic implementation

Infection probabilities in the NLM are statistically independent events. Suppose that I_{y1} and I_{y2} are two distinct infected trees (where $y1 \neq y2$), and that one susceptible tree S_x at location x exists in the domain. In this case, S_x will experience infection pressure from I_{y1} and I_{y2}. Assuming that I_{y1} and I_{y2} survive over the time-step $t \to t+1$, an infection probability follows the inclusion-exclusion principle:

$$Pr(S_x \to I_x; I_{y1}, I_{y2}) = P_{y1} + P_{y2} - \big[P_{y1} \cap P_{yn}\big] \tag{B.8}$$

where P_{y1} and P_{y2} are the individual probabilities of transition due to I_{y1} and I_{y2} respectively. In Chapters 5 and 6, these infection probabilities followed Gaussian and inverse power-law dispersal models. Now, suppose there are yN infected trees, the inclusion-exclusion principle is generalised to give the well-known form:

$$Pr(S_x \to I_x; I_{y1}, I_{y2}...I_{yN}) = \sum_{k=1}^{N} \big(-1\big)^{k+1} \Big[\sum_{1 \le y1 \le y2 \le .... \le P_{yk} \le N} \big| P_{y1} \cap ... \cap P_{yk} \big| \Big] \tag{B.9}$$

where each intersection $P_{y1} \cap ... \cap P_{yk}$ adds a small order correction to the individual transition probabilities $P_{y1}, P_{y2}, ..., P_{yN}$. Equation B.9 is complicated, unsightly and computationally hard to simulate. Moreover, the definition of R_0 becomes more obscure with equation B.9 because secondary can in principle be induced under the influence of multiple sources.

Nevertheless, equation B.9 can be significantly simplified by considering the probability of remaining susceptible (used also in Appendix A.2). Consider once again the system of two infected trees (I_{y1} and I_{y2}) and one susceptible (S_x). The probability of S_x remaining unaffected is given by:

$$Pr(S_x \to S_x; I_{y1}, I_{y2}) = (1 - P_{y1})(1 - P_{y2}) \tag{B.10}$$

and now the probability of S_x becoming infected is:

$$Pr(S_x \rightarrow I_x; I_{y1}, I_{y2}) = 1 - Pr(S_x \rightarrow I_x; I_{y1}, I_{y2}) \tag{B.11}$$

Generalising equation B.11 to a system of N infected trees and M susceptible trees, over a single time-step we have:

$$\begin{aligned} Pr(S_{x1} \rightarrow I_{x1}) &= 1 - \prod_{n=1}^{N}(1 - P_{yn}) \\ &\vdots \\ Pr(S_{xM} \rightarrow I_{xM}) &= 1 - \prod_{n=1}^{N}(1 - P_{yn}) \end{aligned} \tag{B.12}$$

where the repeated product is introduced for brevity, which combines and multiplies each transition probability, i.e $(1 - P_{y1})(1 - P_{y2})\ldots(1 - P_{yN})$.

In pseudo-code, equations B.12 follow:

```
                # IMPLEMENTATION (A): Full inclusion-exclusion
def run_algorithm(S_tree_arr, I_tree_arr, R_tree_arr, run_times):

    for time_step in range(run_times): # iterate over time-steps
        for S_i in S_tree_arr:  # iterate over susceptible trees
            for I_j in I_tree_arr:
                # iterate & store individual probabilities
                Pr(S_i --> I_i; I_j)
            # combine all probabilities as in equation B.12
            Pr(S_i --> I_i; I_1,I_2,..., I_N)
            update I_tree_arr # updated infected tree arr
            update R_tree_arr # update removed tree arr
            if BCDs:
                end
```

However, upon simulation, it was noted that each individual probability $(P_{y1}...P_{yN})$ is small. Otherwise, the degree of epidemic spread becomes unphysical—consider the limiting behaviour when a single infected tree infects every susceptible in one time-step. For the epidemic parameters of interest, it was noted that combining all probabilities with the inclusion-exclusion principle (as per the system of equations B.12) was unnecessary because small-order corrections (intersection $P_{y1} \cap ... \cap P_{yk}$) become trivially small.

A simpler, more intuitive implementation that neglects the probability intersections can be put forward as:

```
                # IMPLEMENTATION (B): Approx inclusion-exclusion
def run_algorithm(S_tree_arr, I_tree_arr, R_tree_arr, run_times):

    for time_step in range(run_times): # iterate over time-steps
        for I_i in I_tree_arr:  # iterate over infected trees
            for S_j in S_tree_arr:
                # compute individual probabilities
                Pr(S_j --> I_j; I_i)
                update I_tree_arr # updated infected tree arr
                update R_tree_arr # update removed tree arr
                if BCDs:
                    end
```

here, we iterate over *infected* trees rather than susceptible trees. In this scheme, individual probabilities between infected and susceptible trees are computed sequentially. Henceforth, both schemes are denoted as ‘implementation’ (A) and (B), i.e. full and approximated inclusion-exclusion formulas respectively.

B.3.1 Contrasting implementations: epidemic spread

Implementations (A) and (B) are compared in Figure B.3. Plots in Figure B.3 show the number of removed trees in R, or tree mortality, for two values of dispersal $\ell = 25$ and $\ell = 50$ and three values of infectivity. In each panel, 25 ensemble realisations were computed, and both tree density and domain size are fixed to $\rho = 0.01$, $L = 350$, respectively. Overall, both implementations give rise to the same epidemic, illustrated by the similarity between orange and blue plots.

Two values of dispersal were contrasted, motivated by the idea that for a larger dispersal value, the intersections of $(P_{y1}...P_{yN})$ might be larger as trees can interact further apart. Although, no divergence can be seen between ℓ values in Figure B.3, except for a slightly more infectious system when $\ell = 50$—consistent with the results shown in Figure 5.4(c). Indeed, even highly infectious epidemic systems with $\beta^* = 25$ and $\beta^* = 100$ $(10 \lessapprox R_0 \lessapprox 100)$ the spread of disease remains the same. Simulations therefore indicate that for physical epidemic regimes, both implementations are virtually identical. At some

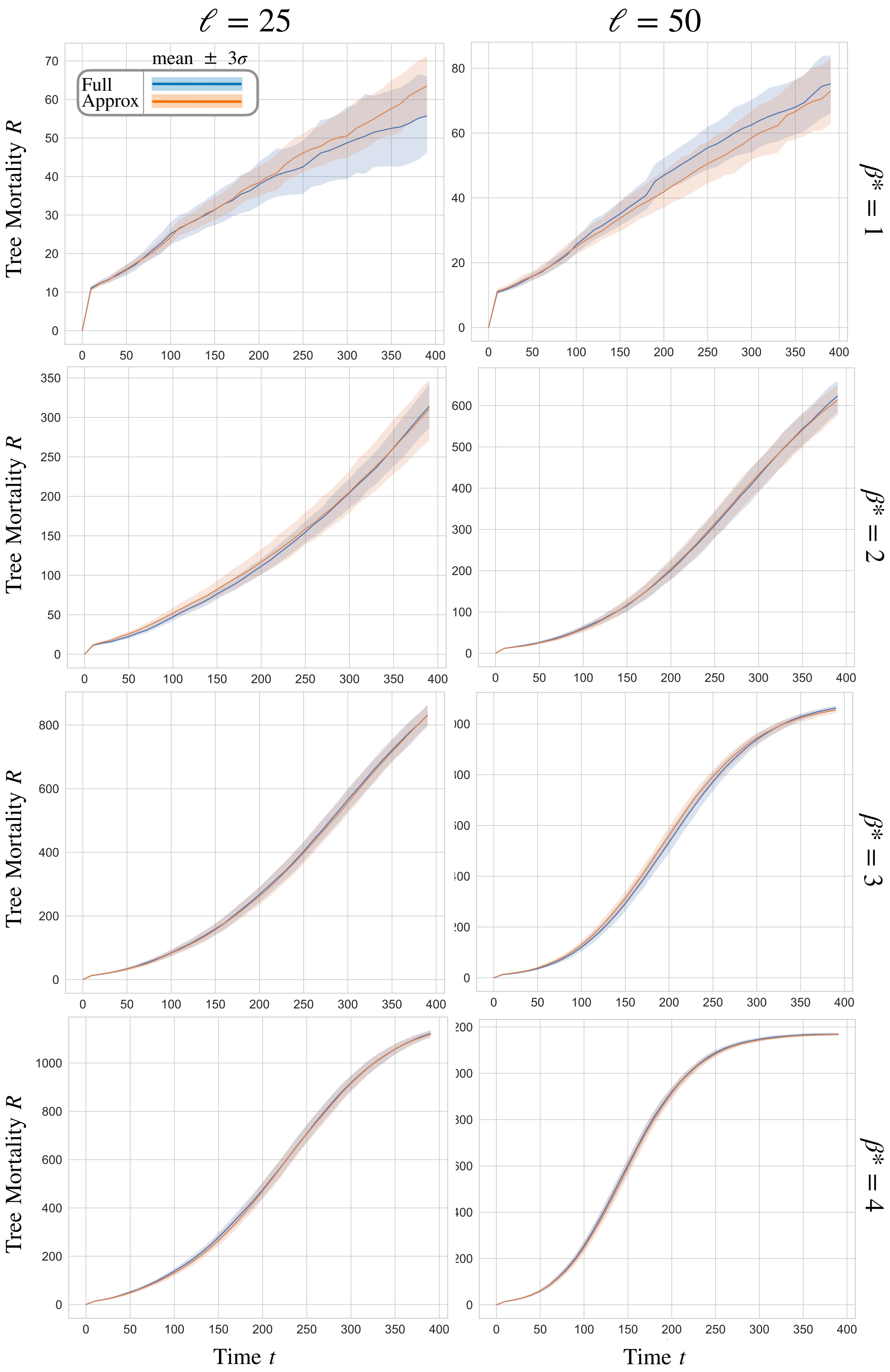


Figure B.3: Comparing both the approximated and full probability implementation methods reveals little to no differences in the tree mortality, or number of removed trees R, over time for different epidemic parameter combinations. All plots show $\rho = 0.01$, $L = 350$, and were seeded by ten initally infected trees.

parameter values, neglecting the intersections $P_{y1} \cap ... \cap P_{yk}$ would undoubtedly deviate from the full inclusion-exclusion formula. However, we have not encountered a realistic epidemic regime were the approximation becomes inaccurate throughout this thesis.

B.3.2 Contrasting implementations: computational cost

Despite epidemic simulations remaining the same, a large difference was noted in the computational cost (and runtime) between both implementations. Figure B.4 contrasts the computer runtime (in seconds) over a range of infectivity parameters and two tree densities. Both panels (a) and (b) depict a system with $\ell = 50$ and $L = 500$ over 10 ensemble realisations. Interestingly, the full inclusion-exclusion formula, as per implementation (A), is more efficient when β^* is large and the number of infected trees is high. Whereas, the approximated inclusion-exclusion formula, as per implementation (B), is more efficient when β^* is low and the number of infected trees is small.

The trends in Figure B.4 can be understood by realising that when infectivity is high, the number of infected trees grows quickly. A large number of computations arise in implementation (A) for lower values of β^*, as we iterate through a large number susceptible trees. Consequently, the runtime is large, particularly when the infection spreads slowly and the number of trees in S remains relatively large throughout the simulation. As β^* is increased, the number of trees in S decrease rapidly, and runtime decreases accordingly. The situation is similar for implementation (B). When β^* is high, a larger number of infected trees entails a larger number of computations as we primarily iterate over the infected trees.

Given the virtually identical spread between both implementation methods (A) and (B), it is alluring to remark about the possibility of switching between implementation (A) for lower and (B) for severe epidemic regimes, as this has the lowest overall runtime and highest computational efficiency. Nevertheless, implementation (B) was chosen throughout the thesis because it is more intuitive and efficient for the epidemic regime we are interested in, typically around the average density of trees in GB and $R_0 < 10$.

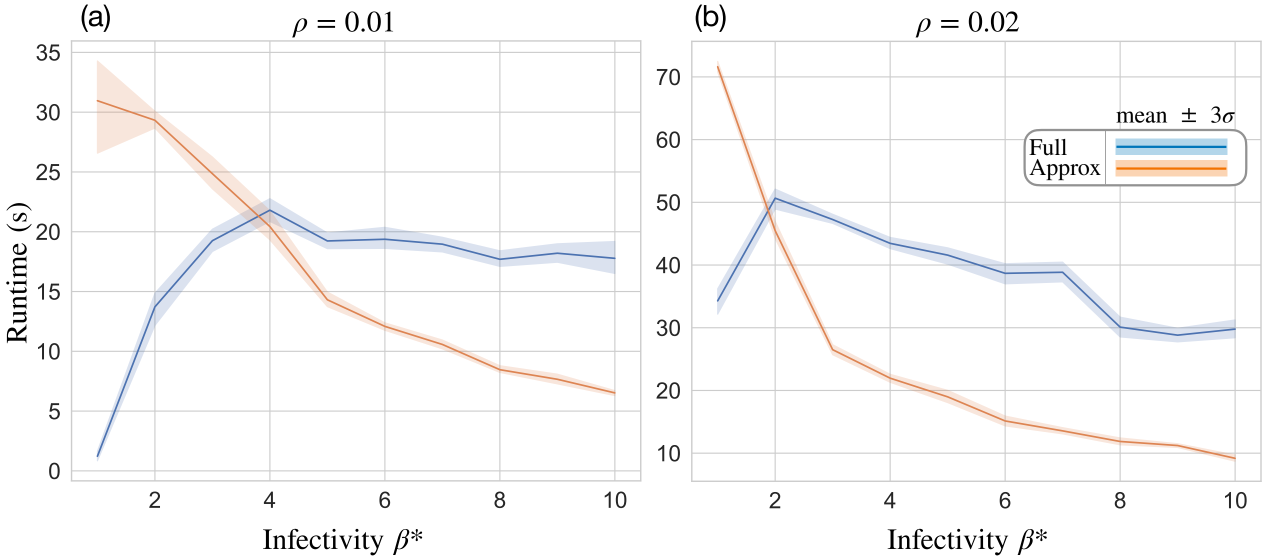


Figure B.4: A comparative look at the difference between simulation runtime (in seconds) between both probability implementations. When the number of infected trees is small for low β^*, the approximated scheme (implementation B) is more efficient and runs quicker. In contrast, when the number of infected trees is high for large β^*, full inclusion-exclusion formula (implementation A) becomes more efficient. Overall, the simulation runtime increases with the tree density.

B.3.3 Codebase

Throughout this thesis, I gravitated heavily toward the software engineering side of the project. As such, a single unified codebase can by found at: `https://github.com/John-Holden/tree_epi_dispersal.git`. In the repository, an epidemic simulator comprising both the NLM and the seasonal $SEIR$ model of ADB can be found and downloaded. I took the approach of developing a flexible, integrated codebase, where the user defines what model is run based on configuration option. In particular, the model configuration is defined by:

- which compartments S, E, I, or R are to be run
- the domain size (L_x, L_y)
- the type of dispersal function: Gaussian/exponential/Inverse power law
- the type of sporulation function: $\phi_0(t) = 1 \ \forall t$ (Ch 5), $\phi_1(t)$, $\phi_2(t)$ (Ch 6,7)
- epidemic parameters, density, infectivity and dispersal (ρ, β, ℓ)

- infection life-times: uniform/exponential dynamics, T steps
- which metrics are to be collected, e.g. time series, R_0, tree mortality, spread velocity
- initial and boundary conditions, e.g. end simulation when all infectives die
- ensemble realisations: iterate a user-defined function N times and store
- whether or not simulations are run on a HPC (Leeds arc3) or the local machine

B.4 Contact-tracing R_0

In Figure B.5(a), we invert the plot (shown in Figure 5.7) and show the range of first generation reproduction ratios $R_0^{(i1)}$ against the ensemble-averaged tree mortality. Inverting the plot gives information about the spread of $R_0^{(1)}$; as we can see, low values of infectiviy produce a skewed distribution which becomes more centered as infectivity increases. A threshold can be seen around $R_0^{(1)} = 1$, although a number of simulations can produce a low-valued $R_0^{(1)}$ for any value of infectivity due to initial extinction events. In Chapter 5 both the contact and analytic values of R_0 were computed/observed for a single infectious tree at the domain center. As elaborated in chapter 5, initial stochastic forces had the tendency to reduce the chance of epidemic by causing early extinction events—thereby reducing the mean tree mortality. However, a number of initial conditions are possible. As such, the plot of Figure 5.7 was re-run with 10 infectious trees in the domain center at $t = 0$ to test how initial stochasticity in the system changes, shown by Figure B.5(b). As expected, increasing the number of infected trees at $t = 0$ reduces stochastic and early extinction events; this is demonstrated by noting that Figure B.5(b) has a smoother ensemble mean, and no instances of zero mortality for highly infectious epidemics (c.f. the bottom right hand side of Figure 5.7 where multiple observations can be seen of zero tree mortality for high β^*). Surprisingly, for later generations the degree of inflexion for $R_0^{(4)}$-$R_0^{(5)}$ is reduced at high β^* in comparison to Figure 5.7.

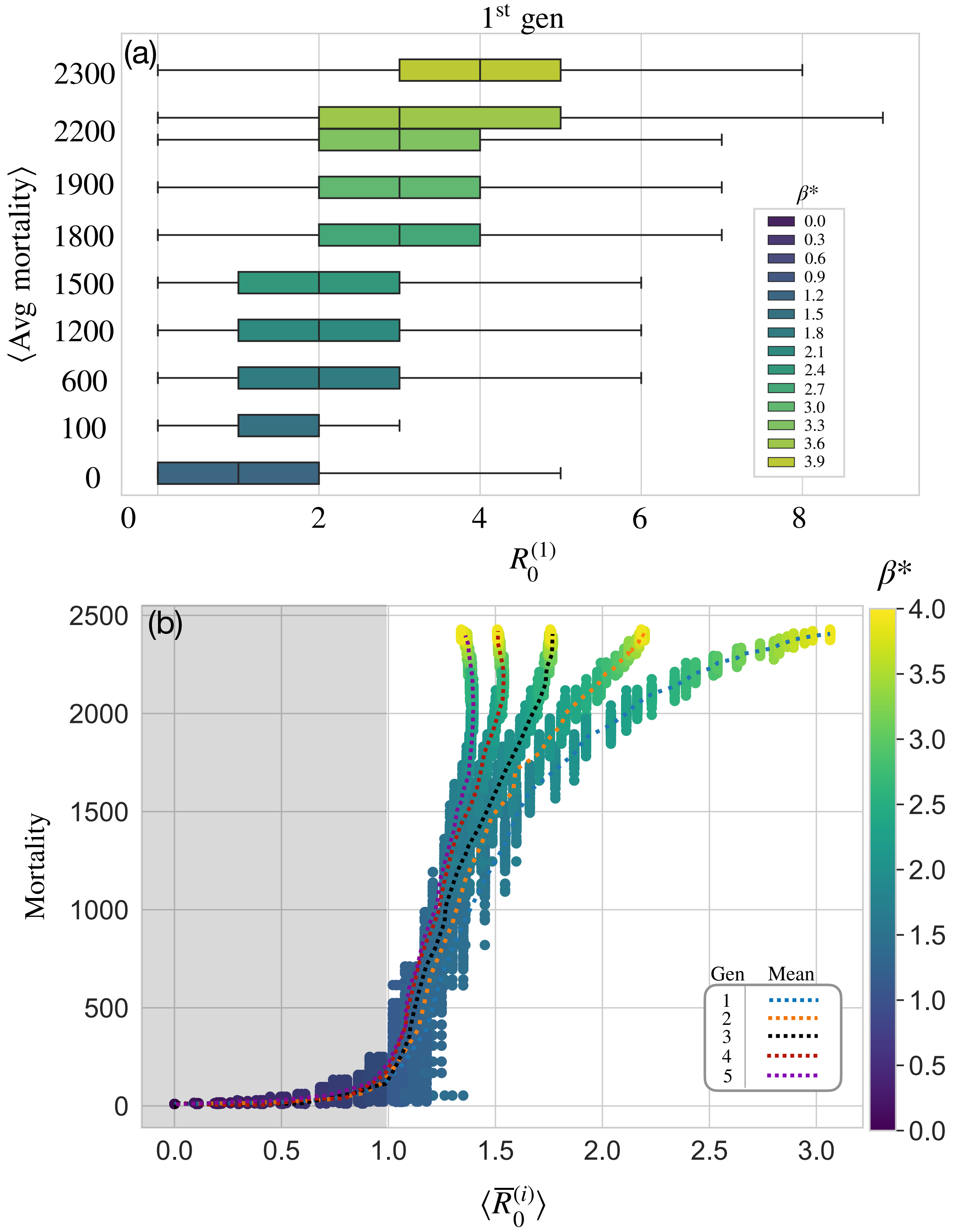


Figure B.5: Comparing the contact-traced reproduction ratio against tree mortality. (a) Inverting the plot of Figure 5.7 to show the spread of $R_0^{(1)}$ against the ensemble averaged tree mortality. (b) Re-running the ensemble shown in Figure 5.7 with 10 initially infected trees. The threshold appears more abrupt and stochasticity is reduced.

Appendix C

Constructing R_0-maps and landscape control

C.1 Connected component analysis

A cluster labelling algorithm was employed to identify and distinguish different clusters of susceptible ($R_0 > 1$) patches in GB. In Python, the function ‘label’ (in the SciPy library) can be used to identify and label different connected clusters. Label requires two pieces of information, the object/image to be analysed and the type of connectivity. Here, we use a 4-connectedness structuring element comprising either Von Neumann or Moore neighbourhoods. The object/image needs to be in the form of a binary matrix to be analysed; this requirement fulfilled through the (Bernoulli trial) tree density implementation, composed of either occupied (1) and unoccupied (0) tree states.

The label function performs the following steps to find the connected sites within the binary matrix:

1. Search for the next unlabelled site, i.
2. Perform a flood-fill algorithm to label all the sites connected to site i.

3. Repeat steps 1 and 2 until all sites are labelled.

The algorithm returns a new structure reflecting the structuring element, the image dimension, the number of distinct clusters found, and the size of each cluster. Hence, a new matrix with labelled clusters is formed, along with a new variable detailing the cluster sizes. The net result is a new image showing each cluster labelled by a unique integer. Through this, all of the sites contained within this cluster have the same unique integer label.

First, the algorithm starts the sweep from the top left-most column then iteratively repeats column-by-column. Two preliminary variables are created to store: A) the number of clusters in the system and B) the size of each cluster found. Both variables are displayed in the output, as separate one-dimensional arrays, denoted by C_s and C_l representing the cluster sizes and the cluster numbers respectively. Using these two variables and the occupancy of sites, the methodology of the algorithm is described in the flowchart shown in Figure 3.1.

Both $\mathbf{C_s}$ and $\mathbf{C_l}$ are an essential part of the process, though they are used in fundamentally different ways. That is, $\mathbf{C_l}$ is used to label each occupied site using the following method:

- An index is set to 1 at the beginning of the algorithm—and increases by 1 upon finding a new cluster.
- If both sites north and west of an occupied site i are unoccupied, site i is given the current index of $\mathbf{C_l}$.
- If the site to the north OR the west of the occupied site i are occupied, the site i is assigned the label of the occupied site.
- If both sites to the north and west of the occupied site i are occupied, then site i and the site to north of site i are both assigned the label of the site to the west of site i.

From $\mathbf{C_l}$, the label j ($j = 1, 2, 3, ...$) of each site can be found through satisfying $\mathbf{C_l}(j) = j$. The label for each site i that will be stored in $\mathbf{C_l}$ is identified using the following argument:

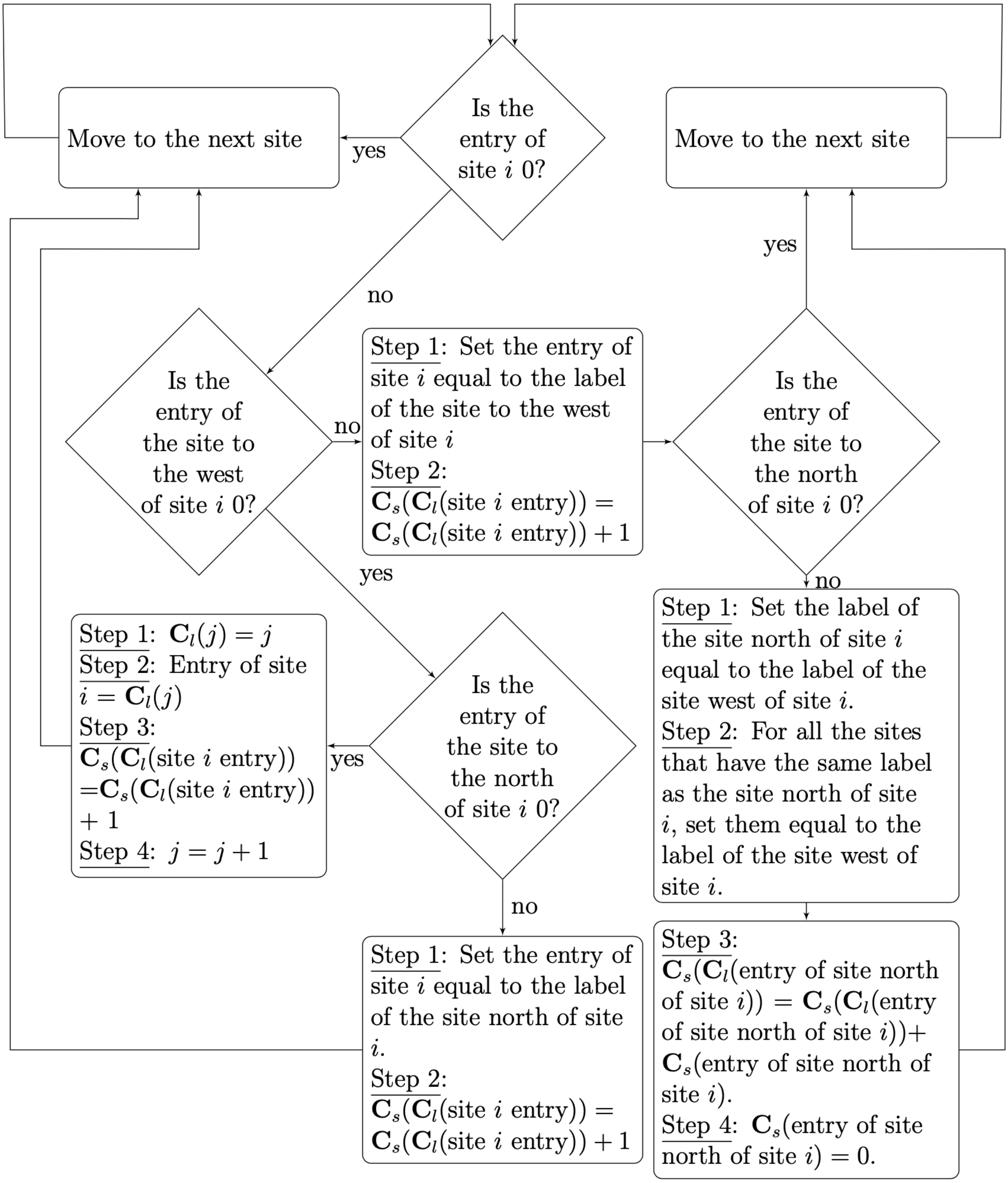


Figure C.1: Flowchart describing the cluster labelling algorithm.

first, let the original entry of site $(i-1)$ be j_1 and the label assigned to site $(i1)$ be j_2. Then, for site i, the following values can be assigned: $\mathbf{C_l}(j1) = j_2 and \mathbf{C_l}(j2) = j_2$, resulting in j_2 being stored as the label for site i. Having established the notion of $\mathbf{C_l}$, the uses of $\mathbf{C_s}$ will now be explored.

The parameter $\mathbf{C_s}$ is used to store the different cluster sizes. Whenever an occupied site is assigned a label, the corresponding entry of C_s is increased by one. When both the sites to the north and west of site i are occupied but have different labels, it is necessary to merge the two clusters together and ensure all sites have identical labels. The cluster size that has the same label as the entry of the site to the north of i is added to the size of the cluster that has the same label as the site to the west of i. The size of the cluster corresponding to the site north of i is then zeroed out. Repeating this method allows the algorithm to perform a complete sweep. This provides an accurate labelling of all occupied sites and the size of all the clusters too. Due to this, only one sweep of the algorithm is necessary [Hoshen and Kopelman, 1976].

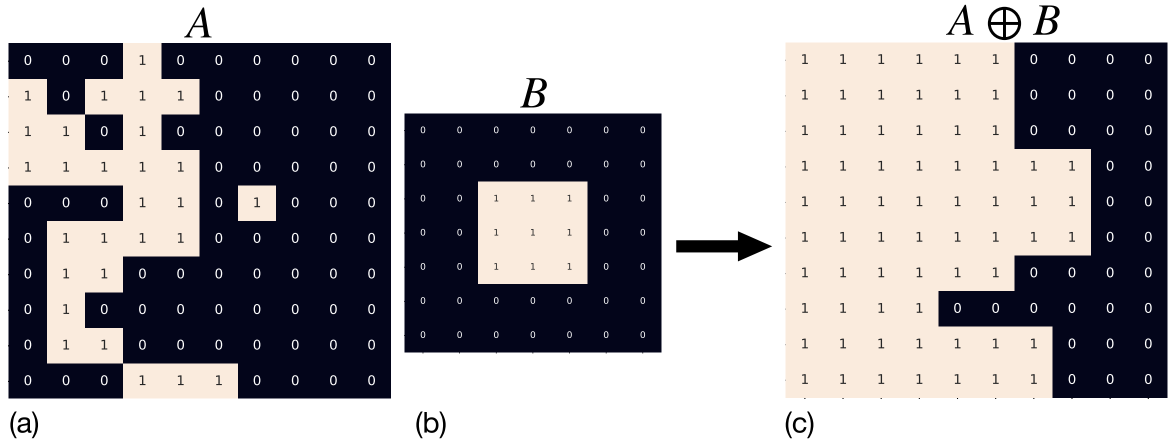
Using the label function creates an efficient way of labelling multiple clusters relatively quickly. The output of the algorithm produces easily distinguishable clusters whilst providing data that can be analysed further in a variety of ways. This demonstrates the importance of a cluster multiple labelling algorithm (or connected component labelling technique) within many applications, be this in percolation theory or other areas of research. One of the main complications of using this method is the computational cost that accompanies it. We will now go on to explore how this has affected choosing the size of the system to analyse.

C.2 Binary dilation

A binary dilator was used to help identify when two distinct clusters ($\mathbf{C_1}$ and $\mathbf{C_2}$) connect. Suppose a we have a binary image in Euclidean space, $\mathbb{R}^{\nvDash}$. For our purpose, $\mathbb{R}^{\nvDash}$ consists of susceptible and insusceptible patches, where $R_0 \geq 1$ and $R_0 < 1$ receptively. The binary dilator operation is defined by:

$$A \oplus B = \bigcup_{b \in B} A_b \tag{C.1}$$

where A_b has underdone a translation by the structuring element b. The meaning of equation C.1 is best described an example:



(a)

A

0	0	0	1	0	0	0	0	0	0
1	0	1	1	1	0	0	0	0	0
1	1	0	1	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0
0	0	0	1	1	0	1	0	0	0
0	1	1	1	1	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	1	1	0	0	0	0	0	0	0
0	0	0	1	1	1	0	0	0	0

(b)

B

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	1	1	1	0	0
0	0	1	1	1	0	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

(c)

$A \oplus B$

1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	0	0	0	0
1	1	1	1	1	1	1	1	0	0
1	1	1	1	1	1	1	1	0	0
1	1	1	1	1	1	1	1	0	0
1	1	1	1	1	1	0	0	0	0
1	1	1	1	0	0	0	0	0	0
1	1	1	1	1	1	1	0	0	0
1	1	1	1	1	1	1	0	0	0

here, A is the binary-valued input image and B is the structuring element, in this case, a Moore neighbourhood. Then, B is super imposed onto every non-zero value in A, thereby forming the binary dilation shown in Figure (c). Binary dilation was implemented in Python, using the 'binary dilation' function within the Scipy.ndimage module [Virtanen et al., 2020].

In section 7.1, I put forward the first steps toward an algorithm to optimise cluster fragmentation. In this method, distinct clusters ($\mathbf{C_1}$ and $\mathbf{C_2}$) connected over some value of ξ—the reader is referred back to section 7.1 for information on ξ. Detecting cluster joins turned out to be non-trivial problem to solve. The naive approach would be a brute-force search to identify which *connecting* patches bridge the gap between $\mathbf{C_1}$ and

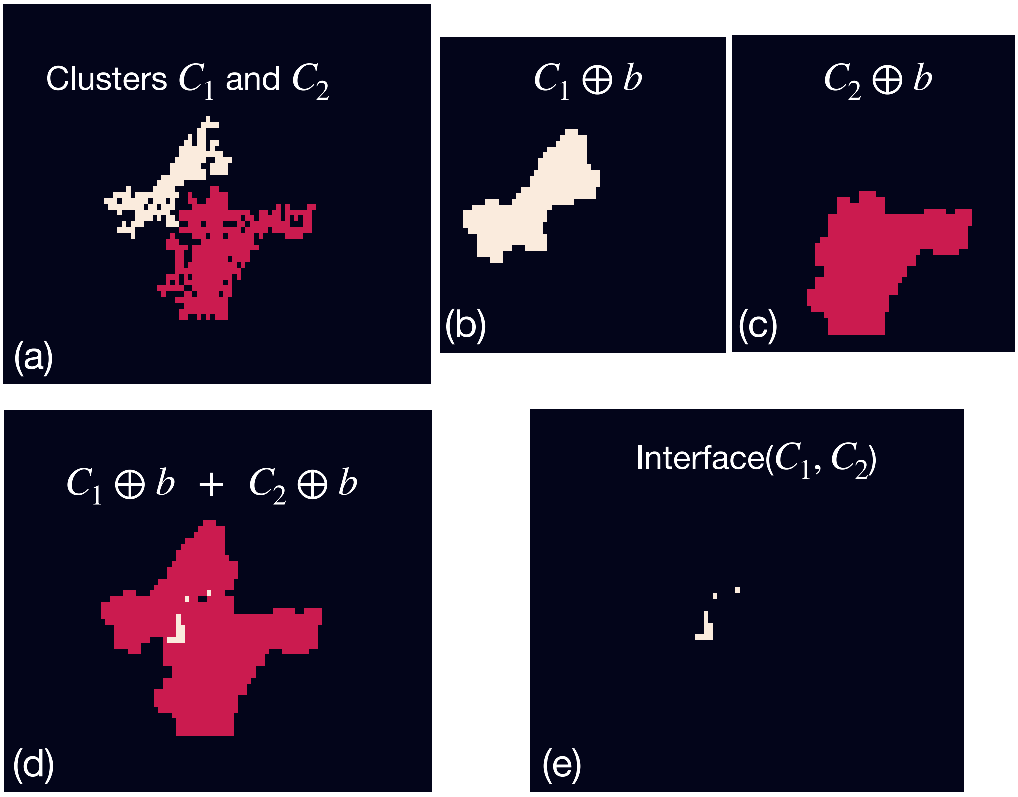


Figure C.2: The binary dilator operator was used to help increase efficiency when identifying which susceptible patches bridge the gap between $\mathbf{C_1}$ and $\mathbf{C_2}$. Here, connectivity is defined with the Moore neighbour, b.

$\mathbf{C_2}$ for each step in ξ. However, this approach quickly became computationally hard to solve, particularly when ξ was low and the number of susceptible patches in the system was large. Below, the following steps detail a more efficient method:

1. Prior to detecting a large rise in cluster size, label clusters $\mathbf{C_1}$ and $\mathbf{C_2}$, shown in Figure C.2(a).
2. Perform a binary dilation on each cluster individually, shown in Figure C.2(b, c).
3. Add both clusters arrays together in a pairwise fashion, shown in Figure C.2(d). The interface between $\mathbf{C_1}$ and $\mathbf{C_2}$ now has the numerical value of 2, easily singled out in Python (e.g. np.where(Arr == 2)), illustrated in Figure C.2(e).
4. One or more of the interface patches must bridge the gap between $\mathbf{C_1}$ and $\mathbf{C_2}$. For each interface patch, test if a Moore neighbourhood falls within both $\mathbf{C_1}$ and $\mathbf{C_2}$; if so, it is a '*connecting*' patch.

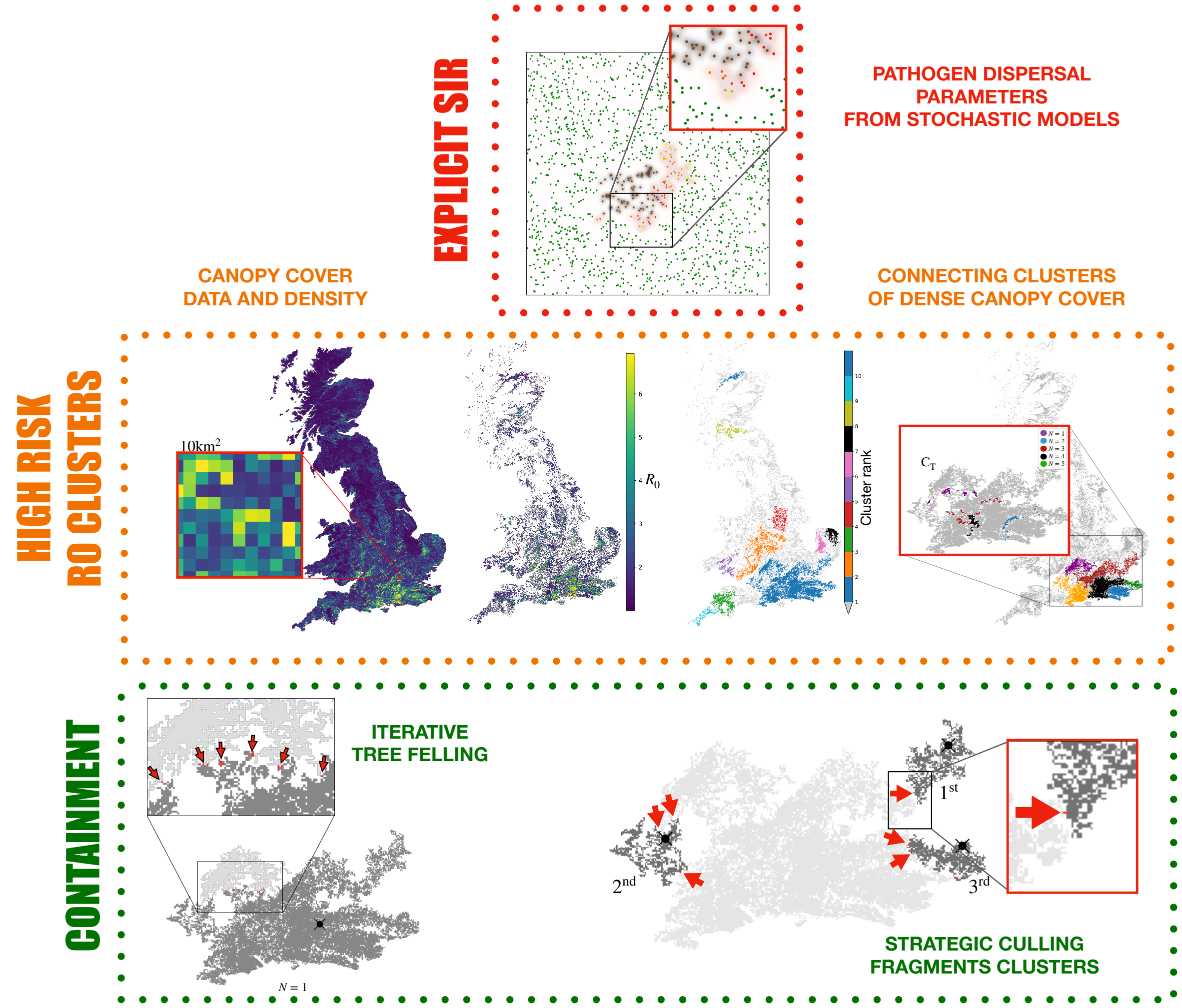


Figure C.3: A graphical abstract showing the end-to-end flow of A) constructing a local-scale spatially-explicit model of pathogen dispersal B) Scaling the small-scale epidemic model over a landscape C) Identifying large susceptible clusters D) Identifying areas for targeted landscape-level control.

.

Appendix D

References

[Abad et al., 2020] Abad, E., Angstmann, C., Henry, B., McGann, A., Le Vot, F., and Yuste, S. (2020). Reaction-diffusion and reaction-subdiffusion equations on arbitrarily evolving domains. *Physical Review E*, 102(3):032111.

[Abramowitz and Stegun, 1948] Abramowitz, M. and Stegun, I. A. (1948). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55. US Government printing office.

[Adler et al., 1983] Adler, J., Zallen, R., Deutscher, G., and ha-fisialit le Yiśrael, A. (1983). *Percolation structures and processes.* Bristol : A. Hilger ; Jerusalem : Israel Physical Society in association with The American Institute of Physics, New York. Includes index.

[Albert et al., 2000] Albert, R., Jeong, H., and Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *nature*, 406(6794):378–382.

[Alonso Chavez et al., 2015] Alonso Chavez, V., Parnell, S., and Van den Bosch, F. (2015). Designing strategies for epidemic control in a tree nursery: the case of ash dieback in the uk. *Forests*, 6(11):4135–4145.

[Alsop et al., 2015] Alsop, J. et al. (2015). Ash dieback (hymenoscyphus fraxineus) experience from continental europe. *Quarterly Journal of Forestry*, 109(3):174–180.

[Atkeson et al., 2020] Atkeson, A. et al. (2020). On using sir models to model disease scenarios for covid-19. *Quarterly Review*, 41(01):1–35.

[Austerlitz et al., 2004] Austerlitz, F., Dick, C. W., Dutech, C., Klein, E. K., Oddou-Muratorio, S., Smouse, P. E., and Sork, V. L. (2004). Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, 13(4):937–954.

[Bailey et al., 2000] Bailey, D. J., Otten, W., and Gilligan, C. A. (2000). Saprotrophic invasion by the soil-borne fungal plant pathogen rhizoctonia solani and percolation thresholds. *The New Phytologist*, 146(3):535–544.

[Baily, 1975] Baily, N. T. (1975). The mathematical theory of infectious diseases. *Griffin, London.*

[Baker et al., 1974] Baker, K., Cook, R. J., et al. (1974). *Biological control of plant pathogens.* WH Freeman and Company.

[Balloux and van Dorp, 2017] Balloux, F. and van Dorp, L. (2017). Q&a: What are pathogens, and what have they done to and for us? *BMC biology*, 15(1):1–6.

[Becker and Ziff, 2009] Becker, A. M. and Ziff, R. M. (2009). Percolation thresholds on two-dimensional voronoi networks and delaunay triangulations. *Phys. Rev. E*, 80:041101.

[Bengtsson et al., 2014] Bengtsson, S. B. K., Barklund, P., von Brömssen, C., and Stenlid, J. (2014). Seasonal pattern of lesion development in diseased fraxinus excelsior infected by hymenoscyphus pseudoalbidus. *PLoS One*, 9(4):e76429.

[Benincà et al., 2020] Benincà, E., Hagenaars, T., Boender, G. J., van de Kassteele, J., and van Boven, M. (2020). Trade-off between local transmission and long-range dispersal drives infectious disease outbreak size in spatially structured populations. *PLoS Computational Biology*, 16(7):e1008009.

[Berger, 1981] Berger, R. (1981). Comparison of the gompertz and logistic equations to describe plant disease progress. *Phytopathology*, 71(7):716–719.

[Bhatt et al., 2013] Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., Drake, J. M., Brownstein, J. S., Hoen, A. G., Sankoh, O., et al. (2013). The global distribution and burden of dengue. *Nature*, 496(7446):504–507.

[Bianco et al., 2013] Bianco, F., Chibbaro, S., Vergni, D., and Vulpiani, A. (2013). Reaction spreading on percolating clusters. *Physical Review E*, 87(6):062811.

[Bock et al., 2010] Bock, C., Graham, J., Gottwald, T., Cook, A., and Parker, P. (2010). Wind speed effects on the quantity of xanthomonas citri subsp. citri dispersed downwind from canopies of grapefruit trees infected with citrus canker. *Plant Disease*, 94(6):725–736.

[Boyd et al., 2013a] Boyd, I., Freer-Smith, P., Gilligan, C., and Godfray, H. (2013a). The consequence of tree pests and diseases for ecosystem services. *Science*, 342(6160).

[Boyd et al., 2013b] Boyd, I. L., Freer-Smith, P. H., Gilligan, C. A., and Godfray, H. C. J. (2013b). The consequence of tree pests and diseases for ecosystem services. *Science*, 342(6160).

[Brasier, 1991] Brasier, C. (1991). Ophiostoma novo-ulmi sp. nov., causative agent of current dutch elm disease pandemics. *Mycopathologia*, 115(3):151–161.

[Brasier and Webber, 2010] Brasier, C. and Webber, J. (2010). Sudden larch death. *Nature*, 466(7308):824–825.

[Brasier et al., 2005] Brasier, C. M., Beales, P. A., Kirk, S. A., Denman, S., and Rose, J. (2005). Phytophthora kernoviae sp. nov., an invasive pathogen causing bleeding stem lesions on forest trees and foliar necrosis of ornamentals in the uk. *Mycological Research*, 109(8):853–859.

[Britton et al., 1986] Britton, N. F. et al. (1986). *Reaction-diffusion equations and their applications to biology.* Academic Press.

[Broadbent and Hammersley, 1957] Broadbent, S. R. and Hammersley, J. M. (1957). Percolation processes: I. crystals and mazes. 53(3):629–641.

[Brown and Hovmøller, 2002] Brown, J. K. and Hovmøller, M. S. (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. *Science*, 297(5581):537–541.

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

[Brugha and Varvasovszky, 2000] Brugha, R. and Varvasovszky, Z. (2000). Stakeholder analysis: a review. *Health policy and planning*, 15(3):239–246.

[Buckle and Smith, 2015] Buckle, A. P. and Smith, R. H. (2015). *Rodent pests and their control*. CABI.

[Bullock et al., 2017] Bullock, J. M., Mallada González, L., Tamme, R., Götzenberger, L., White, S. M., Pärtel, M., and Hooftman, D. A. (2017). A synthesis of empirical plant dispersal kernels. *Journal of Ecology*, 105(1):6–19.

[Bussell and Cunniffe, 2020] Bussell, E. H. and Cunniffe, N. J. (2020). Applying optimal control theory to a spatial simulation model of sudden oak death: ongoing surveillance protects tanoak while conserving biodiversity. *Journal of The Royal Society Interface*, 17(165):20190671.

[Bussell et al., 2019] Bussell, E. H., Dangerfield, C. E., Gilligan, C. A., and Cunniffe, N. J. (2019). Applying optimal control theory to complex epidemiological models to inform real-world disease management. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1776):20180284.

[Butt and Royle, 1974] Butt, D. and Royle, D. (1974). Multiple regression analysis in the epidemiology of plant diseases. In *Epidemics of Plant Diseases*, pages 78–114. Springer.

[Camarretta et al., 2020] Camarretta, N., Harrison, P. A., Bailey, T., Potts, B., Lucieer, A., Davidson, N., and Hunt, M. (2020). Monitoring forest structure to guide adaptive management of forest restoration: a review of remote sensing approaches. *New Forests*, 51(4):573–596.

[Cardy and Grassberger, 1985] Cardy, J. L. and Grassberger, P. (1985). Epidemic models and percolation. *Journal of Physics A: Mathematical and Theoretical*, 18(6):l267.

[Carling and Leiner, 1990] Carling, D. and Leiner, R. (1990). Effect of temperature on virulence of rhizoctonia solani and other rhizoctonia on potato. *Phytopathology (USA).*

[Chami et al., 2017] Chami, G. F., Ahnert, S. E., Kabatereine, N. B., and Tukahebwa, E. M. (2017). Social network fragmentation and community health. *Proceedings of the National Academy of Sciences*, 114(36):E7425–E7431.

[Chandelier et al., 2017] Chandelier, A., Delahaye, L., Claessens, H., Lassois, L., et al. (2017). Ash dieback in wallonia, southern belgium: research on disease development, resistance and management options. pages 53–60.

[Chandelier et al., 2014] Chandelier, A., Helson, M., Dvorak, M., and Gischer, F. (2014). Detection and quantification of airborne inoculum of h ymenoscyphus pseudoalbidus using real-time pcr assays. *Plant pathology*, 63(6):1296–1305.

[Chaudhary, 2020] Chaudhary, R. (2020). Identification of molecular markers associated with fungal resistance in norway spruce and common ash.

[Chela Fenille et al., 2002] Chela Fenille, R., Luiz de Souza, N., and Eurya Kuramae, E. (2002). Characterization of rhizoctonia solani associated with soybean in brazil. *European journal of plant pathology*, 108(8):783–792.

[Chowell et al., 2009] Chowell, G., Hyman, J. M., Bettencourt, L. M., Castillo-Chavez, C., and Nishiura, H. (2009). *Mathematical and statistical estimation approaches in epidemiology.* Springer.

[Chumanová et al., 2019] Chumanová, E., Romportl, D., Havrdová, L., Zahradník, D., Pešková, V., and Černỳ, K. (2019). Predicting ash dieback severity and environmental suitability for the disease in forest stands. *Scandinavian Journal of Forest Research*, 34(4):254–266.

[Cleary et al., 2013] Cleary, M. R., Daniel, G., and Stenlid, J. (2013). Light and scanning electron microscopy studies of the early infection stages of hymenoscyphus pseudoalbidus on fraxinus excelsior. *Plant Pathology*, 62(6):1294–1301.

[Cobb et al., 2017] Cobb, R. C., Hartsough, P., Ross, N., Klein, J., LaFever, D. H., Frankel, S. J., and Rizzo, D. M. (2017). Resiliency or restoration: management of sudden oak death before and after outbreak. *Forest Phytophthoras*, 7(1):1–14.

[Coetsee et al., 2000] Coetsee, C., Wingfield, M., Crous, P., and Wingfield, B. (2000). Xenochalara, a new genus of dematiaceous hyphomycetes for chalara-like fungi with apical wall building conidial development. *South African journal of botany*, 66(2):99–103.

[Coker et al., 2019] Coker, T. L. R., Rozsypálek, J., Edwards, A., Harwood, T. P., Butfoy, L., and Buggs, R. J. A. (2019). Estimating mortality rates of european ash (fraxinus excelsior) under the ash dieback (hymenoscyphus fraxineus) epidemic. *PLANTS, PEOPLE, PLANET*, 1(1):48–58.

[Cook et al., 2008] Cook, A., Gibson, G., Gottwald, T., and Gilligan, C. (2008). Constructing the effect of alternative intervention strategies on historic epidemics. *Journal of the Royal Society Interface*, 5(27):1203–1213.

[Courant et al., 1967] Courant, R., Friedrichs, K., and Lewy, H. (1967). On the partial difference equations of mathematical physics. *IBM Journal of Research and Development*, 11(2):215–234.

[Cowger et al., 2005] Cowger, C., Wallace, L. D., and Mundt, C. C. (2005). Velocity of spread of wheat stripe rust epidemics. *Phytopathology*, 95(9):972–982.

[Cox and Durrett, 1988] Cox, J. and Durrett, R. (1988). Limit theorems for the spread of epidemics and forest fires. *Stochastic Processes and their Applications*, 30(2):171–191.

[Cunniffe et al., 2016] Cunniffe, N. J., Cobb, R. C., Meentemeyer, R. K., Rizzo, D. M., and Gilligan, C. A. (2016). Modeling when, where, and how to manage a forest epidemic, motivated by sudden oak death in california. *Proceedings of the National Academy of Sciences*, 113(20):5640–5645.

[Cunniffe et al., 2015a] Cunniffe, N. J., Koskella, B., Metcalf], C. J. E., Parnell, S., Gottwald, T. R., and Gilligan, C. A. (2015a). Thirteen challenges in modelling plant diseases. *Epidemics*, 10:6 – 10. Challenges in Modelling Infectious DIsease Dynamics.

[Cunniffe et al., 2015b] Cunniffe, N. J., Stutt, R. O. J. H., DeSimone, R. E., Gottwald, T. R., and Gilligan, C. A. (2015b). Optimising and communicating options for the control of invasive plant disease when there is epidemiological uncertainty. *PLOS Computational Biology*, 11(4):1–24.

[Cunniffe et al., 2012] Cunniffe, N. J., Stutt, R. O. J. H., van den Bosch, F., and Gilligan, C. A. (2012). Time-dependent infectivity and flexible latent and infectious periods in compartmental models of plant disease. *Phytopathology®*, 102(4):365–380. PMID: 22106830.

[Dal Maso and Montecchio, 2014] Dal Maso, E. and Montecchio, L. (2014). Risk of natural spread of hymenoscyphus fraxineus with environmental niche modelling and ensemble forecasting technique. *Forest Research*, 3(4):1–11.

[Dandy et al., 2017] Dandy, N., Marzano, M., Porth, E. F., Urquhart, J., and Potter, C. (2017). Who has a stake in ash dieback? a conceptual framework for the identification and categorisation of tree health stakeholders.

[Dangl and Jones, 2001] Dangl, J. L. and Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *nature*, 411(6839):826–833.

[De Bach et al., 1964] De Bach, P. et al. (1964). Biological control of insect pests and weeds. *Biological control of insect pests and weeds.*

[Defra, 2013] Defra (2013). Chalara management plan.

[Delamater et al., 2019] Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T., and Jacobsen, K. H. (2019). Complexity of the basic reproduction number (r0). *Emerging infectious diseases*, 25(1):1.

[Desprez-Loustau et al., 2016] Desprez-Loustau, M.-L., Aguayo, J., Dutech, C., Hayden, K. J., Husson, C., Jakushkin, B., Marçais, B., Piou, D., Robin, C., and Vacher, C. (2016). An evolutionary ecology perspective to address forest pathology challenges of today and tomorrow. *Annals of Forest Science*, 73(1):45–67.

[Devaux et al., 2007] Devaux, c., Lavigne, c., Austerlitz, F., and Klein, E. K. (2007). Modelling and estimating pollen movement in oilseed rape (brassica napus) at the landscape scale using genetic markers. *Molecular Ecology*, 16(3):487–499.

[Dixon, 1976] Dixon, K. (1976). Analysis of seasonal leaf fall in north temperate deciduous forests. *Oikos*, pages 300–306.

[Dobrowolska et al., 2011a] Dobrowolska, D., Hein, S., Oosterbaan, A., Wagner, S., Clark, J., and Skovsgaard, J. P. (2011a). A review of European ash (Fraxinus excelsior L.): implications for silviculture. *Forestry: An International Journal of Forest Research*, 84(2):133–148.

[Dobrowolska et al., 2011b] Dobrowolska, D., Hein, S., Oosterbaan, A., Wagner, S., Clark, J., and Skovsgaard, J. P. (2011b). A review of european ash (fraxinus excelsior l.): implications for silviculture. *Forestry*, 84(2):133–148.

[Drake and Griffen, 2010] Drake, J. M. and Griffen, B. D. (2010). Early warning signals of extinction in deteriorating environments. *Nature*, 467(7314):456–459.

[Eames and Keeling, 2003] Eames, K. T. and Keeling, M. J. (2003). Contact tracing and disease control. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1533):2565–2571.

[(EFSA) et al., 2022] (EFSA), E. F. S. A., Delbianco, A., Gibin, D., Pasinato, L., and Morelli, M. (2022). Update of the xylella spp. host plant database–systematic literature search up to 30 june 2021. *EFSA Journal*, 20(1):e07039.

[Enderle et al., 2017] Enderle, R., Fussi, B., Lenz, H., Langer, G., Nagel, R., Metzler, B., et al. (2017). Ash dieback in germany: research on disease development, resistance and management options. *Dieback of European ash (Fraxinus spp.): consequences and guidelines for sustainable management*, pages 89–105.

[Enderle et al., 2013] Enderle, R., Peters, F., Nakou, A., and Metzler, B. (2013). Temporal development of ash dieback symptoms and spatial distribution of collar rots in a provenance trial of fraxinus excelsior. *European journal of forest research*, 132(5):865–876.

[Enderle et al., 2019] Enderle, R., Stenlid, J., Vasaitis, R., et al. (2019). An overview of ash (fraxinus spp.) and the ash dieback disease in europe. *CAB Rev*, 14:1–12.

[Essam, 1980] Essam, J. W. (1980). Percolation theory. *Reports on Progress in Physics*, 43(7):833–912.

[Fabre et al., 2021] Fabre, F., Coville, J., and 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*, pages 47–72. Springer.

[Family and Vicsek, 1985] Family, F. and Vicsek, T. (1985). Scaling of the active zone in the eden process on percolation networks and the ballistic deposition model. *Journal of Physics A: Mathematical and General*, 18(2):L75–L81.

[Ferrandino, 1993] Ferrandino, F. J. (1993). Dispersive epidemic waves: I. focus expansion within a linear planting. *Phytopathology*, 83(8):795–802.

[Filipe et al., 2004] Filipe, J., Otten, W., Gibson, G. J., and Gilligan, C. A. (2004). Inferring the dynamics of a spatial epidemic from time-series data. *Bulletin of Mathematical Biology*, 66(2):373–391.

[Filipe et al., 2012] Filipe, J. A. N., Cobb, R. C., Meentemeyer, R. K., Lee, C. A., Valachovic, Y. S., Cook, A. R., Rizzo, D. M., and Gilligan, C. A. (2012). Landscape epidemiology and control of pathogens with cryptic and long-distance dispersal: Sudden oak death in northern californian forests. *PLOS Computational Biology*, 8:1–13.

[Fisher, 1969] Fisher, M. E. (1969). Rigorous inequalities for critical-point correlation exponents. *Phys. Rev.*, 180:594–600.

[Fisher, 1937] Fisher, R. A. (1937). The wave of advance of advantageous genes. *Annals of eugenics*, 7(4):355–369.

[Flor, 1971] Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annual review of phytopathology*, 9(1):275–296.

[Fones et al., 2016] Fones, H. N., Mardon, C., and Gurr, S. J. (2016). A role for the asexual spores in infection of fraxinus excelsior by the ash-dieback fungus hymenoscyphus fraxineus. *Scientific Reports*, 6(1):1–10.

[Forster and Gilligan, 2007] Forster, G. A. and Gilligan, C. A. (2007). Optimizing the control of disease infestations at the landscape scale. *Proceedings of the National Academy of Sciences*, 104(12):4984–4989.

[Franklin, 2013] Franklin, J. (2013). Species distribution models in conservation biogeography: developments and challenges.

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

[Fuller et al., 2016] Fuller, L., Marzano, M., Peace, A., Quine, C. P., and Dandy, N. (2016). Public acceptance of tree health management: Results of a national survey in the uk. *Environmental Science & Policy*, 59:18–25.

[Garrett et al., 2006] Garrett, K. A., Dendy, S. P., Frank, E. E., Rouse, M. N., and Travers, S. E. (2006). Climate change effects on plant disease: genomes to ecosystems.

Annu. Rev. Phytopathol., 44:489–509.

[Gaston et al., 2000] Gaston, K. J., Blackburn, T. M., Greenwood, J. J., Gregory, R. D., Quinn, R. M., and Lawton, J. H. (2000). Abundance–occupancy relationships. *Journal of Applied Ecology*, 37:39–59.

[Gaydos et al., 2019] Gaydos, D. A., Petrasova, A., Cobb, R. C., and Meentemeyer, R. K. (2019). Forecasting and control of emerging infectious forest disease through participatory modelling. *Philosophical Transactions of the Royal Society B*, 374(1776):20180283.

[Gibbs, 1902] Gibbs, J. W. (1902). *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics.* C. Scribner's sons.

[Gibson et al., 2006] Gibson, G. J., Otten, W., N Filipe, J., Cook, A., Marion, G., and Gilligan, C. A. (2006). Bayesian estimation for percolation models of disease spread in plant populations. *Statistics and Computing*, 16(4):391–402.

[Gilioli et al., 2013] Gilioli, G., Bodini, A., and Baumgärtner, J. (2013). Metapopulation modelling and area-wide pest management strategies evaluation. an application to the pine processionary moth. *Ecological Modelling*, 260:1 – 10.

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

[Gilligan et al., 2007a] Gilligan, C. A., Truscott, J. E., and Stacey, A. J. (2007a). Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. *Journal of The Royal Society Interface*, 4(16):925–934.

[Gilligan et al., 2007b] Gilligan, C. A., Truscott, J. E., and Stacey, A. J. (2007b). Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. *Journal of the Royal Society Interface*, 4(16):925–934.

[Gilligan and van den Bosch, 2008] Gilligan, C. A. and van den Bosch, F. (2008). Epidemiological models for invasion and persistence of pathogens. *Annu. Rev. Phytopathol.*, 46:385–418.

[Golan and Pringle, 2017] Golan, J. J. and Pringle, A. (2017). Long-distance dispersal of fungi. *Microbiology spectrum*, 5(4):5–4.

[Gottwald et al., 2002] Gottwald, T. R., Sun, X., Riley, T., Graham, J. H., Ferrandino, F., and Taylor, E. L. (2002). Geo-referenced spatiotemporal analysis of the urban citrus canker epidemic in florida. *Phytopathology*, 92(4):361–377.

[Grassberger, 1983] Grassberger, P. (1983). On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157 – 172.

[Grassberger, 1986] Grassberger, P. (1986). Spreading of epidemic processes leading to fractal structures. In PIETRONERO, L. and TOSATTI, E., editors, *Fractals in Physics*, pages 273 – 278. Elsevier, Amsterdam.

[Gregory, 1968] Gregory, P. H. (1968). Interpreting plant disease dispersal gradients. *Annual Review of Phytopathology*, 6(1):189–212.

[Grenfell and Harwood, 1997] Grenfell, B. and Harwood, J. (1997). (meta)population dynamics of infectious diseases. *Trends in Ecology Evolution*, 12(10):395 – 399.

[Grosdidier et al., 2018] Grosdidier, M., Ioos, R., Husson, C., Cael, O., Scordia, T., and Marçais, B. (2018). Tracking the invasion: dispersal of hymenoscyphus fraxineus airborne inoculum at different scales. *FEMS microbiology ecology*, 94(5):fiy049.

[Grosdidier et al., 2020] Grosdidier, M., Scordia, T., Ioos, R., and Marçais, B. (2020). Landscape epidemiology of ash dieback. *Journal of Ecology*, 108(5):1789–1799.

[Gross et al., 2014a] Gross, A., Holdenrieder, O., Pautasso, M., Queloz, V., and Sieber, T. N. (2014a). H ymenoscyphus pseudoalbidus, the causal agent of e uropean ash dieback. *Molecular Plant Pathology*, 15(1):5–21.

[Gross et al., 2014b] Gross, A., Holdenrieder, O., Pautasso, M., Queloz, V., and Sieber, T. N. (2014b). Hymenoscyphus pseudoalbidus, the causal agent of european ash dieback. *Molecular Plant Pathology*, 15(1):5–21.

[Gross et al., 2012] Gross, A., Zaffarano, P. L., Duo, A., and Grünig, C. (2012). Reproductive mode and life cycle of the ash dieback pathogen hymenoscyphus pseudoalbidus. *Fungal genetics and biology*, 49(12):977–986.

[Grünwald et al., 2008] Grünwald, N. J., Goss, E. M., and Press, C. M. (2008). Phytophthora ramorum: a pathogen with a remarkably wide host range causing sudden oak death on oaks and ramorum blight on woody ornamentals. *Molecular Plant Pathology*, 9(6):729–740.

[Grünwald et al., 2012] Grünwald, N. J., Garbelotto, M., Goss, E. M., Heungens, K., and Prospero, S. (2012). Emergence of the sudden oak death pathogen phytophthora ramorum. *Trends in Microbiology*, 20(3):131–138.

[Gubbins et al., 2000] Gubbins, S., Gilligan, C. A., and Kleczkowski, A. (2000). Population dynamics of plant–parasite interactions: thresholds for invasion. *Theoretical Population Biology*, 57(3):219–233.

[Guillera-Arroita et al., 2015] Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., McCarthy, M. A., Tingley, R., and Wintle, B. A. (2015). Is my species distribution model fit for purpose? matching data and models to applications. *Global Ecology and Biogeography*, 24(3):276–292.

[Haack et al., 2010] Haack, R. A., Hérard, F., Sun, J., and Turgeon, J. J. (2010). Managing invasive populations of asian longhorned beetle and citrus longhorned beetle: a worldwide perspective. *Annual review of entomology*, 55.

[Hämäläinen et al., 2020] Hämäläinen, R. P., Miliszewska, I., and Voinov, A. (2020). Leadership in participatory modelling–is there a need for it? *Environmental Modelling & Software*, 133:104834.

[Haňáčková et al., 2017] Haňáčková, Z., Koukol, O., Čmoková, A., Zahradník, D., and Havrdová, L. (2017). Direct evidence of hymenoscyphus fraxineus infection pathway through the petiole-shoot junction. *Forest Pathology*, 47(6):e12370.

[Hanski, 1998] Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396(6706):41–49.

[Harwood et al., 2011] Harwood, T. D., Tomlinson, I., Potter, C. A., and Knight, J. D. (2011). Dutch elm disease revisited: past, present and future management in great britain. *Plant Pathology*, 60(3):545–555.

[Harwood et al., 2009] Harwood, T. D., Xu, X., Pautasso, M., Jeger, M. J., and Shaw, M. W. (2009). Epidemiological risk assessment using linked network and grid based modelling: Phytophthora ramorum and phytophthora kernoviae in the uk. *Ecological Modelling*, 220(23):3353–3361.

[Hauptman et al., 2015] Hauptman, T., Celar, F. A., De Groot, M., and Jurc, D. (2015). Application of fungicides and urea for control of ash dieback. *iForest-Biogeosciences and Forestry*, 8(2):165.

[Hauptman et al., 2013] Hauptman, T., Piškur, B., De Groot, M., Ogris, N., Ferlan, M., and Jurc, D. (2013). Temperature effect on c halara fraxinea: heat treatment of saplings as a possible disease control method. *Forest Pathology*, 43(5):360–370.

[Havrdová et al., 2016] Havrdová, L., Novotná, K., Zahradník, D., Buriánek, V., Pešková, V., Šrtka, P., and Černỳ, K. (2016). Differences in susceptibility to ash dieback in czech provenances of fraxinus excelsior. *Forest pathology*, 46(4):281–288.

[Havrdova et al., 2017] Havrdova, L., Zahradnik, D., Romportl, D., Pešková, V., Černỳ, K., et al. (2017). Environmental and silvicultural characteristics influencing the extent of ash dieback in forest stands. *Baltic Forestry*, 23(1):168–182.

[Hawker, 2016] Hawker, L. E. (2016). *The physiology of reproduction in fungi*. Cambridge University Press.

[He et al., 2017] He, L., Ren, X., Gao, Q., Zhao, X., Yao, B., and Chao, Y. (2017). The connected-component labeling problem: A review of state-of-the-art algorithms. *Pattern Recognition*, 70:25 – 43.

[He et al., 2019] He, Y., Chen, G., Potter, C., and Meentemeyer, R. K. (2019). Integrating multi-sensor remote sensing and species distribution modeling to map the spread of emerging forest disease and tree mortality. *Remote sensing of environment*, 231:111238.

[Hedelin et al., 2017] Hedelin, B., Evers, M., Alkan-Olsson, J., and Jonsson, A. (2017). Participatory modelling for sustainable development: Key issues derived from five cases of natural resource and disaster risk management. *Environmental science & policy*, 76:185–196.

[Heesterbeek and Zadoks, 1987] Heesterbeek, J. and Zadoks, J. (1987). Modelling pandemics of quarantine pests and diseases: problems and perspectives. *Crop protection*, 6(4):211–221.

[Heffernan et al., 2005] Heffernan, J. M., Smith, R. J., and Wahl, L. M. (2005). Perspectives on the basic reproductive ratio. *Journal of the Royal Society, Interface*, 2(4):281–293.

[Heiniger and Rigling, 1994] Heiniger, U. and Rigling, D. (1994). Biological control of chestnut blight in europe. *Annual review of phytopathology*, 32(1):581–599.

[Herring, 1979] Herring, J. R. (1979). Subgrid scale modeling — an introduction and overview. In Durst, F., Launder, B. E., Schmidt, F. W., and Whitelaw, J. H., editors, *Turbulent Shear Flows I*, pages 347–352, Berlin, Heidelberg. Springer Berlin Heidelberg.

[Hibbett et al., 2007] Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., Huhndorf, S., James, T., Kirk, P. M., Lücking, R., et al. (2007). A higher-level phylogenetic classification of the fungi. *Mycological research*, 111(5):509–547.

[Hietala et al., 2013] Hietala, A. M., Timmermann, V., BØrja, I., and Solheim, H. (2013). The invasive ash dieback pathogen hymenoscyphus pseudoalbidus exerts maximal infection pressure prior to the onset of host leaf senescence. *Fungal Ecology*, 6(4):302–308.

[Hill et al., 2017] Hill, L., Hector, A., Hemery, G., Smart, S., Tanadini, M., and Broan, N. (2017). Abundance distributions for tree species in great britain: A two-stage approach to modeling abundance using species distribution modeling and random forest. *Ecology and Evolution*, 7(4):1043–1056.

[Hill et al., 2019] Hill, L., Jones, G., Atkinson, N., Hector, A., Hemery, G., and Brown, N. (2019). The £15 billion cost of ash dieback in britain. *Current Biology*, 29(9):R315 – R316.

[Hillman and Suzuki, 2004] Hillman, B. I. and Suzuki, N. (2004). Viruses in the chestnut blight fungus. *Adv. Virus Res*, 63:423–473.

[Hollings et al., 2017] Hollings, T., Robinson, A., van Andel, M., Jewell, C., and Burgman, M. (2017). Species distribution models: A comparison of statistical approaches for livestock and disease epidemics. *PloS one*, 12(8):e0183626.

[Hopkins and Purcell, 2002] Hopkins, D. and Purcell, A. (2002). Xylella fastidiosa: cause of pierce's disease of grapevine and other emergent diseases. *Plant disease*, 86(10):1056–1066.

[Hoshen and Kopelman, 1976] Hoshen, J. and Kopelman, R. (1976). Percolation and cluster distribution. i. cluster multiple labeling technique and critical concentration algorithm. *Physical Review B*, 14(8):3438.

[Howe and Smallwood, 1982] Howe, H. F. and Smallwood, J. (1982). Ecology of seed dispersal. *Annual review of ecology and systematics*, 13(1):201–228.

[Hryniv and Mykytyuk, 2009] Hryniv, R. and Mykytyuk, Y. (2009). On zeros of some entire functions. *Transactions of the American Mathematical Society*, 361(4):2207–2223.

[Huang et al., 1980] Huang, H., Hoes, J., et al. (1980). Importance of plant spacing and sclerotial position to development of sclerotinia wilt of sunflower. *Plant Disease*, 64(1):81–84.

[Hui et al., 2009] Hui, C., McGeoch, M. A., Reyers, B., Roux, P. C., Greve, M., and Chown, S. L. (2009). Extrapolating population size from the occupancy–abundance relationship and the scaling pattern of occupancy. *Ecological Applications*, 19(8):2038–2048.

[Hultberg et al., 2020] Hultberg, T., Sandström, J., Felton, A., Öhman, K., Rönnberg, J., Witzell, J., and Cleary, M. (2020). Ash dieback risks an extinction cascade. *Biological Conservation*, 244:108516.

[Hunter et al., 2012] Hunter, W. B., Glick, E., Paldi, N., and Bextine, B. R. (2012). Advances in RNA interference: dsRNA Treatment in Trees and Grapevines for Insect Pest Suppression. *Southwestern Entomologist*, 37(1):85 – 87.

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

[James et al., 1990] James, N. D. G. et al. (1990). *A history of English forestry.* Number Ed. 2. Basil Blackwell.

[Janssen et al., 1988] Janssen, H. K., Schaub, B., and Schmittmann, B. (1988). The general epidemic process in a finite environment. *Journal of Physics A: Mathematical and Theoretical*, 21(7):l427.

[Jeger, 1984] Jeger, M. (1984). The use of mathematical models in plant disease epidemiology. *Scientific Horticulture*, 35:11–27.

[Jiménez-Valverde, 2012] Jiménez-Valverde, A. (2012). Insights into the area under the receiver operating characteristic curve (auc) as a discrimination measure in species distribution modelling. *Global Ecology and Biogeography*, 21(4):498–507.

[Jones and Kleczkowski, 2020] Jones, G. and Kleczkowski, A. (2020). Modelling plant health for policy. *Emerging Topics in Life Sciences*, 4(5):473–483.

[Kapitulnik et al., 1983] Kapitulnik, A., Aharony, A., Deutscher, G., and Stauffer, D. (1983). Self similarity and correlations in percolation. *Journal of Physics A: Mathematical and General*, 16(8):L269–L274.

[Kazempour, 2004] Kazempour, M. N. (2004). Biological control of rhizoctonia solani, the causal agent of rice sheath blight by antagonistics bacteria in greenhouse and field conditions. *Plant Pathology Journal*.

[Kearney and Porter, 2009] Kearney, M. and Porter, W. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology letters*, 12(4):334–350.

[Keeling and Eames, 2005] Keeling, M. J. and Eames, K. T. (2005). Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307.

[Kermack and McKendrick, 1927] Kermack, W. O. and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721.

[Keßler et al., 2012] Keßler, M., Cech, T. L., Brandstetter, M., and Kirisits, T. (2012). Dieback of ash (fraxinus excelsior and fraxinus angustifolia) in eastern austria: disease development on monitoring plots from 2007 to 2010. *Journal of Agricultural Extension and Rural Development*, 4(9):223–226.

[Khan et al., 2020] Khan, M. A., Ahmed, L., Mandal, P. K., Smith, R., and Haque, M. (2020). Modelling the dynamics of pine wilt disease with asymptomatic carriers and optimal control. *Scientific reports*, 10(1):1–15.

[King et al., 2015] King, K., Harris, A., and Webber, J. (2015). In planta detection used to define the distribution of the european lineages of phytophthora ramorum on larch (larix) in the uk. *Plant Pathology*, 64(5):1168–1175.

[Kjær et al., 2012] Kjær, E. D., McKinney, L. V., Nielsen, L. R., Hansen, L. N., and Hansen, J. K. (2012). Adaptive potential of ash (fraxinus excelsior) populations against the novel emerging pathogen hymenoscyphus pseudoalbidus. *Evolutionary applications*, 5(3):219–228.

[Klein et al., 2010] Klein, D. J., Hespanha, J., and Madhow, U. (2010). A reaction-diffusion model for epidemic routing in sparsely connected manets. In *2010 Proceedings IEEE INFOCOM*, pages 1–9. IEEE.

[Konadu et al., 2015] Konadu, D. D., Mourão, Z. S., Allwood, J. M., Richards, K. S., Kopec, G., McMahon, R., and Fenner, R. (2015). Land use implications of future energy system trajectories—the case of the uk 2050 carbon plan. *Energy Policy*, 86:328–337.

[Kowalski, 2001] Kowalski, T. (2001). O zamieraniu jesionów. *Trybuna Leśnika*, (04):6–7.

[Kowalski, 2006] Kowalski, T. (2006). Chalara fraxinea sp. nov. associated with dieback of ash (fraxinus excelsior) in poland. *Forest Pathology*, 36(4):264–270.

[Kowalski and Holdenrieder, 2009] Kowalski, T. and Holdenrieder, O. (2009). The teleomorph of chalara fraxinea, the causal agent of ash dieback. *Forest Pathology*, 39(5):304–308.

[Kraemer et al., 2016] Kraemer, M. U., Hay, S. I., Pigott, D. M., Smith, D. L., Wint, G. W., and Golding, N. (2016). Progress and challenges in infectious disease cartography. *Trends in Parasitology*, 32(1):19–29.

[Kramer and Botterweg, 1991] Kramer, K. J. and Botterweg, J. (1991). Aquatic biological early warning systems: an overview. *Bioindicators and environmental management*, pages 95–126.

[Kushalappa and Ludwig, 1982] Kushalappa, A. C. and Ludwig, A. (1982). Calculation of apparent infection rate in plant diseases: development. *Phytopathology*, 72:1373–1377.

[Langer et al., 2015] Langer, G., Harriehausen, U., Bressem, U., et al. (2015). Ash dieback. *AFZ/Der Wald, Allgemeine Forst Zeitschrift für Waldwirtschaft und*

Umweltvorsorge, 70(20):22–34.

[Leclerc et al., 2014] Leclerc, M., Doré, T., Gilligan, C. A., Lucas, P., and Filipe, J. A. (2014). Estimating the delay between host infection and disease (incubation period) and assessing its significance to the epidemiology of plant diseases. *PloS one*, 9(1):e86568.

[Li et al., 2011] Li, J., Blakeley, D., et al. (2011). The failure of r 0. *Computational and mathematical methods in medicine*, 2011.

[Liang et al., 1989] Liang, J., Piper, J., and Tang, J.-Y. (1989). Erosion and dilation of binary images by arbitrary structuring elements using interval coding. *Pattern Recognition Letters*, 9(3):201–209.

[Littmann, 2005] Littmann, F. (2005). Entire approximations to the truncated powers. *Constructive approximation*, 22(2):273–295.

[Liu et al., 2007] Liu, D., Kelly, M., Gong, P., and Guo, Q. (2007). Characterizing spatial–temporal tree mortality patterns associated with a new forest disease. *Forest Ecology and Management*, 253(1-3):220–231.

[Liu and Stechlinski, 2012] Liu, X. and Stechlinski, P. (2012). Infectious disease models with time-varying parameters and general nonlinear incidence rate. *Applied Mathematical Modelling*, 36(5):1974–1994.

[Lõhmus and Runnel, 2014] Lõhmus, A. and Runnel, K. (2014). Ash dieback can rapidly eradicate isolated epiphyte populations in production forests: a case study. *Biological Conservation*, 169:185–188.

[Lovett et al., 2016] Lovett, G. M., Weiss, M., Liebhold, A. M., Holmes, T. P., Leung, B., Lambert, K. F., Orwig, D. A., Campbell, F. T., Rosenthal, J., McCullough, D. G., et al. (2016). Nonnative forest insects and pathogens in the united states: Impacts and policy options. *Ecological Applications*, 26(5):1437–1455.

[Ludlam et al., 2012] Ludlam, J. J., Gibson, G. J., Otten, W., and Gilligan, C. A. (2012). Applications of percolation theory to fungal spread with synergy. *Journal of the Royal Society Interface*, 9(70):949–956.

[Luo et al., 2021] Luo, Y., Wu, J., Wang, X., Zhao, Y., and Feng, Z. (2021). Understanding ecological groups under landscape fragmentation based on network theory. *Landscape and Urban Planning*, 210:104066.

[MacKay and Jan, 1984] MacKay, G. and Jan, N. (1984). Forest fires as critical phenomena. *Journal of Physics A: Mathematical and General*, 17(14):L757–L760.

[Madden and Hughes, 1995] Madden, L. and Hughes, G. (1995). Plant disease incidence: distributions, heterogeneity, and temporal analysis. *Annual Review of Phytopathology*, 33(1):529–564.

[Madden et al., 2007] Madden, L. V., Hughes, G., and Van Den Bosch, F. (2007). The study of plant disease epidemics.

[Mamet et al., 2015] Mamet, S., Chun, K., Metsaranta, J., Barr, A., and Johnstone, J. (2015). Tree rings provide early warning signals of jack pine mortality across a moisture gradient in the southern boreal forest. *Environmental Research Letters*, 10(8):084021.

[Mansfield et al., 2018] Mansfield, J. W., Galambos, N., and Saville, R. (2018). The use of ascospores of the dieback fungus hymenoscyphus fraxineus for infection assays reveals a significant period of biotrophic interaction in penetrated ash cells. *Plant Pathology*, 67(6):1354–1361.

[Marçais et al., 2017] Marçais, B., Husson, C., Cael, O., Dowkiw, A., Saintonge, F.-X., Delahaye, L., Collet, C., and Chandelier, A. (2017). Estimation of ash mortality induced by hymenoscyphus fraxineus in france and belgium. *Baltic forestry*, 23(1):159–167.

[Marciulyniene et al., 2017] Marciulyniene, D., Davydenko, K., Stenlid, J., and Cleary, M. (2017). Can pruning help maintain vitality of ash trees affected by ash dieback in urban landscapes? *Urban Forestry & Urban Greening*, 27:69–75.

[McKinney et al., 2014] McKinney, L., Nielsen, L., Collinge, D., Thomsen, I., Hansen, J., and Kjær, E. (2014). The ash dieback crisis: genetic variation in resistance can prove a long-term solution. *Plant Pathology*, 63(3):485–499.

[McRoberts et al., 2019] McRoberts, N., Figuera, S. G., Olkowski, S., McGuire, B., Luo, W., Posny, D., and Gottwald, T. (2019). Using models to provide rapid programme support for california's efforts to suppress huanglongbing disease of citrus. *Philosophical Transactions of the Royal Society B*, 374(1776):20180281.

[Meentemeyer et al., 2004] Meentemeyer, R., Rizzo, D., Mark, W., and Lotz, E. (2004). Mapping the risk of establishment and spread of sudden oak death in california. *Forest Ecology and Management*, 200(1-3):195–214.

[Meentemeyer et al., 2011] Meentemeyer, R. K., Cunniffe, N. J., Cook, A. R., Filipe, J. A., Hunter, R. D., Rizzo, D. M., and Gilligan, C. A. (2011). Epidemiological modeling of invasion in heterogeneous landscapes: spread of sudden oak death in california (1990–2030). *Ecosphere*, 2(2):1–24.

[Messina et al., 2016] Messina, J. P., Kraemer, M. U., Brady, O. J., Pigott, D. M., Shearer, F. M., Weiss, D. J., Golding, N., Ruktanonchai, C. W., Gething, P. W., Cohn, E., et al. (2016). Mapping global environmental suitability for zika virus. *elife*, 5:e15272.

[Metcalf and Luckmann, 1994] Metcalf, R. L. and Luckmann, W. H. (1994). *Introduction to insect pest management*, volume 101. John Wiley & Sons.

[Mikaberidze et al., 2016] Mikaberidze, A., Mundt, C. C., and Bonhoeffer, S. (2016). Invasiveness of plant pathogens depends on the spatial scale of host distribution. *Ecological applications*, 26(4):1238–1248.

[Milne et al., 2020] Milne, A. E., Gottwald, T., Parnell, S. R., Alonso Chavez, V., and Van den Bosch, F. (2020). What makes or breaks a campaign to stop an invading plant pathogen? *PLoS computational biology*, 16(2):e1007570.

[Mitchell et al., 2014] Mitchell, R., Beaton, J., Bellamy, P., Broome, A., Chetcuti, J., Eaton, S., Ellis, C., Gimona, A., Harmer, R., Hester, A., et al. (2014). Ash dieback in the uk: a review of the ecological and conservation implications and potential management options. *Biological conservation*, 175:95–109.

[Mollison, 1977] Mollison, D. (1977). Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(3):283–313.

[Moré, 1978] Moré, J. J. (1978). The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer.

[Mundt et al., 2009a] Mundt, C. C., Sackett, K. E., Wallace, L. D., Cowger, C., and Dudley, J. P. (2009a). Aerial dispersal and multiple-scale spread of epidemic disease. *EcoHealth*, 6(4):546–552.

[Mundt et al., 2009b] Mundt, C. C., Sackett, K. E., Wallace, L. D., Cowger, C., and Dudley, J. P. (2009b). Long-distance dispersal and accelerating waves of disease: empirical relationships. *The American Naturalist*, 173(4):456–466.

[Murray, 2002] Murray, J. D. (2002). *MathematicalBiology I. An Introduction.* Springer.

[Muñoz et al., 2016] Muñoz, F., Marçais, B., Dufour, J., and Dowkiw, A. (2016). Rising out of the ashes: Additive genetic variation for crown and collar resistance to hymenoscyphus fraxineus in fraxinus excelsior. *Phytopathology®*, 106(12):1535–1543. PMID: 27349738.

[Nachtegael and Kerre, 2001] Nachtegael, M. and Kerre, E. E. (2001). Connections between binary, gray-scale and fuzzy mathematical morphologies. *Fuzzy sets and systems*, 124(1):73–85.

[Nathan et al., 2012] Nathan, R., Klein, E., Robledo-Arnuncio, J. J., and Revilla, E. (2012). Dispersal kernels. *Dispersal ecology and evolution*, pages 187–210.

[Neher and Campbell, 1992] Neher, D. A. and Campbell, C. L. (1992). Underestimation of disease progress rates with the logistic, monomolecular, and gompertz models when maximum disease intensity is less than 100 percent. *Phytopathology*, 82(8):811–814.

[Nemesio-Gorriz et al., 2020a] Nemesio-Gorriz, M., Menezes, R. C., Paetz, C., Hammerbacher, A., Steenackers, M., Schamp, K., Höfte, M., Svatoš, A., Gershenzon, J., and Douglas, G. C. (2020a). Canditate metabolites for ash dieback tolerance in fraxinus excelsior. *Journal of Experimental Botany*, 71(19):6074–6083.

[Nemesio-Gorriz et al., 2020b] Nemesio-Gorriz, M., Menezes, R. C., Paetz, C., Hammerbacher, A., Steenackers, M., Schamp, K., Höfte, M., Svatoš, A., Gershenzon, J., and Douglas, G. C. (2020b). Metabolomics in fraxinus excelsior; identification and validation of biochemical markers for tolerance to ash dieback. *Journal of Experimental Botany*.

[Neri et al., 2014] Neri, F. M., Cook, A. R., Gibson, G. J., Gottwald, T. R., and Gilligan, C. A. (2014). Bayesian analysis for inference of an emerging epidemic: citrus canker in urban landscapes. *PLoS Computational Biology*, 10(4):e1003587.

[Nickmans et al., 2019] Nickmans, H., Jonard, M., Verheyen, K., and Ponette, Q. (2019). Modelling leaf dispersal and nutrient return in tree species mixtures. *Forest Ecology and Management*, 436:68–78.

[Ódor, 2004] Ódor, G. (2004). Universality classes in nonequilibrium lattice systems. *Rev. Mod. Phys.*, 76:663–724.

[Oerke, 2006] Oerke, E. (2006). Crop losses to pests. *The Journal of Agricultural Science*, 144:31.

[Ohtsuki and Keyes, 1986] Ohtsuki, T. and Keyes, T. (1986). Kinetic growth percolation: Epidemic processes with immunization. *Physical Review A*, 33(2):1223–1232.

[Orozco-Fuentes et al., 2019] Orozco-Fuentes, S., Griffiths, G., Holmes, M., Ettelaie, R., Smith, J., Baggaley, A., and Parker, N. (2019). Early warning signals in plant disease outbreaks. *Ecological Modelling*, 393:12 – 19.

[Otieno et al., 2021] Otieno, F. T., Gachohi, J., Gikuma-Njuru, P., Kariuki, P., Oyas, H., Canfield, S. A., Blackburn, J. K., Njenga, M. K., and Bett, B. (2021). Modeling the spatial distribution of anthrax in southern kenya. *PLoS neglected tropical diseases*, 15(3):e0009301.

[Otten et al., 2004] Otten, W., Bailey, D. J., and Gilligan, C. A. (2004). Empirical evidence of spatial thresholds to control invasion of fungal parasites and saprotrophs. *New Phytologist*, 163(1):125–132.

[O'Neill et al., 1992] O'Neill, R. V., Gardner, R. H., Turner, M. G., and Romme, W. H. (1992). Epidemiology theory and disturbance spread on landscapes. *Landscape Ecology*, 7(1):19–26.

[Pal and Gardener, 2006] Pal, K. K. and Gardener, B. M. (2006). Biological control of plant pathogens.

[Palmieri and Frankel, 2006] Palmieri, K. and Frankel, S. J. (2006). California oak mortality task force. 2006 sudden oak death & phytophthora ramorum summary report. a compendium of comtf monthly newsletters.

[Papavizas, 1970] Papavizas, G. C. (1970). Colonization and growth of rhizoctonia solani in soil. *Rhizoctonia solani, biology and pathology*, pages 108–122.

[Papaïx et al., 2014] Papaïx, J., Touzeau, S., Monod, H., and Lannou, C. (2014). Can epidemic control be achieved by altering landscape connectivity in agricultural systems? *Ecological Modelling*, 284:35 – 47.

[Park et al., 2001] Park, A. W., Gubbins, S., and Gilligan, C. A. (2001). Invasion and persistence of plant parasites in a spatially structured host population. *Oikos*, 94(1):162–174.

[Park et al., 2002] Park, A. W., Gubbins, S., and Gilligan, C. A. (2002). Extinction times for closed epidemics: the effects of host spatial structure. *Ecology Letters*, 5(6):747–755.

[Parnell et al., 2012] Parnell, S., Gottwald, T., Gilks, W., and Van den Bosch, F. (2012). Estimating the incidence of an epidemic when it is first discovered and the design of early detection monitoring. *Journal of theoretical biology*, 305:30–36.

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

[Parnell et al., 2009a] Parnell, S., Gottwald, T., Van Den Bosch, F., and Gilligan, C. (2009a). Optimal strategies for the eradication of asiatic citrus canker in heterogeneous host landscapes. *Phytopathology*, 99(12):1370–1376.

[Parnell et al., 2009b] Parnell, S., Gottwald, T. R., van den Bosch, F., and Gilligan, C. A. (2009b). Optimal strategies for the eradication of asiatic citrus canker in heterogeneous host landscapes. *Phytopathology®*, 99(12):1370–1376. PMID: 19900003.

[Parnell et al., 2017] Parnell, S., van den Bosch, F., Gottwald, T., and Gilligan, C. A. (2017). Surveillance to inform control of emerging plant diseases: An epidemiological perspective. *Annual Review of Phytopathology*, 55(1):591–610. PMID: 28637378.

[Pautasso et al., 2013] Pautasso, M., Aas, G., Queloz, V., and Holdenrieder, O. (2013). European ash (fraxinus excelsior) dieback–a conservation biology challenge. *Biological conservation*, 158:37–49.

[Pecchi et al., 2019] Pecchi, M., Marchi, M., Burton, V., Giannetti, F., Moriondo, M., Bernetti, I., Bindi, M., and Chirici, G. (2019). Species distribution modelling to support forest management. a literature review. *Ecological Modelling*, 411:108817.

[Pérez-Reche et al., 2011] Pérez-Reche, F. J., Ludlam, J. J., Taraskin, S. N., and Gilligan, C. A. (2011). Synergy in spreading processes: From exploitative to explorative foraging strategies. *Phys. Rev. Lett.*, 106:218701.

[Perrings, 2016] Perrings, C. (2016). Options for managing the infectious animal and plant disease risks of international trade. *Food Security*, 8(1):27–35.

[Picard et al., 2019] Picard, C., Soubeyrand, S., Jacquot, E., and Thébaud, G. (2019). Analyzing the influence of landscape aggregation on disease spread to improve management strategies. *Phytopathology*, 109(7):1198–1207.

[Pietzsch et al., 2021] Pietzsch, B. W., Peter, F. J., and Berger, U. (2021). The effect of sanitation felling on the spread of the european spruce bark beetle-an individual-based modeling approach. *Frontiers in Forests and Global Change*, 4:103.

[Plotkin et al., 2002] Plotkin, J. B., Chave, J., and Ashton, P. S. (2002). Cluster analysis of spatial patterns in malaysian tree species.. *The American Naturalist*, 160(5):629–644. PMID: 18707513.

[Plumb et al., 2020] Plumb, W. J., Coker, T. L. R., Stocks, J. J., Woodcock, P., Quine, C. P., Nemesio-Gorriz, M., Douglas, G. C., Kelly, L. J., and Buggs, R. J. A. (2020). The viability of a breeding programme for ash in the british isles in the face of ash dieback. *PLANTS, PEOPLE, PLANET*, 2(1):29–40.

[Poggi et al., 2013] Poggi, S., Neri, F., Deytieux, V., Bates, A., Otten, W., Gilligan, C. A., and Bailey, D. J. (2013). Percolation-based risk index for pathogen invasion: application to soilborne disease in propagation systems. *Phytopathology*, 103(10):1012–1019.

[Potter et al., 2011] Potter, C., Harwood, T., Knight, J., and Tomlinson, I. (2011). Learning from history, predicting the future: the uk dutch elm disease outbreak in relation to contemporary tree disease threats. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1573):1966–1974.

[Potter and Urquhart, 2017] Potter, C. and Urquhart, J. (2017). Tree disease and pest epidemics in the anthropocene: A review of the drivers, impacts and policy responses in the uk. *Forest Policy and Economics*, 79:61 – 68. Forest governance in the Anthropocene: a challenge for theory and practice.

[Pybus et al., 2012] Pybus, O. G., Suchard, M. A., Lemey, P., Bernardin, F. J., Rambaut, A., Crawford, F. W., Gray, R. R., Arinaminpathy, N., Stramer, S. L., Busch, M. P., et al. (2012). Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the national academy of sciences*, 109(37):15066–15071.

[Queloz et al., 2011] Queloz, V., Grünig, C. R., Berndt, R., Kowalski, T., Sieber, T. N., and Holdenrieder, O. (2011). Cryptic speciation in hymenoscyphus albidus. *Forest Pathology*, 41(2):133–142.

[Ravera et al., 2020] Ravera, F., Tarrasón, D., Hubacek, K., Molowny-Horas, R., and Sendzimir, J. (2020). Participatory modelling in adaptive environmental management: a case study in semi-arid northern nicaragua. In *Environmental Assessments*. Edward Elgar Publishing.

[Ray et al., 2021] Ray, D., Marchi, M., Rattey, A., and Broome, A. (2021). A multi-data ensemble approach for predicting woodland type distribution: Oak woodland in britain. *Ecology and Evolution*.

[Redak et al., 2004] Redak, R. A., Purcell, A. H., Lopes, J. R., Blua, M. J., Mizell Iii, R. F., and Andersen, P. C. (2004). The biology of xylem fluid–feeding insect vectors of xylella fastidiosa and their relation to disease epidemiology. *Annual Reviews in Entomology*, 49(1):243–270.

[Reed et al., 2018] Reed, M. S., Vella, S., Challies, E., De Vente, J., Frewer, L., Hohenwallner-Ries, D., Huber, T., Neumann, R. K., Oughton, E. A., Sidoli del Ceno, J., et al. (2018). A theory of participation: what makes stakeholder and public engagement in environmental management work? *Restoration ecology*, 26:S7–S17.

[Refaeilzadeh et al., 2009] Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, 5:532–538.

[Reimer et al., 2017] Reimer, J. R., Bonsall, M. B., and Maini, P. K. (2017). The critical domain size of stochastic population models. *Journal of mathematical biology*, 74(3):755–782.

[Rieux et al., 2014] Rieux, A., Soubeyrand, S., Bonnot, F., Klein, E. K., Ngando, J. E., Mehl, A., Ravigne, V., Carlier, J., and de Lapeyre de Bellaire, L. (2014). Long-distance wind-dispersal of spores in a fungal plant pathogen: Estimation of anisotropic dispersal kernels from an extensive field experiment. *PLOS ONE*, 9(8):1–13.

[Rigling and Prospero, 2018] Rigling, D. and Prospero, S. (2018). Cryphonectria parasitica, the causal agent of chestnut blight: invasion history, population biology and disease control. *Molecular Plant Pathology*, 19(1):7–20.

[Rizzo et al., 2002] Rizzo, D., Garbelotto, M., Davidson, J., Slaughter, G., and Koike, S. (2002). Phytophthora ramorum as the cause of extensive mortality of quercus spp. and lithocarpus densiflorus in california. *Plant disease*, 86(3):205–214.

[Rodoni, 2009] Rodoni, B. (2009). The role of plant biosecurity in preventing and controlling emerging plant virus disease epidemics. *Virus research*, 141(2):150–157.

[Rogers et al., 2018] Rogers, B. M., Solvik, K., Hogg, E. H., Ju, J., Masek, J. G., Michaelian, M., Berner, L. T., and Goetz, S. J. (2018). Detecting early warning signals of tree mortality in boreal north america using multiscale satellite data. *Global change biology*, 24(6):2284–2304.

[Roy et al., 2014] Roy, B. A., Alexander, H. M., Davidson, J., Campbell, F. T., Burdon, J. J., Sniezko, R., and Brasier, C. (2014). Increasing forest loss worldwide from invasive pests requires new trade regulations. *Frontiers in Ecology and the Environment*, 12(8):457–465.

[Ryle et al., 1969] Ryle, G. et al. (1969). Forest service; the first forty-five years of the forestry commission of great britain. *Forest Service; the first forty-five years of the Forestry Commission of Great Britain.*

[Sall, 1980] Sall, M. A. (1980). Epidemiology of grape powdery mildew: a model. *Phytopathology*, 70(4):338–342.

[Samet and Tamminen, 1988] Samet, H. and Tamminen, M. (1988). Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):579–586.

[Sangare and Adler, 2009] Sangare, D. and Adler, P. (2009). Continuum percolation of isotropically oriented circular cylinders. *Physical Review E*, 79(5):052101.

[Scheffer et al., 2009] Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., Van Nes, E. H., Rietkerk, M., and Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, 461(7260):53–59.

[Scherm, 1996] Scherm, H. (1996). On the velocity of epidemic waves in model plant disease epidemics. *Ecological Modelling*, 87(1-3):217–222.

[Schubert et al., 2001] Schubert, T. S., Rizvi, S. A., Sun, X., Gottwald, T. R., Graham, J. H., and Dixon, W. N. (2001). Meeting the challenge of eradicating citrus canker in florida—again. *Plant disease*, 85(4):340–356.

[Schumacher, 2011] Schumacher, J. (2011). The general situation regarding ash dieback in germany and investigations concerning the invasion and distribution strategies of chalara fraxinea in woody tissue 1. *EPPO Bulletin*, 41(1):7–10.

[Sconyers et al., 2005] Sconyers, L. E., Brenneman, T. B., Stevenson, K. L., and Mullinix, B. G. (2005). Effects of plant spacing, inoculation date, and peanut cultivar on epidemics of peanut stem rot and tomato spotted wilt. *Plant Disease*, 89(9):969–974. PMID: 30786630.

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

[Sellke, 1983] Sellke, T. (1983). On the asymptotic distribution of the size of a stochastic epidemic. *Journal of Applied Probability*, 20(2):390–394.

[Severns et al., 2019] Severns, P. M., Sackett, K. E., Farber, D. H., and Mundt, C. C. (2019). Consequences of long-distance dispersal for epidemic spread: Patterns, scaling, and mitigation. *Plant disease*, 103(2):177–191.

[Shabani et al., 2016] Shabani, F., Kumar, L., and Ahmadi, M. (2016). A comparison of absolute performance of different correlative and mechanistic species distribution models in an independent area. *Ecology and evolution*, 6(16):5973–5986.

[Shaw et al., 2006] Shaw, M., Harwood, T., Wilkinson, M., and Elliott, L. (2006). Assembling spatially explicit landscape models of pollen and spore dispersal by wind for risk assessment. *Proceedings of the Royal Society B: Biological Sciences*, 273(1594):1705–1713.

[Shih and Mitchell, 1989] Shih, F. and Mitchell, O. (1989). Threshold decomposition of gray-scale morphology into binary morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(1):31–42.

[Shure and Gottschalk, 1985] Shure, D. J. and Gottschalk, M. R. (1985). Litter-fall patterns within a floodplain forest. *American Midland Naturalist*, pages 98–111.

[Sidda et al., 2020] Sidda, J. D., Song, L., Parker, J. L., Studholme, D. J., Sambles, C., and Grant, M. (2020). Diversity of secoiridoid glycosides in leaves of uk and danish ash provide new insight for ash dieback management. *Scientific reports*, 10(1):1–12.

[Sixsmith, 2011] Sixsmith, D. (2011). Entire functions for which the escaping set is a spider's web. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 151, pages 551–571. Cambridge University Press.

[Skellam, 1951] Skellam, J. G. (1951). Random dispersal in theoretical populations. *Biometrika*, 38(1/2):196–218.

[Skov et al., 2016] Skov, H., Heinänen, S., Thaxter, C. B., Williams, A. E., Lohier, S., and Banks, A. N. (2016). Real-time species distribution models for conservation and management of natural resources in marine environments. *Marine Ecology Progress Series*, 542:221–234.

[Skovsgaard et al., 2017] Skovsgaard, J. P., Wilhelm, G. J., Thomsen, I. M., Metzler, B., Kirisits, T., Havrdová, L., Enderle, R., Dobrowolska, D., Cleary, M., and Clark, J. (2017). Silvicultural strategies for fraxinus excelsior in response to dieback caused by hymenoscyphus fraxineus. *Forestry: An International Journal of Forest Research*, 90(4):455–472.

[Smart, 2020] Smart, S.M.;Andrews, C. E. R. A. P. E. D. D. R. R. P. J. M. B. C. (2020). Vegetation plot data from the ukceh countryside survey, great britain, 2019.

[Smith et al., 2004] Smith, M. T., Tobin, P. C., Bancroft, J., Li, G., and Gao, R. (2004). Dispersal and spatiotemporal dynamics of asian longhorned beetle (coleoptera: Cerambycidae) in china. *Environmental Entomology*, 33(2):435–442.

[Solheim and Hietala, 2017] Solheim, H. and Hietala, A. M. (2017). Spread of ash dieback in norway.

[Soubeyrand et al., 2009] Soubeyrand, S., Laine, A.-L., Hanski, I., and Penttinen, A. (2009). Spatiotemporal structure of host-pathogen interactions in a metapopulation. *The American Naturalist*, 174(3):308–320.

[Srivastava et al., 2019] Srivastava, V., Lafond, V., Griess, V. C., et al. (2019). Species distribution models (sdm): applications, benefits and challenges in invasive species management. *CAB Rev*, 14(10.1079).

[Staelens et al., 2003] Staelens, J., Nachtergale, L., Luyssaert, S., and Lust, N. (2003). A model of wind-influenced leaf litterfall in a mixed hardwood forest. *Canadian Journal of Forest Research*, 33(2):201–209.

[Stauffer, 1979] Stauffer, D. (1979). Scaling theory of percolation clusters. *Physics Reports*, 54(1):1 – 74.

[Stauffer and Aharony, 2018] Stauffer, D. and Aharony, A. (2018). *Introduction to percolation theory.* CRC press.

[Stener, 2013] Stener, L.-G. (2013). Clonal differences in susceptibility to the dieback of fraxinus excelsior in southern sweden. *Scandinavian Journal of Forest Research*, 28(3):205–216.

[Stocks et al., 2017] Stocks, J. J., Buggs, R. J., and Lee, S. J. (2017). A first assessment of fraxinus excelsior (common ash) susceptibility to hymenoscyphus fraxineus (ash dieback) throughout the british isles. *Scientific reports*, 7(1):1–7.

[Suffert and Thompson, 2018] Suffert, F. and Thompson, R. (2018). Some reasons why the latent period should not always be considered constant over the course of a plant disease epidemic. *Plant Pathology*, 67(9):1831–1840.

[Suzuki et al., 2003] Suzuki, R. O., Kudoh, H., and Kachi, N. (2003). Spatial and temporal variations in mortality of the biennial plant, lysimachia rubida: effects of intraspecific competition and environmental heterogeneity. *Journal of Ecology*, 91(1):114–125.

[Svensson, 2007] Svensson, Å. (2007). A note on generation times in epidemic models. *Mathematical biosciences*, 208(1):300–311.

[Swinton and Gilligan, 1996a] Swinton, J. and Gilligan, C. A. (1996a). Dutch elm disease and the future of the elm in the uk: a quantitative analysis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1340):605–615.

[Swinton and Gilligan, 1996b] Swinton, J. and Gilligan, C. A. (1996b). Dutch elm disease and the future of the elm in the u.k.: a quantitative analysis. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1340):605–615.

[Tang and Lin, 2011] Tang, Q. and Lin, Z. (2011). The asymptotic analysis of an insect dispersal model on a growing domain. *Journal of mathematical analysis and applications*, 378(2):649–656.

[Tankam-Chedjou et al., 2020] Tankam-Chedjou, I., Touzeau, S., Mailleret, L., Tewa, J. J., and Grognard, F. (2020). Modelling and control of a banana soilborne pest in a multi-seasonal framework. *Mathematical biosciences*, 322:108324.

[Taylor et al., 1979] Taylor, PHELPS, K., and DUDLEY, C. L. (1979). Epidemiology and strategy for the control of halo-blight of beans. *Annals of Applied Biology*, 93(2):167–172.

[Teng, 1985] Teng, P. S. (1985). A comparison of simulation approaches to epidemic modeling. *Annual Review of Phytopathology*, 23(1):351–379.

[Thomas, 2016] Thomas, P. A. (2016). Biological flora of the british isles: Fraxinus excelsior. *Journal of Ecology*, 104(4):1158–1209.

[Thompson et al., 2019] Thompson, R., Stockwin, J., van Gaalen, R. D., Polonsky, J., Kamvar, Z., Demarsh, P., Dahlqwist, E., Li, S., Miguel, E., Jombart, T., et al. (2019). Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics*, 29:100356.

[Thrall and Burdon, 2003] Thrall, P. H. and Burdon, J. J. (2003). Evolution of virulence in a plant host-pathogen metapopulation. *Science*, 299(5613):1735–1737.

[Thuiller et al., 2016] Thuiller, W., Georges, D., Engler, R., Breiner, F., Georges, M. D., and Thuiller, C. W. (2016). Package 'biomod2'. *Species distribution modeling within an ensemble forecasting framework*.

[Tildesley and Keeling, 2009] Tildesley, M. J. and Keeling, M. J. (2009). Is r(0) a good predictor of final epidemic size: foot-and-mouth disease in the uk. *Journal of theoretical biology*, 258(4):623–629.

[Timmermann et al., 2011] Timmermann, V., Børja, I., Hietala, A. M., Kirisits, T., and Solheim, H. (2011). Ash dieback: pathogen spread and diurnal patterns of ascospore dispersal, with special emphasis on norway*. *EPPO Bulletin*, 41(1):14–20.

[Toffoli and Margolus, 1987] Toffoli, T. and Margolus, N. (1987). *Cellular automata machines: a new environment for modeling*. MIT press.

[Tol, 2021] Tol, R. S. (2021). Europe's climate target for 2050: An assessment. *Intereconomics*, 56(6):330–335.

[Tomlinson, 2016] Tomlinson, I. (2016). The discovery of ash dieback in the uk: The making of a focusing event. *Environmental Politics*, 25(4):709–728.

[Tomlinson and Potter, 2010] Tomlinson, I. and Potter, C. (2010). 'too little, too late'? science, policy and dutch elm disease in the uk. *Journal of Historical Geography*, 36(2):121 – 131.

[Tomlinson et al., 2015] Tomlinson, I., Potter, C., and Bayliss, H. (2015). Managing tree pests and diseases in urban settings: the case of oak processionary moth in london, 2006–2012. *Urban Forestry & Urban Greening*, 14(2):286–292.

[Tonini et al., 2017] Tonini, F., Shoemaker, D., Petrasova, A., Harmon, B., Petras, V., Cobb, R. C., Mitasova, H., and Meentemeyer, R. K. (2017). Tangible geospatial modeling for collaborative solutions to invasive species management. *Environmental modelling & software*, 92:176–188.

[Tooley et al., 2004] Tooley, P., Kyde, K., and Englander, L. (2004). Susceptibility of selected ericaceous ornamental host species to phytophthora ramorum. *Plant Disease*, 88(9):993–999.

[Torres et al., 2021] Torres, P., Rodes-Blanco, M., Viana-Soto, A., Nieto, H., and García, M. (2021). The role of remote sensing for the assessment and monitoring of forest health: A systematic evidence synthesis. *Forests*, 12(8):1134.

[Tsror, 2010] Tsror, L. (2010). Biology, epidemiology and management of rhizoctonia solani on potato. *Journal of Phytopathology*, 158(10):649–658.

[Tyrväinen et al., 2005] Tyrväinen, L., Pauleit, S., Seeland, K., and de Vries, S. (2005). Benefits and uses of urban forests and trees. In *Urban forests and trees*, pages 81–114. Springer.

[Valachovic et al., 2008] Valachovic, Y., Lee, C., Marshall, J., and Scanlon, H. (2008). Wildland management of phytophthora ramorum in northern california forests. In *In: Frankel, Susan J.; Kliejunas, John T.; Palmieri, Katharine M., tech. coords. 2008. Proceedings of the sudden oak death third science symposium. Gen. Tech. Rep. PSW-GTR-214. Albany, CA: US Department of Agriculture, Forest Service, Pacific Southwest Research Station. pp. 305-312*, volume 214.

[van den Berg et al., 2011] van den Berg, F., Bacaer, N., Metz, J. A., Lannou, C., and van den Bosch, F. (2011). Periodic host absence can select for higher or lower parasite transmission rates. *Evolutionary Ecology*, 25(1):121–137.

[Van den Bosch et al., 1988] Van den Bosch, F., Frinking, H., Metz, J., and Zadoks, J. (1988). Focus expansion in plant disease. iii: Two experimental examples. *Phytopathology*, 78(7):919–925.

[van den Bosch and Gilligan, 2008] van den Bosch, F. and Gilligan, C. A. (2008). Models of fungicide resistance dynamics. *Annual Review of Phytopathology*, 46(1):123–147. PMID: 18680425.

[Van der Plank, 1963] Van der Plank, J. E. (1963). *Plant diseases: epidemics and control*. Elsevier.

[Van Leeuwen et al., 1959] Van Leeuwen, J., Groeneveld, J., and De Boer, J. (1959). New method for the calculation of the pair correlation function. i. *Physica*, 25(7-12):792–808.

[Viljanen-Rollinson et al., 2007] Viljanen-Rollinson, S., Parr, E., and Marroni, M. (2007). Monitoring longdistance spore dispersal by wind a review. *New Zealand Plant Protection*, 60:291–296.

[Virtanen et al., 2020] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

[Wadkin et al., 2022] Wadkin, L. E., Branson, J., Hoppit, A., Parker, N. G., Golightly, A., and Baggaley, A. W. (2022). Inference for epidemic models with time-varying infection rates: Tracking the dynamics of oak processionary moth in the uk. *Ecology and evolution*, 12(5):e8871.

[Waggoner et al., 1969] Waggoner, P. E., Horsfall, J. G., et al. (1969). Epidem: a simulator of plant disease written for a computer. *Bulletin. Connecticut Agricultural Ex-*

periment Station, 698.

[Walker, 1963] Walker, J. C. (1963). The future of plant pathology. *Annual Review of Phytopathology*, 1(1):1–4.

[Weiss, 2013] Weiss, H. H. (2013). The sir model and the foundations of public health. *Materials matematics*, pages 0001–17.

[White et al., 2017] White, S. M., Bullock, J. M., Hooftman, D. A., and Chapman, D. S. (2017). Modelling the spread and control of xylella fastidiosa in the early stages of invasion in apulia, italy. *Biological Invasions*, 19(6):1825–1837.

[Wiegand et al., 2007] Wiegand, T., Gunatilleke, S., Gunatilleke, N., and Okuda, T. (2007). Analyzing the spatial structure of a sri lankan tree species with multiple scales of clustering. *Ecology*, 88(12):3088–3102.

[Wingen et al., 2013] Wingen, L., Shaw, M., and Brown, J. (2013). Long-distance dispersal and its influence on adaptation to host resistance in a heterogeneous landscape. *Plant Pathology*, 62(1):9–20.

[Wingfield et al., 2010] Wingfield, J., M., Slippers, Bernard, Jolanda, R., Wingfield, and D., B. (2010). *Fifty Years of Tree Pest and Pathogen Invasions, Increasingly Threatening World Forests*, chapter 8, pages 89–99. John Wiley Sons, Ltd.

[Wingfield et al., 2012] Wingfield, M. J., De Beer, Z. W., Slippers, B., Wingfield, B. D., Groenewald, J. Z., Lombard, L., and Crous, P. W. (2012). One fungus, one name promotes progressive plant pathology. *Molecular plant pathology*, 13(6):604–613.

[Wittmann et al., 2016] Wittmann, M. E., Barnes, M. A., Jerde, C. L., Jones, L. A., and Lodge, D. M. (2016). Confronting species distribution model predictions with species functional traits. *Ecology and Evolution*, 6(4):873–879.

[Wood et al., 2017] Wood, C. M., Smart, S. M., Bunce, R. G., Norton, L. R., Maskell, L. C., Howard, D. C., Scott, W. A., and Henrys, P. A. (2017). Long-term vegetation monitoring in great britain–the countryside survey 1978–2007 and beyond. *Earth System Science Data*, 9(2):445–459.

[Worthen et al., 2010] Worthen, L. M., Woeste, K. E., and Michler, C. H. (2010). *Breeding American Chestnuts for Blight Resistance*, chapter 7, pages 305–339. John Wiley Sons, Ltd.

[Wulfert et al., 2010] Wulfert, I., Gundermann, G., Er, H., et al. (2010). The implementation of the eu plant passport system in turkey. *Julius-Kuhn-Archiv*, (428).

[Wylder et al., 2018] Wylder, B., Biddle, M., King, K., Baden, R., and Webber, J. (2018). Evidence from mortality dating of fraxinus excelsior indicates ash dieback (hymenoscyphus fraxineus) was active in england in 2004–2005. *Forestry: An International Journal of Forest Research*, 91(4):434–443.

[Yamamura et al., 2016] Yamamura, K., Katsumata, H., Yoshioka, J., Yuda, T., and Kasugai, K. (2016). Sampling inspection to prevent the invasion of alien pests: statistical theory of import plant quarantine systems in japan. *Population Ecology*, 58(1):63–80.

[Yang et al., 2005] Yang, X., Zhang, K., Jia, B., and Ci, L. (2005). Desertification assessment in china: An overview. *Journal of Arid Environments*, 63(2):517–531.

[Yano, 2017] Yano, R. (2017). Kinetic modeling of local epidemic spread and its simulation. *Journal of scientific computing*, 73(1):122–156.

[Zadoks and Kampmeijer, 1977] Zadoks, J. and Kampmeijer, P. (1977). The role of crop populations and their deployment, illustrated by means of a simulator, epimul76. *Annals of the New York Academy of Sciences*, 287(1):164–190.

[Zadoks et al., 1979] Zadoks, J. C., Schein, R. D., et al. (1979). Epidemiology and plant disease management. *Epidemiology and plant disease management.*

[Zhang, 2017] Zhang, G.-B. (2017). Non-monotone traveling waves and entire solutions for a delayed nonlocal dispersal equation. *Applicable Analysis*, 96(11):1830–1866.

[Zhang et al., 2014] Zhang, H., Yuan, W., Dong, W., and Liu, S. (2014). Seasonal patterns of litterfall in forest ecosystem worldwide. *Ecological Complexity*, 20:240–247.

[Zhang et al., 2019] Zhang, Z., Xu, S., Capinha, C., Weterings, R., and Gao, T. (2019). Using species distribution model to predict the impact of climate change on the potential distribution of japanese whiting sillago japonica. *Ecological Indicators*, 104:333–340.

[Zhao et al., 2013] Zhao, Y.-J., Hosoya, T., Baral, H.-O., Hosaka, K., and Kakishima, M. (2013). Hymenoscyphus pseudoalbidus, the correct name for lambertella albida reported from japan. *Mycotaxon*, 122(1):25–41.

[Zhu et al., 2000] Zhu, Y., Chen, H., Fan, J., Wang, Y., Li, Y., Chen, J., Fan, J., Yang, S., Hu, L., Leung, H., Mew, T. W., Teng, P. S., Wang, Z., and Mundt, C. C. (2000). Genetic diversity and disease control in rice. *Nature*, 406(6797):718–722.