Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Exploring Possible Causes of a Dilution Effect Using Agent-Based Models

Jason Archer

December 2024

Abstract

The relationship between infectious diseases and ecosystems has become a much-discussed topic, with many emerging infectious diseases originating in animal reservoirs. A controversial subject within ecological epidemiology is the dilution effect. It supposes a negative relationship exists between a given ecosystem's biodiversity and the risk of infectious disease transmission. An appealing prospect in theory, the dilution effect provides a public health incentive for conservation. This idea is particularly relevant for infectious diseases caused by pathogens circulated within natural reservoirs, such as influenza and ebolaviruses. The converse phenomenon, where biodiversity increases the risk, is known as the amplification effect.

The debate over the dilution effect has been fervent, particularly surrounding its generalisability and scale dependence. We address these concerns by combining popular Ordinary Differential Equation (ODE) models with two classes of agent-based models, one set on a lattice and the other spatially explicit. Agent-based models are uncommon in ecoepidemiology and have yet to be applied to the topic of dilution. We apply existing methods for quantifying the dilution effect in ODEs to numerical data from simulations and highlight circumstances where observations of dilution and amplification are sensitive to the selected definitions of biodiversity and infectious risk. We find that a lattice-based approach is well-suited to capturing spatiotemporal dynamics on longer time scales, while the spatially explicit method effectively describes outbreaks on shorter time scales. We also use a lattice-based model to explore the possible mechanisms of the dilution effect in horticulture.

We also highlight two infectious diseases to analyse in more detail: toxoplasmosis and Lyme disease. For toxoplasmosis, we discuss how pathogen influence on prey behaviour can affect infection risk, showing an amplification effect for an example system. With Lyme disease, we discuss the influence of a vector and show how control strategies associated with long-term mitigation of risk can cause a short-term increase in infection. We also show the diluting influence of adding a less competent host, concluding that the dilution effect is not a general phenomenon but a product of scale and the unique properties of individual infections and ecologies.

Acknowledgements

My PhD experience has been an arduous but rewarding process, made all of the more worthwhile by the fantastic people I had the opportunity to meet and the institution and community of which I could be a part. The first thanks must go to my supervisory team, David Aguirre, David Hayman, and my primary supervisor, Mick Roberts, for the academic guidance and endless patience that I had been afforded throughout the difficulties of the pandemic and beyond. Mick's supervision has been a valuable asset, both in developing my skills as an academic, providing levity when needed, and helping me to turn a broad topic into a coherent plan for a thesis. I am beyond thankful to have been your PhD student. Thanks also to James McCaw, my Master's supervisor, without whom I would not have had this opportunity.

I am also grateful to the Royal Society of New Zealand and its Marsden Fund for providing the funds for my study and for its wisdom in providing a grant for research in what I believe to be an exciting field. I also humbly acknowledge the financial support I received from ANZIAM to attend their conferences and present my work. Thanks must also go to Massey University for their financial support and to the entire university community at Albany, especially those at the Halls of Residence, for housing me throughout my studies.

I must also thank my friends for their advice and for providing distraction when needed, especially those at the Massey Albany Tabletop Club for helping me to relax on campus. Special mention must go to Jonathon for agreeing to live with me despite the thesis all-nighters and to Ryan, not just for your support but also for agreeing to read this thesis despite a misguided indifference to the field of mathematics. Most importantly, I must thank my family, who have supported me in everything I do, and my parents, who worked hard to ensure I had the freedom to choose the future I wanted for myself.

Finally, I sincerely thank my examiners, Prof. Kathryn Glass, Dr. Dion O'Neale and Prof. Thomas Pfeiffer, for their insightful feedback and enlightening discussion.

Contents

1 Introduction

In infectious disease ecology, the dilution effect is a subject of much controversy. It refers to a hypothetical relationship between infectious disease transmission and biodiversity in a given ecosystem, by which a greater degree of biodiversity reduces the risk of disease transmission. Conversely, an amplification effect refers to a possible relationship where more biodiversity results in a greater risk of transmission [70]. The premise of a dilution effect is an appealing prospect in theory, incentivising conservation efforts by linking biodiversity with a decrease in the risk of zoonotic diseases emerging in humans. The concept is particularly relevant for diseases that circulate within natural reservoirs, such as ebolaviruses and influenza, two pathogens that pose a significant threat to public health [152, 153]. Analysis by Taylor et al. found that while 61% of known pathogenic infectious organisms are zoonotic, this rises to 75% among those considered "emerging" at the turn of the century [132]. This overrepresentation, combined with the relevance of zoonotic pathogens such as influenza, ebolaviruses and coronaviruses, means that the existence and nature of the dilution effect are of considerable interest from both epidemiological and ecological perspectives.

A sizable body of research exists surrounding the dilution and amplification effects. Debate on this topic primarily focused on how generalisable the dilution effect is. Discourse initially centred on studies of tick-borne, zoonotic Lyme disease. Mice and deer in forest biomes host these ticks. Some authors argued that reforestation would improve biodiversity and reduce incursions by hosts into urban areas, causing fewer interactions with humans. Other authors claimed that there is some scale dependency, where reforestation could give rise to an amplification effect on a small scale, providing a suitable habitat for host and vector alike where one had not existed previously.

The dilution effect posits a relationship between biodiversity and the risk of emerging zoonotic infectious diseases in humans. However, it does not prescribe a particular measure or definition of either, posing a substantial challenge for comparing studies. Here, we will discuss possible ways of quantifying biodiversity and risk. We will also provide several examples of experimental and observational literature on both dilution and amplification effects to give a theoretical background for our research, illustrating the contentious nature of attempts to argue for a generalised dilution effect. The concepts discussed in this section will underpin our chosen modelling techniques and then our analysis of systems throughout.

1.1 Measures of Biodiversity

Biodiversity is measured in various ways. At the species level, these measurements all encompass two factors: species richness and species evenness [147]. Species richness is defined simply as the number of species in a given area. Mathematically, species richness is typically denoted by S, though to avoid conflict with the epidemiological convention of S referring to susceptible individuals, we will represent this with R. Species evenness refers to how close the species in an ecosystem are in numbers. Calculating species evenness requires knowledge of the relative abundance of each species. In ecology, evenness can be found observationally, but in a mathematical context, it can be calculated readily from the numerical output of a system of equations or a simulation.

Various metrics exist to quantify biodiversity by incorporating both richness and evenness. One such measure is the Simpson index, first proposed by Simpson in 1949 [125]. Simpson's index assumes that all individuals in an ecosystem can be classified into R distinct species, where R is the species richness and not to be confused with the removed compartment in epidemiological models. The Simpson index λ is given by

$$\lambda = \sum_{i=1}^{R} \left(\frac{n_i}{N}\right)^2 \tag{1}$$

where n_i is the number of individuals present of species i, and N is the total population of all species. The quantity n_i/N is the proportional abundance of each species, which we denote p_i.

The Simpson index is equal to the probability that any two randomly selected individuals in the ecosystem are of the same species. Accordingly, this probability is higher when one species dominates and reaches a minimum of $1/R$, where all species are represented equally. Therefore, the Simpson Index has a negative relationship to biodiversity, so the reciprocal of the Simpson Index is sometimes taken to create the inverse Simpson Index, denoted as D, to create a positively correlated measure. The Simpson Index can be subtracted from one to accomplish the same goal, creating the Gini-Simpson index. The Gini-Simpson index is thus the probability that two randomly selected population members belong to different groups.

Another index used to quantify biodiversity is the Shannon index. Introduced by Shannon in 1948, the index calculates the entropy (or uncertainty, more colloquially) associated with predicting the species of a randomly selected ecosystem member [123]. The Shannon index H is given by

$$H = -\sum_{i=1}^{R} p_i \ln p_i \tag{2}$$

where the base of the logarithm can be selected freely. A relative advantage of the Shannon index is that since $0 \leq p_i \leq 1$, the index will be greater where biodiversity is greater, with no transformation required. The maximum value where all species are equally present is $\ln R$, and when one species dominates, $H \to 0$.

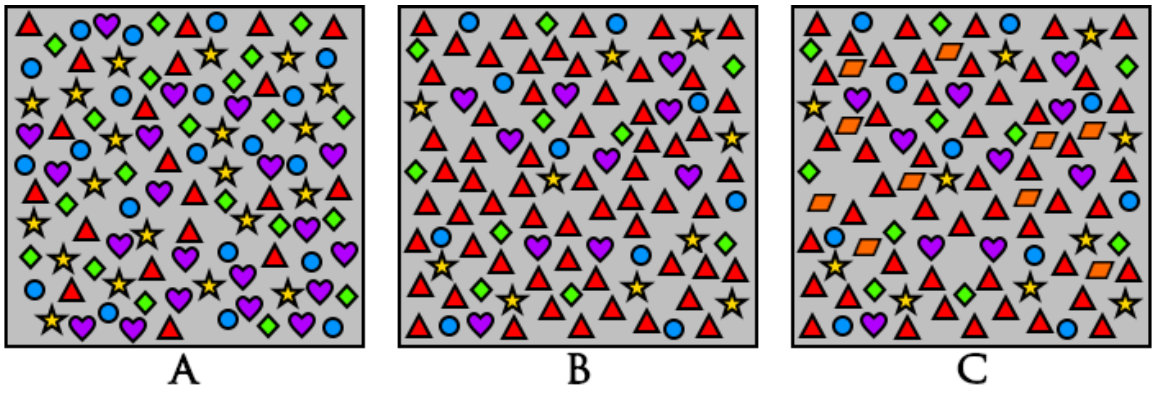

Figure 1: Three toy "ecosystems" of 100 individuals, where each coloured shape represents a different species. The species counts and biodiversity indices for each are presented in Table 1.

Ecosystem	**R**	**Counts**	$\boldsymbol{\lambda}$	$\mathbf{1-\lambda}$	**H**
A	5	$(20, 20, 20, 20, 20)$	$\frac{1}{5}$	$\frac{4}{5}$	1.609
B	5	$(60, 10, 10, 10, 10)$	$\frac{2}{5}$	$\frac{3}{5}$	1.227
C	6	$(50, 10, 10, 10, 10, 10)$	$\frac{3}{10}$	$\frac{7}{10}$	1.498

Table 1: Diversity index values for the ecosystems in Figure 1.

To demonstrate these indices in practice and their differences, we use the ecosystems represented in Figure 1, as shown in Table 1. Notably, the Simpson index λ can be observed to decrease with biodiversity, unlike with the Gini-Simpson index and the Shannon index. Additionally, system C's species richness is only an advantage to biodiversity where species distribution is comparatively more even; for all three indices, its biodiversity measurement falls between that of A and B, showing an advantage to

using an index rather than richness alone.

These biodiversity indices are underutilised in the context of dilution effect modelling, and so we aim to integrate them into our dilution effect analysis.

1.2 Measures of Infection Risk

Like biodiversity, infection risk is, in practice, a vague term. In particular, the term does not presuppose a subject to whom the risk applies, leaving it free for different studies to select different subjects along with different metrics, which poses a particular challenge for meta-analysis of experimental and observation work [119].

Popular measures of infection risk are infection prevalence or infection incidence. Incidence is the rate of new cases of infection, while prevalence is the proportion of cases within the population at any one time. In terms of risk, these can be characterised as probabilities. Incidence rates relate to the probability of an individual in a community becoming infected in a given period. In contrast, prevalence relates to the probability of a randomly sampled individual in a community being infected. Both quantities can be taken on a community level or applied to particular species as needed. These measures have the advantage of their relative ease of mathematical calculation and experimental observation. We particularly concern ourselves with prevalence P, calculated as $P = I/N$, where I is the number of infected individuals, and N is the total population.

For specific pathogens, infection risk could be measured directly by using the population of a target species. An example of where this might be appropriate is a vector or parasite's population. These are represented as N_i, where N is the population of the vector or parasite. Like with prevalence and incidence, these measures have the advantage of their mathematical simplicity. In the case of multiple vectors or a vector with a complex lifecycle, such as ticks, these populations can be weighted in terms of their role in transmission to a host of interest, such as humans, or an intermediate host, such as domestic cats, for toxoplasmosis.

Another measure of infection risk is the basic reproduction number R_0, defined as variously as the number [37], average number [43], or expected number [35] of secondary cases stemming from one infected individual in a fully susceptible population, typically in the absence of human intervention. It is helpful as it provides a threshold for whether an infectious disease can spread. If $R_0 < 1$, the disease will die out (except under certain conditions), whereas the disease can persist if $R_0 > 1$ [112]. In common

parlance, R_0 can be cited as a property of a particular pathogen, even though it also depends on other factors, such as the environment and the population, which may differ spatially, demographically or behaviorally. Mathematically, the definition is equivalent to $R_0 = \beta\tau$ where β is the number of contacts producing infection per unit time (often simplified to contact rate), and τ is the expected length of the infectious period. The quantity R_0 can be calculated using the parameters of a given mathematical model or estimated through observations or simulations. It provides another possible metric of infection risk, though its mathematical complexity can hinder direct comparisons with practical data.

This list is far from exhaustive. Similar quantities can be used, such as the force of infection Λ, which measures the rate at which susceptible individuals acquire infection. In discussing the literature surrounding the dilution effect, many measures of both biodiversity and infection risk appear, and our mathematical modelling can help demonstrate the role of metric choices in observing the dilution or amplification effect.

1.3 Quantifying the Dilution Effect

With a metric chosen for biodiversity and infection risk, it remains to use them to quantify the dilution effect. In 2018 Roberts and Heesterbeek addressed the question of how to quantify the dilution and amplification effects [116]. They introduced the *elasticity* of biodiversity x and a measure infection risk y, given by

$$\mathcal{D}y = \frac{x}{y}\frac{dy}{dx}. \tag{3}$$

Elasticity, in a mathematical sense, is a non-dimensionalisation of the derivative of infection risk with respect to biodiversity, allowing for easier comparison of different metrics. A dilution effect occurs when $\mathcal{D}y < 0$, while an amplification effect occurs when $\mathcal{D}y > 0$, provided that both measures used are positive and increasing. The opposite is true for a measure such as the Simpson Index, which decreases with biodiversity.

Analytical expressions for these quantities are generally difficult to derive for complex systems, especially where there are more than two species. When biodiversity and infection risk are affected by the change of a third parameter, the chain rule can be used to obtain an expression for dy/dx. The elasticity is readily applicable to all kinds of ecoepidemiological systems.

1.4 Examples of the Dilution and Amplification Effects

The study of the dilution effect has been mainly concentrated on its theoretical basis, with limited experimental testing to accompany hypotheses and observational work [148]. Nevertheless, a significant amount of research has found both positive and negative relationships between biodiversity and the risk of infection, using various metrics for each quality. This work spans a range of ecological and epidemiological contexts, with various modes of transmission and different interspecies relationships involved.

In support of the dilution effect, one study by Zhu et al. investigated the role of genetic diversity in rice crops in protecting against rice blast [154]. This fungal infection significantly suppresses yields in an agricultural setting with major economic ramifications. They investigated the influence of crop heterogeneity on the severity of rice blast outbreaks by enlisting the help of rice farmers in China's Yunnan Province in comparing mixed plots of glutinous and disease-resistant hybrid varieties of rice with control plots comprised of monocultures of each type. The more susceptible glutinous varieties were found to benefit substantially from the presence of the other variety, and even the more disease-resistant hybrid variety saw a reduction in disease severity. These findings support a dilution effect in this context, though the mechanisms that cause this are not immediately apparent. For mathematical modelling, the study suggests the possibility of investigating mechanisms using a stochastic, agent-based model.

One notable example of a study of the dilution effect in animals is in *Mycoplasma* infection among *Gerbillus* rodents in the Negev Desert. A study by Kedem et al. tested the relationship between host species richness and host infection probabilities through experiments and field observations [69]. The species in question each had different competencies for certain lineages of *Mycoplasma* bacteria. They concluded that the relationship between host-species richness and infection prevalence depended on the host-specificity of the relevant bacterium, with a dilution effect observed when the increase in richness was driven by the increasing prevalence of the least competent host. A later study found that the primary transmission route was through direct contact rather than vector-borne and also highlighted the significance of chronic infection in the continued persistence of the pathogen in the wild [30].

Examples of the amplification effect also exist in the literature, though they are less represented overall [150]. One example of an amplification effect concerns malaria caused by the protozoan *Plasmodium vivax* in the rainforests of Brazil's Atlantic coast. A key vector of malaria transmission in humans is the *Kerteszia* subgenus of *Anopheles*

mosquitos. The larvae of these mosquitos develop in water that gathers on the leaves of bromeliad plants, and their lifecycle depends strongly on the presence of these plants in their rainforest habitats, suggesting a positive relationship between the presence of high biodiversity rainforest biomes and malaria transmission to humans [90]. Subsequent modelling by Valle and Clark that concentrated on the Brazillian Amazon found that the strongest predictor of malaria risk was increased forest cover, with the effect being 25 times greater than that of land clearing [136].

In another study, Hechinger and Lafferty used videography to assess the relationship between bird communities and parasitic *Trematoda*, commonly known as flukes, that infect birds through snails [53]. They found a positive correlation between the richness, heterogeneity and abundance of birds with the same quantities measured in host snails on the same spatial scale. Bird species abundance, richness and heterogeneity were each positively correlated with the respective quantities in trematodes, suggesting an amplification effect. While the broad community focus with broad classes of parasites is a level of complexity beyond what we seek to model mathematically, it provides an important data point in considering the effect of scale on observations of dilution and amplification effects.

On an even larger scale, Wood et al. investigated the change in infectious disease burden over time in countries, using estimates of mortality and morbidity from Murray et al. [95, 151]. They compared these data to changes in biodiversity over time and found no correlation with infectious disease burden. In fact, they found reforestation to be positively correlated with infectious disease burden. Though the relationships behind the association were not identified, it highlights the need to consider the role of factors like urbanisation that could be a causal link between changes in infection risk and biodiversity on large scales.

Overall, the dilution and amplification effects have been extensively examined, and there is no shortage of examples of both. As such, our examples illustrate the wide range of ecologies that have been studied and demonstrate the importance of developing modelling tools that can apply to different modes of transmission, ecological scales, and interspecies interactions.

1.4.1 Lyme Disease Debate

Attempts to apply and generalise the dilution effect gave rise to debate in the literature. This debate is best embodied in the literature in the correspondence around Lyme Disease between Lafferty and Wood and Ostfeld and Keesing. In a paper titled

"Biodiversity loss and the rise of zoonotic pathogens", Ostfeld argues that human Lyme disease infection is connected to biodiversity through the nature of its hosts [102]. He argues that the preferred hosts of the nymphs responsible for the majority of infections are species that overwhelmingly dominate in urban and urban-adjacent biomes and that biodiversity there is negatively impacted by human factors. Conversely, where native species are present in a more diverse ecosystem, the wider pool of available, less ideal hosts mitigates the spread of disease and, thus, the risk to humans, an example of the dilution effect.

Wood and Lafferty, in 2013, provided a review on the subject and reached a different conclusion. They attempted to unify perspectives they label as "traditional", arguing that increased forestation increases the prevalence of competent hosts and Lyme disease, and "Dilution effect", which posits an opposite relationship similar to Ostfeld in 2009. Their solution to these apparently conflicting ideas was to argue for scale dependence, where increased forestation on a small level increases risk by providing the hosts with a habitat in the first place, while on a larger scale, the ecosystem can support more species and produce a dilution effect. They conclude that this example should result in reevaluating the dilution effect, de-emphasising generalisability in favour of thoroughly considering the ecosystems and diseases in question.

The conflicting arguments of each party led to a series of letters exchanging rebuttals on various topics [81, 104]. These include (but are certainly not limited to)

- Whether reforestation and species richness can be used as proxies for biodiversity
- Whether the initial arguments of Ostfeld painted an overly optimistic picture of how frequently the dilution effect is observed
- The precise nature of the mathematical relationship between biodiversity and the dilution effect (e.g. linear, asymptotic)
- How best to measure biodiversity and risk of human infection
- The effect of the order in which species are added or removed

In justifying their arguments, both parties identify that the broad acceptance of the dilution effect could profoundly influence policymakers and find common ground in cautioning against premature generalisations and arguing for the importance of a deep understanding of the many factors at play in the spread of Lyme disease. The various points made in this correspondence are illuminating, and together, the two parties provide a comprehensive examination of the issues that must be kept in mind

when modelling the dilution effect. It is also important to note that there are other discussions dilution effects as they apply to Lyme disease in the literature [101]. This debate has informed future work, for example, in examining the role of deer [59, 108].

1.4.2 Generality of the Dilution Effect

The examples of dilution and amplification effects paint a picture of a complex phenomenon that cannot be generalised, though research has yielded conflicting results, with numerous competing claims [102, 106]. Naturally, a number of meta-analyses have sought to reconcile these perspectives and resolve debate.

A meta-analysis by Civitello et al. used more than 200 biodiversity assessments in more than 60 ecological systems where an infection is present in animal hosts and found "overwhelming evidence of dilution" [28]. However, the results of the analysis are likely to be affected by selection and sampling biases as discussed by Salkeld et al. [119, 120], though this was disputed [29]. They also examined plant-herbivore systems and reached a similar conclusion. While their discussion allowed for exceptions, the clear conclusion of their research was that human-induced reductions in biodiversity were likely to increase disease burden in many species, with negative consequences for public health and agricultural and horticultural productivity [28].

A different meta-analysis in 2013 used fewer datasets, with stricter criteria for inclusion, excluding results using less direct measures of infection risk or biodiversity, such as the number of individuals of a reservoir host in an ecosystem. Analysis of their dataset of 16 found no statistically significant relationship between the measure of biodiversity used and the measure of infection risk. They also argued that a difference in the calculated p-value using all of their datasets vs. those drawn from published works was suggestive of some level of publication bias in favour of results showing a dilution effect. In concluding that there was a significant degree of heterogeneity in the systems studied, they suggested that different mechanisms were at play in the observed dilution effect in each system [119].

We conclude from these studies that there is a paucity of directly comparable datasets and that observation and experimental studies should be interpreted with the debate surrounding the dilution effect in mind. Mechanisms behind dilution and amplification are considered to be a promising future research direction, though discussion and commentary on the subject continues [50, 51, 71, 117].

The dilution effect is an appealing hypothesis to test, given its association of increased biodiversity with dual benefits in conservation and public health, and the role of publication bias can not be discounted. Given the costs and limitations of experimental and observational studies, a mathematical approach to assessing the dilution effect can be illuminating, and our research is motivated by a desire to use modelling tools to tackle the outstanding questions of the relationship between biodiversity and infection risk.

1.5 Thesis Overview

Having motivated our work by presenting key topics and definitions in ecoepidemiology, we conclude by outlining the content and key results of the chapters to come for the benefit of the reader.

Chapter 2 will discuss existing models that will form the basis of our work. We review ODE-based mathematical models that inform later chapters, such as the SIR model in epidemiology and the Lotka-Volterra models in ecology, alongside extensions that incorporate phenomena of interest, such as demography. Finally, we describe the relevant agent-based simulations in the literature. These models employ various spatial structures, including networks and lattices, as well as spatially explicit settings on a two-dimensional plane.

Chapters 3 and 4 then introduce novel agent-based models on lattices inspired by dilution and amplification effects cited in the literature. The third chapter develops a model for rice blast in an agricultural context by considering arrangements of two cultivars on a square lattice and shows a dilution effect can arise from two separate causes. The fourth chapter then considers a cellular automaton describing animals on a lattice, where one species is a predator and the other prey, both of which are hosts of the same infectious disease. This model shows how varying one parameter can result in either dilution or amplification, depending on the infection risk substitute and the chosen initial value of the parameter.

Chapter 5 introduces a novel spatially explicit model with multiple species under competition reflected in the choice of initial conditions. This model is presented with two possible infection mechanisms, namely infection through direct contact, and through environmental contamination. Using this model, we investigate the possibility of measuring the basic reproduction number R_0 through inference, before turning attention to a two-host system where a third, more competent host is introduced, counterintuitively resulting a dilution effect for many population configurations.

Chapters 6 and 7 use both novel and extended variations of existing ODE models to consider two relevant case studies that have yet to be examined in the literature: toxoplasmosis and Lyme disease. For the predator-prey toxoplasmosis model, we find a robust amplification effect when the pathogen evolves to increase predation of infected prey. The Lyme disease model in Chapter 7 describes ticks and a number of their host species, and considers the effectiveness of a deer cull in lowering infection risk in humans. Further work extends the model to include a third host species, and considers the role of that species' competence and interspecific competition on infection dynamics to address dilution questions outlined in the previous section.

Finally, Chapter 8 presents conclusions, comparing the two agent-based approached used and contrasting both with ODE approaches, and commenting on the meaning of our results in terms of the generality of the dilution effect. The appendices detail aspects of the agent-based models that were not used in the results presented, alongside code for selected agent-based models.

2 Epidemiological and Ecological Modelling

In assessing the dilution effect, a wide variety of existing mathematical and computational tools are available, which this chapter catalogues. The interdisciplinary nature of the dilution effect demands a substantial breadth of ideas from ecology and epidemiology. As such, this section will provide a basic introduction to models from both fields, in the form of differential equation models and stochastic simulations, and the tools used to analyse their behaviour. Depending on their background, the reader is encouraged to skip or read parts of this section as required for their understanding.

2.1 Epidemiological ODE Models

Deterministic epidemiological models prove very useful in our attempts to understand the dynamics behind infectious diseases. These models use a system of Ordinary Differential Equations (ODEs) to describe the spread of infectious disease, which can be solved to simulate its dynamics when combined with initial conditions. These models can help inform decision-making in infection control, which can prove invaluable [78, 145]. We provide an overview of the relevant models in the coming section.

2.1.1 SIR Model

One of the simplest epidemiological models is the Susceptible-Infectious-Recovered (SIR) model, first proposed by Kermack and McKendrick [74]. The eponymous variables are $S(t)$, $I(t)$, and $R(t)$. $S(t)$ denotes the number of individuals in a given population susceptible to a given infectious disease at any given time t. $I(t)$ denotes the number of individuals in this population infected with this disease at time t. Finally, $R(t)$ denotes the number of individuals who have recovered from the disease at time t and are assumed to be immune to further infection.

The rate at which individuals contract a disease per susceptible is denoted β, while the recovery rate is denoted γ. The population comprises N individuals. The resulting model is

$$\frac{dS}{dt} = -\frac{\beta IS}{N} \tag{4}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \tag{5}$$

$$\frac{dR}{dt} = \gamma I. \tag{6}$$

The $\beta IS/N$ term represents the flow of susceptible hosts into the infected state resulting from disease transmission between the infected and susceptible populations. The γI term signifies the flow of infected hosts into the recovered state. This model has a conserved quantity: $S(t) + I(t) + R(t) = constant = N$. Since N is constant, these equations can be rescaled so that the variables S, I, and R refer to the proportion of the population in each state, in which case $S(t) + I(t) + R(t) = 1$. The equations will then take the following form:

$$\frac{dS}{dt} = -\beta IS \tag{7}$$

$$\frac{dI}{dt} = \beta IS - \gamma I \tag{8}$$

$$\frac{dR}{dt} = \gamma I. \tag{9}$$

This system can also be simplified by substituting the basic reproduction number given by $R_0 = \beta/\gamma$, or $\beta N/\gamma$ in the rescaled system. If $R_0 < 1$, the infection will die out. If the second equation's right-hand side is negative initially, the disease will also die out. Since immunity is not lost following infection in this model, it is an assured outcome that the number of infected individuals eventually decreases to zero, i.e. $I(t) \to 0$ as $t \to \infty$. The final size equation can be calculated using this fact, substituting $I = 0$ while the left-hand side of each differential equation is also 0.

$$R_\infty = 1 - S(0)e^{-R_0(R_\infty - R(0))}, \tag{10}$$

where $R_\infty = \lim_{t\to\infty} R(t)$, the proportion of recovered individuals [74]. This equation is transcendental, meaning it cannot be solved algebraically using elementary functions. Solving it numerically can find the final size of an outbreak, $R_\infty - R(0)$, or the number of individuals who become infected over the course of the outbreak, including the initial infected population.

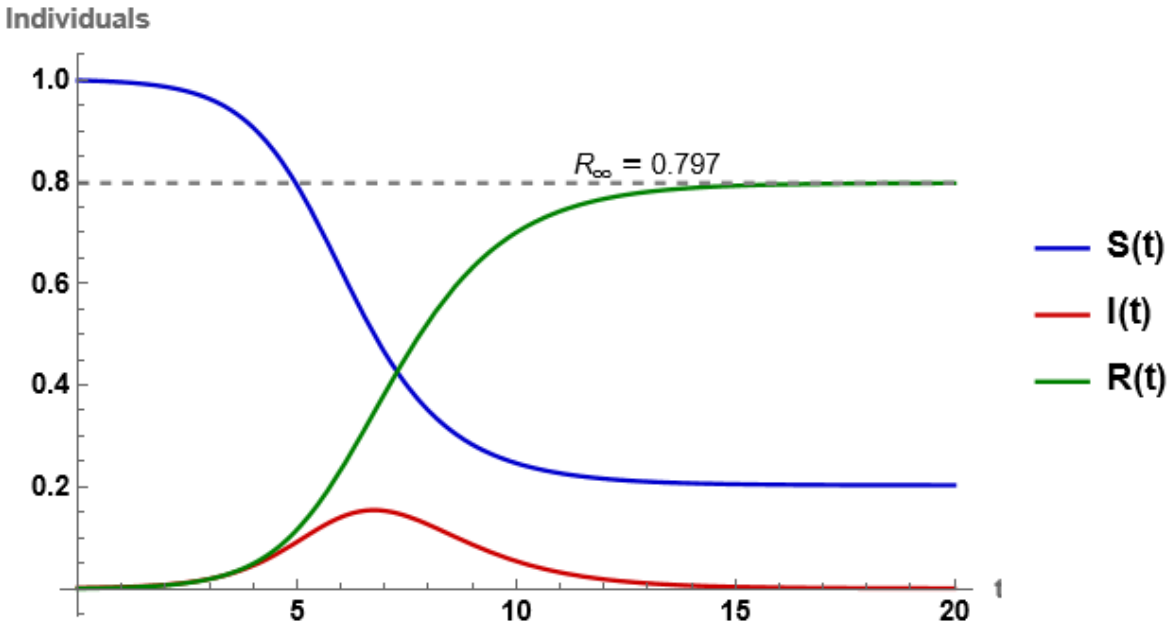

Figure 2: The SIR model with $N = 1$, $\beta = 2$, $\gamma = 1$, $R_0 = 2$. Initially, there is a very low proportion of infected individuals, with the rest of the population susceptible. The dashed line shows the final size.

Solving the system of ODEs gives the number of susceptible, infected and recovered individuals over time, hence describing the dynamics of the disease outbreak. An example solution to the SIR model where $R_0 = 2$ is shown in Figure 2.

The model can be expanded upon or even reduced in numerous ways, most obviously by the addition of states for exposed but not infectious individuals (SEIR model), eventual loss of immunity in recovered individuals (SIRS model), or the complete absence of immunity in recovered individuals (SIS model), among many others. These compartmental models can be selected to reflect the epidemiology of the infectious disease in question. For example, the SIR model might be appropriate for an infection for which immunity is near-total and lasts for a long time, such as measles, but the SIRS model might be more appropriate for an infection such as influenza.

2.1.2 Demography

The basic SIR model excludes the effects of births and deaths on the population; however, these can be easily introduced. If births occur at a rate μ, and deaths occur at a rate ν, then the frequency-dependent SIR equations become

$$\frac{dS}{dt} = -\frac{\beta IS}{N} + \mu N - \nu S \tag{11}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \nu I \tag{12}$$

$$\frac{dR}{dt} = \gamma I - \nu R, \tag{13}$$

where $N(t) = S(t) + I(t) + R(t)$, and is not constant. Since all newborns are placed into the susceptible category, these equations assume the absence of vertical transmission, i.e. transmission from parent to child through the reproductive process. Examples of infections transmitted vertically include the rubella virus (*Rubivirus rubellae*) and *Toxoplasma gondii*, the parasite that causes toxoplasmosis. Where vertical transmission is a factor, the system above can be amended to place an appropriate amount of newborns into the infected category.

A further assumption is that births and deaths are equal for each compartment. A real-world interpretation of this assumption would be that the lethality of infection would be negligible and that the uninfected and infected populations' life expectancy and reproductive rates are approximately equal. Should the population be at equilibrium, meaning the number of births and deaths are equal, we can take $\mu = \nu$. With a now constant population N, we can rescale our equations to remove a factor of N. The equations become

$$\frac{dS}{dt} = -\beta IS + \mu(1 - S) \tag{14}$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \mu I \tag{15}$$

$$\frac{dR}{dt} = \gamma I - \mu R. \tag{16}$$

To examine the long-term behaviour of the model, we can linearise the system about its equilibria. Solving for when the rate of change of each compartment is 0 (i.e., when $S'(t) = I'(t) = R'(t) = 0$), there are two equilibria, one infection-free non-endemic state, and one endemic state where infection persists. The non-endemic equilibrium is at $(S^*, I^*, R^*) = (1, 0, 0)$, and the endemic equilibrium is given by

$$S^* = \frac{\mu + \gamma}{\beta} \tag{17}$$

$$I^* = \frac{\mu}{\beta}\left(\frac{\beta}{\mu + \gamma} - 1\right) \tag{18}$$

$$R^* = \frac{\gamma}{\beta}\left(\frac{\beta}{\mu + \gamma} - 1\right). \tag{19}$$

Unlike the SIR Model without demography, there is a steady state where the infection persists, as births supply new susceptible individuals, as seen in Figure 3.

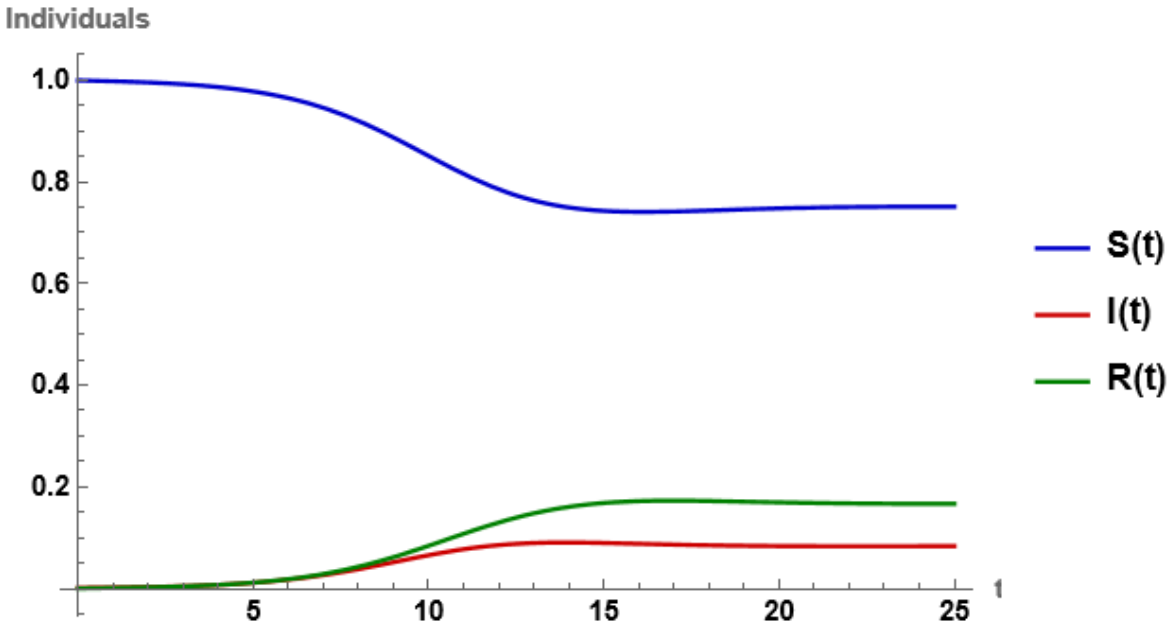

Figure 3: The SIR model with demography where $N = 1$, $\beta = 2$, $\gamma = 1$, $\mu = \nu = 0.5$. Initially, there is a very low proportion of infected individuals, with the rest of the population susceptible.

To determine a threshold for the stability of the equilibria (analogous to finding the basic reproduction number R_0), we must calculate the Jacobian J. If the right-hand sides of each differential equation are denoted f_S, f_I, and f_R then

$$J = \begin{bmatrix} \frac{\partial f_S(t)}{\partial S} & \frac{\partial f_S(t)}{\partial I} & \frac{\partial f_S(t)}{\partial R} \\ \frac{\partial f_I(t)}{\partial S} & \frac{\partial f_I(t)}{\partial I} & \frac{\partial f_I(t)}{\partial R} \\ \frac{\partial f_R(t)}{\partial S} & \frac{\partial f_R(t)}{\partial I} & \frac{\partial f_R(t)}{\partial R} \end{bmatrix} \tag{20}$$
$$= \begin{bmatrix} \beta I - \mu & -\beta S & 0 \\ \beta I & \beta S - \mu - \gamma & 0 \\ 0 & \gamma & \mu \end{bmatrix}.$$

The Hartman-Grobman theorem states that the stability criterion for each fixed point is when all of its corresponding eigenvalues λ have a negative real part, i.e. $Re(\lambda) < 0, \ \forall\lambda$. Substituting the non-endemic equilibrium $S = 1, I = 0$ into the Jacobian and finding the eigenvalues, we have an eigenvalue of multiplicity two where $\lambda = -\mu$ and a non-repeated eigenvalue where $\lambda = \beta - \mu - \gamma$. Since $\mu > 0$ by definition, all of the eigenvalues have negative real parts if $\gamma + \mu < \beta$.

This method is broadly applicable to most ODE systems and will be helpful for the models we create. The exact process can be followed for this model using the endemic equilibrium to show that it is stable where the non-endemic equilibrium is unstable. Indeed, it can be shown that $R_0 = \beta/(\mu + \gamma)$. This model provides an elementary way of including population and infectious disease dynamics in the same system. However, more elaborate models can be constructed by incorporating more complex ecological dynamics, as detailed in a later chapter.

2.1.3 Environmental Transmission

The previous models all assume that transmission occurs by contact between susceptible and infected hosts. However, one way this model can be elaborated upon is by introducing an environmental mode of transmission. A significant number of infections are not transmitted directly from host to host, and are spread via other means, such as a vector, or the environment itself. Malaria, for example, is caused by *Plasmodium* protozoa spread by mosquitos, while Lyme disease is caused by *Borrelia* bacteria and spread by ticks [23, 103]. Other diseases are spread via soil, water or other such environmental features. Waterborne diseases include the protozoal disease giardiasis and the bacterial diseases cholera and dysentery, while the *Toxoplasma gondii* parasite can be soilborne by way of the faeces of an infected cat [25, 24].

Incorporating these into our single-species system of differential equations can be done in various ways depending on the role of the environment in the infection cycle. We present a simple example in the following system of differential equations that includes simple demography, adapted from Kelting [72]:

$$\frac{dS}{dt} = -\omega ES + \mu - \nu S \tag{21}$$

$$\frac{dI}{dt} = \omega ES - \gamma I - \nu I \tag{22}$$

$$\frac{dR}{dt} = \gamma I - \nu R \tag{23}$$

$$\frac{dE}{dt} = \lambda I - \eta E, \tag{24}$$

where $E(t)$ is the amount of the pathogen which persists in the environment at time t. The pathogen is spread to the environment at a rate of λ per infected individual, while η is the rate at which the pathogen in the environment is lost. Taking the role of β in the standard SIR model, ωES is the rate at which the pathogen spreads from the environment to susceptible individuals. Figure 4 shows a possible solution for the above system of ODEs. There remains a conserved quantity with $S + I + R = 1$ only when $\mu = \nu$.

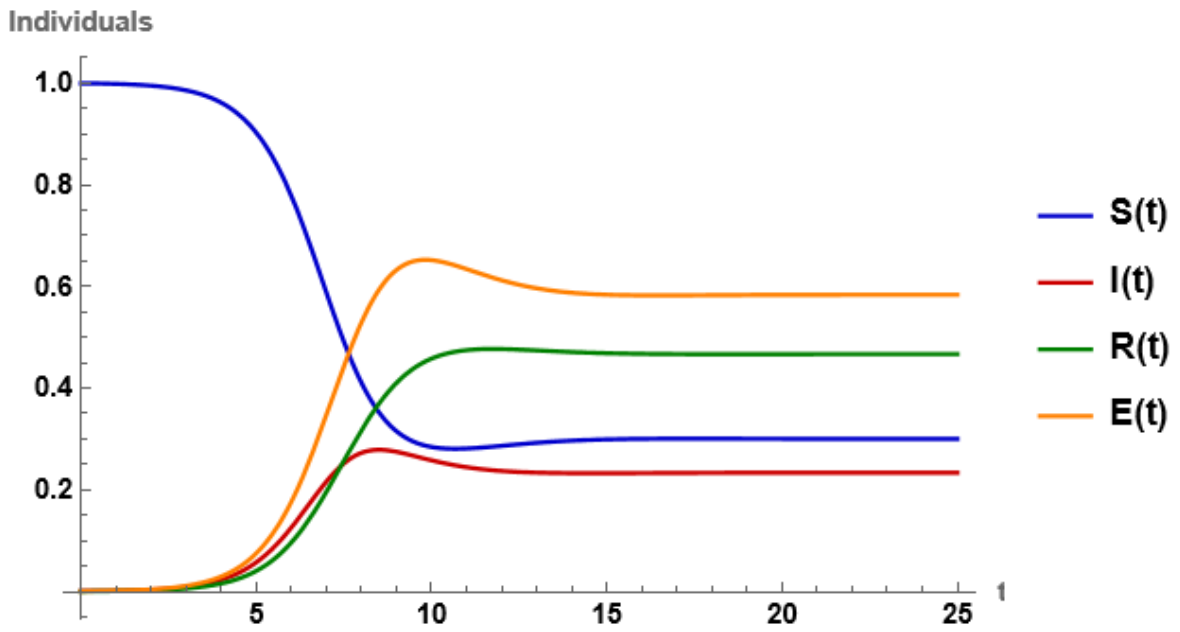

Figure 4: The SIR model with environmental transmission where $\omega = 2$, $\gamma = 1$, $\lambda = 2.5$, $\eta = 1$, $\mu = \nu = 0.5$. Initially, there is a very low proportion of infected individuals, with the rest of the population susceptible and no pathogen present in the environment.

Even this simple formulation applies to multiple potential contexts. We could consider a system where E represented infected individuals of a vector species that does not interact ecologically with the host population and persists at a constant population with a turnover rate of η. Such a vector species would transmit infection from infected hosts to susceptible hosts without transmission between the vector. We could equivalently consider a system in which a pathogen is shed by infected individuals into the environment, where it cannot survive indefinitely and eventually decays. Infection propagates when susceptible individuals come into contact with the pathogen in the environment and become infected. It would be assumed in this context that susceptibles consuming the pathogen would not significantly reduce the amount of it in the environment; this would be because vastly many infectious agents such as bacteria, protozoa and viruses are shed into the environment by each infected individual, while it typically only takes a comparatively low number of these agents to enter a host for infection to take hold [72]. We also assume that recovered individuals cease to be infectious in both of these contexts. However, it is essential to note here in particular that this is not always the case and that chronic infection exists in many pathogens.

The model has four compartments and shows that even a modest increase in sophistication can require different methods to determine quantities like R_0. Calculating

the steady states, we can show that there remain two equilibria, one of which is non-endemic. These are given by $S^* = \mu/\nu, I^* = R^* = E^* = 0$, and

$$S^* = \frac{\eta(\gamma + \nu)}{\lambda\omega} \tag{25}$$

$$I^* = \frac{\mu}{\gamma + \nu} - \frac{\eta\nu}{\lambda\omega} \tag{26}$$

$$R^* = \gamma\left(\frac{\mu}{\nu(\gamma + \nu)} - \frac{\eta}{\lambda\omega}\right) \tag{27}$$

$$E^* = \frac{\lambda\mu}{\eta(\gamma + \nu)} - \frac{\nu}{\omega}. \tag{28}$$

Evaluating the Jacobian and calculating eigenvalues with a matrix of dimension three is difficult, requiring the analytical solution of a cubic equation. In this case, we can use the Routh-Hurwitz criteria for a third-order polynomial to find the root of the cubic expression. [60, 118] These criteria determine whether all of the eigenvalues have negative real parts, but this is not always analytically possible or convenient, so we use this model as a toy example to introduce a different tool to do so, known as the **next-generation matrix**.

The next-generation matrix K describes the next generation of infections from a particular host to another host type, as first introduced by Diekmann et al. in 1990 [35]. Though subsequent work has moved away from this intuitive construction, the process outlined here produces an equivalent derivation of R_0 [36, 113, 115]. For K, the host types need not be different species or even different hosts at all. Rather, they account for distinct epidemiological categories. For example, the types may be hosts in different locations or that acquired their infections through different transmission routes. For the previous systems, there was only one host type. However, in this case, the addition of environmental transmission constitutes an additional host type, considering the environment as part of the transmission route.

To construct the next-generation matrix, we consider a submatrix of the Jacobian J containing only the elements corresponding to the ODEs for the infectious states where the derivative is taken with respect to one of the infectious states [36, 115]. The submatrix is denoted by H, which is evaluated at the non-endemic equilibrium, or indeed equilibria, for models where multiple exist. For the model with environmental transmission

$$H = \begin{bmatrix} \frac{\partial f_I(t)}{\partial I} & \frac{\partial f_I(t)}{\partial E} \\ \frac{\partial f_E(t)}{\partial I} & \frac{\partial f_E(t)}{\partial E} \end{bmatrix} \tag{29}$$
$$= \begin{bmatrix} \gamma - \nu & \omega S^* \\ \lambda & -\eta \end{bmatrix},$$

where S^* is the **non-endemic** equilibrium, in this case μ/ν. The matrix H can be broken down into the sum of two other matrices. The first is the transmission matrix, T, containing the terms relating to the production of new infections. The second of these, Σ, is the transition matrix, containing state changes like the death, removal or acquisition of immunity. The decomposition of $H = T + \Sigma$ for our system is

$$T = \begin{bmatrix} \gamma & \frac{\omega\mu}{\nu} \\ \lambda & 0 \end{bmatrix} \tag{30}$$

$$\Sigma = \begin{bmatrix} -\nu & 0 \\ 0 & -\eta \end{bmatrix}. \tag{31}$$

Using T and Σ, the next-generation matrix is defined $K := -T\Sigma^{-1}$, and R_0 is given as the spectral radius (dominant eigenvalue) of K, $R_0 = \rho(K)$. This decomposition is important, as it reconciles the original idea of tracking the next generation of infections with other dynamics present in the model, such as demography. Since Σ is diagonal, K can be simply found to be

$$K = \begin{bmatrix} \frac{\gamma}{\nu} & \frac{\omega\mu}{\eta\nu} \\ \frac{\lambda}{\nu} & 0 \end{bmatrix} \tag{32}$$

and its two eigenvalues, which we denote with subscripts to make them distinct from the shedding rate λ, are

$$\lambda_{1,2} = \frac{\gamma\eta \pm \sqrt{\eta}\sqrt{\gamma^2\eta + 4\lambda\mu\omega}}{2\eta\nu}. \tag{33}$$

Since the last term on the numerator must always be positive, we find that R_0 is the eigenvalue with the plus sign taken. Considering that all parameters are strictly

positive, we obtain two criteria for the stability of the endemic equilibrium. Where $\gamma > 2\nu$, $R_0 > 1$ will always be satisfied, but if $0 < \gamma \leq 2\nu$, then the following must also hold:

$$\lambda\mu\omega > \eta\nu(\nu - \gamma). \tag{34}$$

This model demonstrates that even relatively simple epidemiological models can be complex to analyse, with the next-generation matrix being a vital tool when multiple host types are present. Epidemiological models beyond those reviewed here describe all manner of possible infection dynamics, but in the interest of time, we withhold those until they are needed.

2.2 Ecological ODE Models

In producing models that describe an infectious disease in multiple species, we must also account for possible interactions between species. Models for species within ecosystems have their roots in population modelling, which originated in the 19th century. We present here some ecological models that describe the population dynamics of different numbers of species, beginning with a single species before moving on to multiple species.

2.2.1 Exponential and Logistic Growth

For one species, consider the most basic way of modelling a population.

$$\frac{dx}{dt} = rx. \tag{35}$$

Here, x is the population of a particular species, while r is the intrinsic growth rate, equal to the birth rate minus the death rate. This population has exponential growth if $r > 0$; it remains constant if $r = 0$, and decays exponentially if $r < 0$. Though simple, this model is unrealistic since individuals of the same species typically must compete for space or resources in a given environment, meaning that exponential growth cannot continue without bound. Throughout the 1830s and 1840s, Verhulst created the logistic population model, which adjusts for this observation by introducing a carrying capacity, or the maximum possible population density of a species that an environment can support [26]. For an environment with carrying capacity K, the logistic population model is given by

$$\frac{dx}{dt} = rx(1 - \frac{x}{K}). \tag{36}$$

This differential equation has the solution

$$x(t) = \frac{K x_0 e^{rt}}{K + x_0(e^{rt} - 1)}, \tag{37}$$

where x_0 is the initial population size [139, 141]. Verhulst labelled this solution the logistic curve [140]. The population density approaches K for as long as $x_0 \neq 0$, providing a stable endemic equilibrium, unlike the exponential growth equation. There is also an extinction solution where $x = 0$ for all r. It can be observed from the solution that the endemic solution is stable where $r > 0$, and the non-endemic solution is stable where $r < 0$.

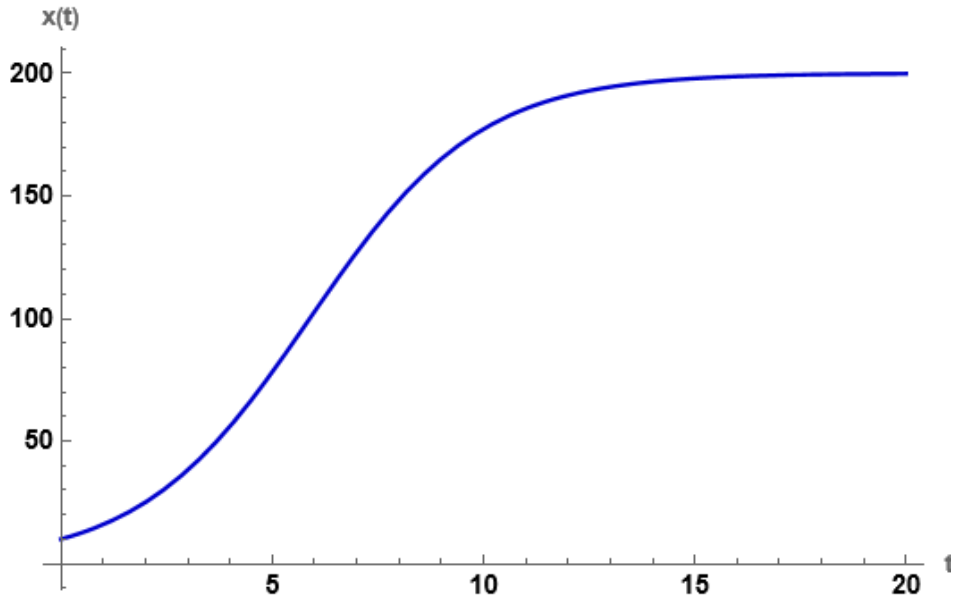

Figure 5: A logistic population model with $K = 200$ and $r = 0.5$. The initial population $x_0 = 10$.

Figure 5 shows a solution of the model, with $x(t)$ asymptotically approaching the carrying capacity $K = 200$. Though simple, this model formed the basis for more sophisticated later population models.

2.2.2 Predator and Prey

Multiple species present in an ecosystem can interact in numerous possible ways. They can interact as predator and prey, as competitors for food and other resources, or through mutualism, in which both species benefit from the presence of the other. We begin by considering a two-species system described by the Lotka-Volterra predator-prey model. Developed concurrently and independently by Lotka and Volterra, it

originally was intended for application in periodic chemical reactions before being applied to an ecological context [15, 86, 87, 143]. The model is given by

$$\frac{dx_1}{dt} = r_1 x_1 - \phi_{12} x_1 x_2 \tag{38}$$

$$\frac{dx_2}{dt} = r_2 x_2 - \phi_{21} x_1 x_2, \tag{39}$$

where x_1 is the prey species, and x_2 is the predator species. The first term in the x_1 differential equation represents the natural growth of the prey population in the absence of predators, while the second term represents the reduction of the prey population as a result of consumption by predators. In keeping with this interpretation, we restrict the parameters by having $\phi_{12} < 0$ and $r_1 > 0$ to ensure that predators consume prey and that the prey would reproduce in their absence. In the second equation, the first term represents the natural growth of the predator population, while the second represents the increase in fecundity that depends on the prey population. Since the predator population should depend on the prey to persist, we require that $r_2 < 0$ and $\phi_{21} > 0$. This model assumes that the prey is assumed to have plentiful food and that the predators are assumed to consume only prey and have a constant appetite; i.e. the rate at which they consume prey per individual is constant [32].

This model has two fixed points, the first of which is $x_1^* = x_2^* = 0$, and the second of which is $x_1^* = -r_2/\phi_{21}$, $x_2^* = r_1/\phi_{12}$. The Jacobian of the system is given by

$$\begin{aligned} J &= \begin{bmatrix} \frac{\partial f_{x_1}(t)}{\partial x_1} & \frac{\partial f_{x_1}(t)}{\partial x_2} \\ \frac{\partial f_{x_2} t)}{\partial x_1} & \frac{\partial f_{x_2}(t)}{\partial x_2} \end{bmatrix} \\ &= \begin{bmatrix} r_1 + \phi_{12} x_2^* & \phi_{12} x_1^* \\ \phi_{21} x_2^* & r_2 + \phi_{21} x_1^* \end{bmatrix}. \end{aligned} \tag{40}$$

For the extinction equilibrium, the eigenvalues λ are $\lambda_1 = r_1$ and $\lambda_2 = r_2$. Therefore, extinction is always unstable, with one positive and one negative real eigenvalue corresponding to a saddle point in the phase plane. For the other equilibrium, we see that $\lambda^2 = r_1 r_2$. Since r_2 is negative, we conclude that the two eigenvalues have zero real part and are thus purely imaginary, corresponding to the long-term behaviour of closed oscillations about the fixed point. Our eigenvalue analysis reveals that for as long as the initial populations of both species are non-zero, the species will coexist indefinitely, with stable periodic behaviour, as shown in Figure 6.

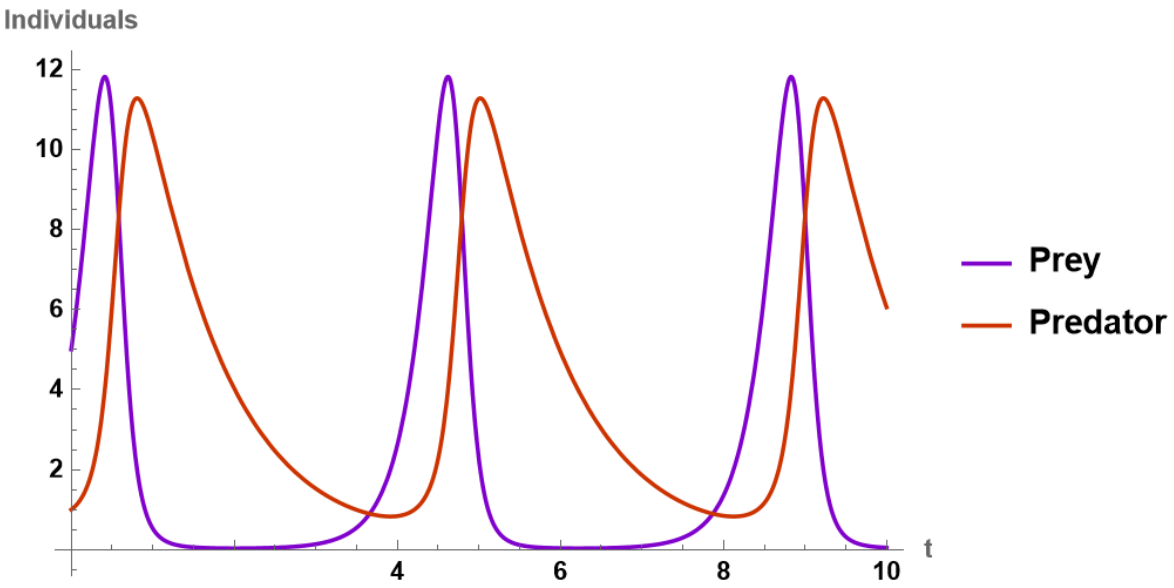

Figure 6: A solution to the predator-prey model with species 1 being the prey, and species 2 being the predator. The parameter values are $r_1 = 4$, $r_2 = -1$, $\phi_{12} = 1$, $\phi_{21} = -0.5$. The initial conditions are $x_1(0) = 5$, $x_2(0) = 1$.

The exact nature of the oscillations can be found by examining the phase plane. Dividing one equation by the other and performing separation of variables gives two readily integrable expressions, given by

$$r_1 \ln(x_2) - r_2 \ln(x_1) - \phi_{12} x_2 + \phi_{21} x_1 = \text{const.} \tag{41}$$

In the phase plane, the expressions form closed trajectories that comprise the limit cycles, with their exact trajectory dependent on the initial conditions. These trajectories can reveal some of the limitations of this model for a predator-prey system, as some oscillations result in a species repeatedly approaching extinction before rebounding, an unrealistic phenomenon referred to as the "atto-fox problem", and can be observed in the prey population of Figure 6 [93].

2.2.3 Resource Competition

Another possible way two species could interact is by competing for a shared pool of resources. In such a system, interspecies competition for resources would typically occur alongside intraspecies competition. This system is described by the competitive Lotka-Volterra equations, which are given by:

$$\frac{dx_1}{dt} = r_1 x_1 \left(1 - \left(\frac{x_1 + \alpha_{12} x_2}{K_1}\right)\right) \tag{42}$$

$$\frac{dx_2}{dt} = r_2 x_2 \left(1 - \left(\frac{x_2 + \alpha_{21} x_1}{K_2}\right)\right), \tag{43}$$

where α_{ij} is the effect of species j on the birth rate of species i relative to the effect of the intraspecies competition for species i. In these equations, it is assumed that members of each species must not only compete among themselves for resources but also with the other species. In the absence of this competition, growth for each species is exponential. Like in the one-species logistic model, each species i has a carrying capacity K_i in the absence of the other. The system of equations can be rewritten as

$$\frac{dx_1}{dt} = r_1 x_1 - \phi_{12} x_1 x_2 - \phi_{11} x_1^2 \tag{44}$$

$$\frac{dx_2}{dt} = r_2 x_2 - \phi_{21} x_1 x_2 - \phi_{22} x_2^2 I, \tag{45}$$

where the coefficients ϕ_{ij} are defined equivalently to how they were with the predatory-prey model. Since all of the logistic interactions are competitive, $\phi_{ij} > 0 \ \forall i, j$, and the growth rates $r_i > 0 \ \forall i$. The system has four solutions, each allowing a different combination of species to persist. These states are $\{x_1^*, x_2^*\} = \{0, 0\}, \{r_1/\phi_{11}, 0\}, \{0, r_2/\phi_{22}\}$, and are analogous to the solutions for the one-species model. The fourth equilibrium state is given by

$$x_1^* = \frac{r_1 \phi_{22} - r_2 \phi_{12}}{\phi_{11}\phi_{22} - \phi_{12}\phi_{21}} \tag{46}$$

$$x_2^* = \frac{r_2 \phi_{11} - r_1 \phi_{21}}{\phi_{11}\phi_{22} - \phi_{12}\phi_{21}}. \tag{47}$$

To consider stability, we once again find the Jacobian

$$J = \begin{bmatrix} r_1 - 2\phi_{11}x_1^* - \phi_{12}x_2^* & -\phi_{12}x_1^* \\ -\phi_{21}x_2^* & r_2 - 2\phi_{22}x_2^* - \phi_{21}x_1^* \end{bmatrix} \tag{48}$$

and derive its eigenvalues at the fixed points. At $(0, 0)$, the eigenvalues $\{r_1, r_2\}$ are both positive by definition, meaning it is always unstable. The two cases where one species exists are symmetric, with the eigenvalues for species i surviving and species j becoming extinct given by $\{-r_i, r_j - r_i\phi_{ji}/\phi_{ii}\}$. For the second eigenvalue to be negative, and hence the fixed point to be stable, it is required that $r_j < r_i\phi_{ji}/\phi_{ii}$. This condition effectively describes a species outcompeting the other. For the coexistence equilibrium, a different technique must be used.

The **invasibility criterion** states that for two species to coexist stably, both must be able to invade a monoculture of the other species. The criterion is necessary and sufficient for the stability of the coexistence equilibrium [18, 48]. If either species is absent, then the other species will be at its carrying capacity, equal to its equilibrium in the absence of the other species. For x_1, this is r_1/ϕ_{11}. The condition is satisfied if the per capita growth rates are positive when the relevant species is absent and the other at its carrying capacity. These growth rates are obtained by simply taking out a factor of x_i in the respective ODE. Following this and noting the symmetry of the model, it is required for species i to have a positive growth rate, then $r_i < r_j\phi_{ij}/\phi_{jj}$. If this holds for both species, then the coexistence equilibrium is stable.

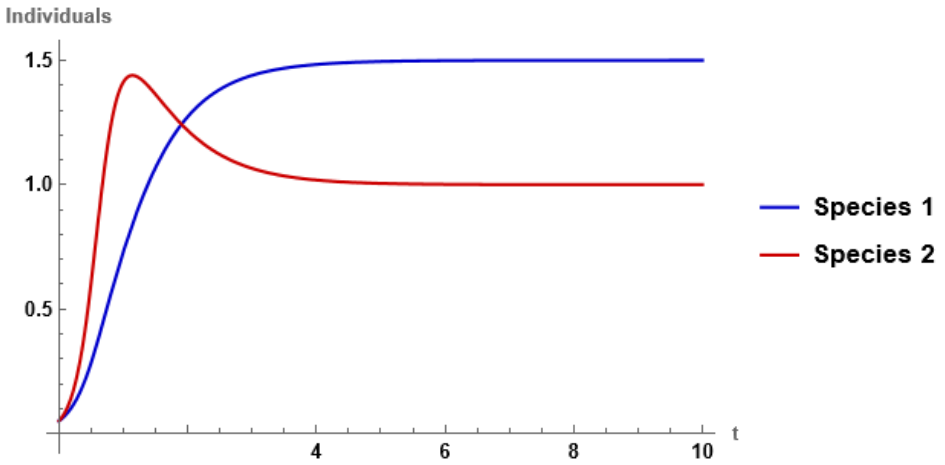

Figure 7: A two-species competition model with $r_1 = 4$, $r_2 = 6$, $\phi_{12} = 1$, $\phi_{21} = 2$, $\phi_{11} = 2$, $\phi_{22} = 3$. The initial population $x_1(0) = x_2(0) = 0.05$.

In ecological terms, this condition is the equivalent of stating that each species must have a stronger suppression effect on itself than on the other species. [48] All of these conditions account for the entire parameter space, excluding bifurcation at the critical point, though a catalogue of the entire space for the two-dimensional Lotka-Volterra equations exists through an analogue with the replicator equation used in game theory [16]. Analysis of this model provides a valuable tool for our models, particularly with a threshold for changes in species richness.

2.2.4 A Generalised Model for Multiple Species

There are various ways species can interact, and complex interrelationships can be formed for ecosystems of more than two species. For example, one predator species can consume two species as prey while these two species compete for resources like space or

a food source. Even interspecies relationships are not limited to solely predator-prey and competitive dynamics, as numerous symbiotic relationships exist between animals and plants. Until this point, we have used x_i to denote the population of species i, as is common in ecological texts, but for our purposes, we use the epidemiological standard for population N. For n species, a generalised system of n equations that can capture a wide range of dynamics is described by the following system of differential equations:

$$\frac{dN_k}{dt} = r_k N_k - N_k \sum_{i=1}^{n} \phi_{k,i} N_i. \tag{49}$$

The parameter $\phi_{k,i}$ describes the interactions between each species. A positive value corresponds to a decrease in growth due to these interactions, while conversely, a negative value corresponds to an increase in growth. In addition, the $\phi_{i,i}$ terms in these equations correspond to the intraspecies resource competition resulting in logistic growth discussed in the previous section. All together, ϕ is known as the **interaction matrix** [14]. Some of the dynamics it captures are presented in Table 2.

$\phi_{i,j}$	$\phi_{j,i}$	Corresponding relationship
> 0	< 0	Predator-prey
> 0	> 0	Resource competition
$= 0$	$= 0$	No interaction
$= 0$	< 0	Commensalism
$= 0$	> 0	Resource destruction
< 0	< 0	Mutualism

Table 2: Real-world interpretations of the relationship between two species i and j, denoted by the parameter $\phi_{i,j}$.

Our review of two-dimensional Lotka-Volterra models showed long-term dynamics involving convergence to equilibrium and stable oscillatory behaviour. However, Lotka-Volterra models with higher species richness can have a wide range of behaviour. For example, in four dimensions, Vano et al. described a purely competitive system that is chaotic, with the behaviour of a strange attractor [137]. However, the literature imposes conditions for the possible dynamics of these systems in specific cases. Smale proved that a five-species competitive Lotka-Volterra system could feature any possible asymptotic behaviour, including attractors, fixed points, cycles and limit toruses [128]. Going further than this, Hirsch showed in a series of papers that attractor dynamics

occur on a manifold of $n-1$ dimensions, meaning that cycles can only feature where $n \geq 3$, and toruses and chaos where $n \geq 4$, providing a useful guide to system behaviour for models with more species [54, 55, 56].

The Lotka-Volterra family of models has limitations, such as the lack of a non-linear functional response, i.e. the relationship between food density and reproduction rates. These ecological models can be combined with various possible infection dynamics, creating a class of ODE models that describe an ecosystem both epidemiologically and ecologically.

2.3 Agent-Based Models

Though useful, ODE-based models have their limitations. ODE models operate in what is known as the thermodynamic limit, meaning that the behaviour of individuals and infectious agents where populations are assumed to be effectively infinite [66]. Consequently, the system's population density is held constant, and it is impossible to examine the role of scale by adjusting the total population alone. Additionally, the ODE models discussed previously are deterministic, meaning that their behaviour always evolves the same way no matter the initial conditions, with no heterogeneity in host infection outcomes beyond what is specified in the existing compartments. The assumption that all hosts are equally capable of infecting all other hosts is known as homogeneous mixing. Social and spatial factors render this assumption unrealistic in many contexts.

In contrast with compartmental ODE models, agent-based models are stochastic, with each host usually modelled independently on a series of discrete time steps rather than over a continuous time interval. Stochasticity allows for the effects of randomness to be examined and provides a tool for considering systems with the same population density but different sizes. There is a gap in the literature regarding applications of agent-based techniques to the dilution effect that we aim to address here, so we give a brief account of agent-based modelling techniques. It should be stressed that agent-based models are not the only approach to addressing the shortcomings of ODE models; other approaches are briefly summarised in Section 2.4.

We discuss two distinct approaches for handling infectious agents in agent-based models. One of these models the progression of an infection on a lattice, while another occurs on a 2D plane in continuous space. We assess model behaviour using the Monte Carlo method for our agent-based models. The method involves repeating the same stochastic process with a particular set of inputs a large number of times and deriving

results from the outputs of each simulation. Aggregating the results of all simulations yields a statistical approximation of desired quantities. Monte Carlo methods are an effective way of analysing model behaviour but have limitations, including the inherent trade-off of model detail and sample size with the associated computation expense [92].

2.3.1 Network and Lattice-Based Models

One approach to agent-based modelling assumes a spatial structure of its population. For infectious diseases, we are primarily interested in infectious disease modelling on a 2D lattice using cellular automata. Cellular automata are models involving a regular structure of cells (representing individuals) that exist in a finite number of discrete states and simulate dynamics by iterating over a series of time steps. In the context of infectious diseases, the cells represent individual hosts, and the states of the cells specify their infection status. In ecology, the cells represent a point in space, and their states represent the species of the agent occupying the cell. Regular lattice structures can be triangular, square or hexagonal, analogous to tesselations of 2D space. The SIR and SIS models have already been applied to lattices for single hosts on various structures. Some models opt for synchronous simulation by updating all sites simultaneously. In contrast, others opt for asynchronous simulation by sequentially performing an algorithm on randomly selected sites and updating the system, one agent at a time.

A paper by Arashiro and Tomé is of particular interest, as their agent-based model provides a predator-prey model on a 2D lattice and then models infection using a special case of the same model [5]. The cells in their base model update synchronously and exist in three states: one prey state (denoted X), one predator state (denoted Y), and one empty state (denoted Z). Each cell's neighbourhood contains the four other cells to the North, South, East, and West, referred to as the von Neumann neighbourhood. If a cell is in the empty state, it is occupied by prey at a maximum rate a, proportional to the number of prey in the neighbourhood. A predator can also replace prey in a cell it occupies at a maximum rate b, proportional to the number of predators in the neighbourhood. If a predator occupies a cell, it can become empty at probability c. This model features stable steady states where predator and prey coexist, unlike the Lotka-Volterra predator-prey ODEs. The model also features local oscillations in predator-prey populations, akin to the asymptotic behaviour of the ODE model [21, 122].

For $a = 0$, Arashiro and Tomé consider that the prey state X and predator state Y are analogous to the susceptible state S and infected state I, respectively. If the empty state Z is equivalent to the removed/recovered state R, then the model can produce

SIR dynamics on a 2D lattice. Their model uses the constraint $a + b + c = 1$, finding a threshold for outbreaks to occur at $c \approx 0.220$. This paper is highly relevant as it demonstrates similar approaches to ecology and epidemiology, suggesting a method for combining the mechanisms.

Other similar epidemic models exist, using different lattice structures and obtaining different outbreak thresholds. These include regular lattices on triangles and hexagons and feature less stringent thresholds for infection probabilities for lattices with larger local neighbourhoods [134]. More recent models extend this by deploying irregular lattice structures, such as the quasiperiodic Penrose lattice, by analysing using tools associated with percolation problems [3, 121].

Though our models will largely focus on lattices, for completeness, we briefly discuss other approaches to agent-based infection that use networks in a more general sense. In applying the SIR framework to an agent-based simulation, one approach first popularised by Pastor-Satorras and Vespignani in the context of computer viruses involves considering a system represented by a graph. Each node on the graph represents an infectious agent, and the links between the nodes represent a potential route of infection [105]. Each node on a graph has a degree k, where k is the number of other connected nodes. The graph as a whole is referred to as a contact network [73, 96].

From the contact network, a secondary network can be formed from the path that infection travels through throughout the stochastic process, known as the transmission matrix. For networks that satisfy certain conditions, quantities like R_0 can be mathematically estimated. For example, Barabási shows that for the SIR model on a network

$$R_0 \approx \frac{\langle k^2 \rangle}{\langle k \rangle} - 1, \tag{50}$$

where $\langle k^2 \rangle$ is the second moment of the transmission network and $\langle k \rangle$ is the first moment, i.e. the mean degree of the network [10]. Depending on the network in question and the distribution of its nodes and links, this can be adapted for various structures. The structure of the network can have a significant impact on dynamics, with the possibility of local saturation where degree distribution is heterogeneous. Furthermore, for certain power-law degree distribtions, the mean degree can be infinite [10, 40].

While substantial modelling exists on lattices and networks for ecologies and infectious diseases, there is a paucity of studies combining the two. Examples exist, but their scope is limited to specific scenarios with no explicit examination of dilution,

for example, in Li and Guo's paper [83]. Nevertheless, the earlier work can inform the use of our agent-based models in analysing the dilution effect.

2.3.2 Off-Lattice Models

Spatially explicit agent-based models exist for both ecology and epidemiology. Rather than contact occurring along a pre-defined network or lattice, spatially explicit models eschew lattices and model the movement of each agent with a particular algorithm to give the desired behaviour. Given that the agents in our models are primarily animals, it makes sense to review models for animal movement before embarking on model construction.

Random walks are well-studied and provide a solid basis to underpin a model. In addition to numerous applications in many fields, correlated random walks are used in ecology to model the movements of animals [42, 126, 127]. This was originally applied to simulating diffusion but has been extended to include various distributions of movement directions and biological phenomena such as food, water and shelter needs. We employ a correlated random walk to underpin our spatially explicit model. The most basic random walk in one dimension occurs on the number line of integers $\mathbb{Z}$, where a single walker moves left or right with equal probability at each time step. As it takes place on a locally finite lattice with equal probabilities of moving in any particular direction, this is referred to as a *simple symmetric* random walk. Restricting such a walk to a finite domain is referred to as *bordered* [79, 110]. In one dimension, only two movement directions are possible. Moving to two dimensions allows infinitely many possible angles of movement, permitting the model to leave the confines of a discretised lattice, no longer being a simple random walk.

A correlated random walk is one where the direction of movement depends on the previous direction of movement [109]. This correlation can occur in several ways, the most basic of which is simply to retain the same direction of movement with a certain preset probability and randomise it otherwise. Other approaches include using a chosen distribution for the direction depending on the previous one at every time step. For example, the change in angle for the next step could be normally distributed with a median of 0 and a lower or higher variance chosen to give the desired amount of correlation. We use a landmark paper by Kareiva and Shigesada as an illustrative example [68]. They modelled insect flight with steps of a varying length and angle, with the length and angle of each flight correlated with that of the previous flight according to a general probability distribution function, and found an expression for the expected displacement after n time-steps in terms of the expected length and expected angles.

Correlated random walks feature less frequently in epidemiological research. Most epidemiological study concerns spread in human communities, which is borne out in the range of agent-based models in the literature. These models tend to favour human social structures in their implementation of movement, for example, in the RAW-ALPS model of epidemic spread in the community developed for lockdowns in the COVID-19 pandemic [129]. Models such as these employ defined social structures such as households and common gathering places like shopping venues, schools and workplaces. These are generally not appropriate for ecosystems unless adapted to the social structure of the species in question. However, such models tend to implement proximity-based disease spread, where susceptible individuals within a specific range of an infected individual have a certain probability of becoming infected themselves. This aspect of network-based approaches is applicable to ecosystems and will form a key component of our agent-based models that seek to combine animal movement with disease spread.

An implementation of a model with one species and proximity-based disease spread can be found in a 2023 article by Nguyen, though the movement in this model is a purely random walk rather than correlated [98]. Overall, agent-based modelling approaches are less explored in the field of ecoepidemiology. Our goal in the coming chapters is to address this gap in the literature, with ODE models serving as a key point of comparison.

2.4 Alternative Models and Summary

While this chapter has largely compared ODE approaches to agent-based models as they will form the primary basis of later models, this is not a strict dichotomy. There are a number of other approaches employed in the literature, each with distinct advantages and disadvantages that should be addressed for completeness.

Partial Differential Equations (PDEs) are one method of introducing spatial heterogeneity to differential equation models. They retain the deterministic nature of ODEs, but introduce spatial dynamics. In epidemiology, PDE variants of the SIR model have been used to model the spread of plagues such as the black death in 14th-century Europe [99]. One example is presented by Jones et al., who assume that where $f(x)$ is a function describing the transmission rate of suceptible to infected individuals at a distance x, which is less than a maximum possible distance δ [66]. For small values of δ, approximation yields

$$\frac{\partial S}{\partial t} = -\theta SI - \phi S\nabla^2 I \tag{51}$$

$$\frac{\partial I}{\partial t} = \theta SI + \phi S\nabla^2 I - \gamma I \tag{52}$$

$$\frac{\partial R}{\partial t} = \gamma I, \tag{53}$$

which is a reaction-diffusion system, where the parameters $\theta = \int_{-\delta}^{\delta} f(|u|)du$ and $\phi = \frac{1}{2}\int_{-\delta}^{\delta} u^2 f(|u|)du$ describe the spatiotemporal spread of infection, and ∇^2 is the Laplacian operator. The system can be further developed to account for host movement through the addition of diffusion to all compartments. Similar diffusive PDE extensions to Lotka-Volterra equations exist, in predator-prey and competitive contexts, as highlighted in a review by Cherniha and Davydovych [27].

While PDEs incorporate spatial dynamics, they are also often difficult to solve, with finite element methods for numerical solution being computationally expensive. This complexity is particularly burdensome when applying numerical methods to the ecoepidemiological systems we seek to analyse, making them impractical for our purposes.

Stochastic models can address the limitations associated with the deterministic nature of ODE models. Two classes of stochastic model of interest are **Continuous-Time Markov Chains** (CTMCs) and **Stochastic Differential Equations** (SDEs). CTMCs describe process changes for discrete random variables over continuous time [1]. The Markov property of future behaviour depending solely on the current state implies that the time between state changes is exponentially distributed, while the state change itself is governed by a stochastic matrix P. The transition probabilities can be used to produce Kolmogorov differential equations that can be solved for sufficiently simple systems, or the process can be simulated numerically using the Gillespie algorithm [2, 46].

As an illiustrative example, CTMC implementation of the SIR model treats the variables $S(t)$, $I(t)$ and $R(t)$ as discrete and non-negative over a continuous time period $t \in [0, \infty)$ [2]. The two state changes are the infection of a susceptible individual, and the recovery of an infected individual, occurring at rates $\beta SI/N$ and γI, analogous to the rates at which equivalent processes occur in the SIR ODEs. Ecological CTMCs treat the births and deaths similarly with probabilities derived from the interaction matrix in Table 2, and numerous examples exist in the literature [47, 57]. Of particular note is that the stochastic nature of the models produces different behaviour to ODE

approaches. If $R_0 > 1$, there is a probability $0 < p < 1$ of a major outbreak where more than $\sqrt{N}$ individuals are infected, and if $R_0 < 1$, $p \to 0$ as $N \to \infty$ [2, 34]. Stochasticity also allows for the possibility of species/infection extinction during the minima associated with the atto-fox problem [20].

In contrast with CTMCs, SDEs treat the random variables as continuous [2]. In relation to formulation from analogous ODEs, transitions between states in continuous-time are approximated to be the sum of many independent and identical discrete periods, the outcome of which is normally distributed as per the central limit theorem. The state transitions are therefore treated as Wiener processes, occurring at rates governed by a covariance matrix describing each process. As with CTMCs, numerical simulation produces sample paths, solved using stochastic methods such as the Euler-Maruyama method [6].

As with CTMCs, SDEs have been in use for decades in epidemiology [2, 20, 89] and ecology [9, 135], where they are well-suited to extinction-time problems. An advantage of SDEs is that they are faster to simulate than CTMCs for large populations, owing to their fixed time-step [2], though the continuous state variables can pose a limitation for low population sizes.

While SDEs and CTMCs are stochastic, like agent-based models, the addition of spatial dynamics is impractical compared to agent-based models, so we do not employ these methods in future chapters. Going forward, we consider ODEs and agent-based models exclusively. We begin by developing some novel agent-based models, before considering some ODE systems with more complex dynamics.

3 An Agent-Based Approach With Plants

Here, we seek to apply the SIRS model in an agent-based system on a lattice, where every site contains a designated rice cultivar. This premise is inspired by Zhu's work, which found a dilution effect when rice cultivars were arranged in configurations that maximised the diversity of the cultivars in each plot, minimising the impact of rice blast, a fungal infection [154]. While a dilution effect was found, a cause for the observation was not ascribed in the paper. We can attempt to use an agent-based model to explore possible causes of a dilution effect in a plant-based system with no movement. As a basis for this model, we use the SIRS framework.

In the SIRS model, susceptibles become infected through contact with infected individuals and then recover at a constant rate. Recovered individuals eventually lose immunity and become susceptible again at a different constant rate. It assumes homogeneous mixing, meaning it neglects any dynamics arising from spatial phenomena, such as inhomogeneity in the distributions of infected individuals or the different species present in a community. With this fact in mind, some deficiencies can be seen with the model, particularly concerning the agricultural context to which it is applied. In this case, we are considering two varieties of rice distributed in particular arrangements within a relatively small plot, so this approach is inappropriate. Instead, we consider an agent-based model, a computational model that considers the interactions and states of individual plants rather than the plot as a whole.

3.1 The Agent-Based Model

First, we discuss a simple SIRS model that only incorporates a single type of plant. Consider a $N \times N$ square array of sites, which we use to represent individual plants. Their statuses (susceptible, infected and recovered) are represented by a matrix $S(t)$. Each element S_{ij} describes the disease status of the plant in the ith row and jth column thusly

$$S_{i,j} = \begin{cases} 0 & \text{plant in position (i, j) is susceptible} \\ 1 & \text{plant (i, j) is infected} \\ 2 & \text{plant (i, j) is recovered.} \end{cases} \tag{54}$$

This array is updated on each iteration using two further matrices. Firstly, $P_{i,j}(t)$ describes the probability of a susceptible plant in position (i, j) becoming infected based on the status of the plants within its neighbourhood. The neighbourhood used is Von Neumann, meaning that it considers the status of the four nearest neighbours above,

below, left and right of the plant in question. Each infected neighbour is assumed to contribute equally to the total probability of a susceptible becoming infected; if no infected sites surround the site in question, there is no chance of infection, whereas if infected sites completely surround a site, it becomes infected at a rate β. Explicitly, this can be given by

$$P_{i,j}(t) = \frac{\beta}{4}\left(\delta(S_{i,j-1}(t)-1)+\delta(S_{i+1,j}(t)-1)+\delta(S_{i,j+1}(t)-1)+\delta(S_{i-1,j}(t)-1)\right) \quad (55)$$

where $\delta(x)$ is the Dirac delta function, which is defined here as

$$\delta(x) \coloneqq \begin{cases} 1 & x = 0 \\ 0 & x \neq 0. \end{cases} \quad (56)$$

The choice of additive probabilities in the entries of $P(t)$ serves to crudely reflect contributions to the probability from possible wind directions. A more elaborate system for describing wind is a potential extension of this model.

Secondly, $Q_{i,j}(t)$ describes the number of iterations for which an infected or recovered individual has remained in that state. For infected sites, once $Q_{i,j}(t)$ exceeds a certain threshold I_{max}, the site becomes recovered. Similarly, once a threshold R_{max} is exceeded for recovered sites, the site becomes susceptible. Using Q means that the infection time and duration of immunity are constant, in contrast with the SIR model's ODEs that do not track individuals and, therefore, have infected individuals recovering at a constant rate, producing an exponential distribution of infection times . In addition to its more realiastic behaviour, this approach provides a high degree of flexibility; analogues to other ODE models can be recovered through judicious choices of these thresholds. For example, setting $R_{max} = 0$ gives the SIS model, where infected sites immediately become susceptible upon recovery, or setting $R_{max} > n$ where n is the number of iterations gives the SIR model. The timer matrix elements $Q_{i,j}$ update as

$$Q_{i,j}(t+1) = \begin{cases} 0 & \text{if } Q_{i,j}(t) = I_{max} \text{ and } S_{i,j}(t) = 1 \\ 0 & \text{if } Q_{i,j}(t) = R_{max} \text{ and } S_{i,j}(t) = 2 \\ Q_{i,j}(t)+1 & \text{otherwise.} \end{cases} \quad (57)$$

Using P and Q, the matrix elements $S_{i,j}$ update as

$$S_{i,j}(t+1) = \begin{cases} 1 & \text{if } S_{i,j}(t) = 0 \wedge P_{i,j}(t) > p \\ 2 & \text{if } S_{i,j}(t) = 1 \wedge Q_{i,j}(t) = I_{max} \\ 0 & \text{if } S_{i,j}(t) = 2 \wedge Q_{i,j}(t) = R_{max} \\ S_{i,j}(t) & \text{otherwise} \end{cases} \tag{58}$$

where $0 \leq p \leq 1$ is a uniformly distributed random number generated separately for each site and $\wedge$ is the logical AND operator. Since these arrays are of a finite $N \times N$ size, boundary conditions must also be considered. One option is Dirichlet boundary conditions, where sites outside the $N \times N$ array are considered unoccupied and do not contribute to disease transmission. Another second option is periodic boundary conditions, in which the sites on the boundary are considered neighbours with their counterparts on the opposite side. For sufficiently large arrays, periodic boundary conditions approximate an infinitely large plot.

These models display behaviour that is typically analogous to their deterministic counterparts. Where the recovery time is short, and immunity persists for an extended period, the disease will likely die out, equivalent to the scenario where $R_0 < 1$ in the SIR model. Likewise, when the disease causes an outbreak, it eventually dies out. The outbreak concludes not with the complete infection of the entire population, but with some remaining uninfected, corresponding to the $R_0 > 1$ case of the SIR model. This agent-based model is stochastic, so rather than a clear threshold being present at $R_0 = 1$ dividing the cases where an outbreak does or does not occur, the early behaviour of the system can have a profound effect on the long-term dynamics of the system. This results in cases where separate realisations of the model can result in vastly different behaviour for a given set of parameters, with outbreaks occurring some but not all of the time. This contrast is a crucial difference between the deterministic differential equation and stochastic agent-based approaches.

3.2 Dilution Effect Causes

We use the agent-based model previously discussed as a platform for investigating possible mechanisms that may bring about a dilution effect. First, we must amend the equations that dictate how the system evolves to incorporate multiple varieties of "plants" occupying the various sites. We suggest two possible causes for a dilution effect; one is a direct reduction in pathogen transmission between plants of a different variety. Another is rooted in the ecological concept of complementarity; sites neighbouring sites of a different type are "fitter" and, therefore, are less vulnerable to infection.

Represent the n different varieties in a plot simply by the following $N \times N$ matrix, A, which is assumed not to change with time and to be totally occupied, though this is not necessarily the case.

$$A_{i,j} = \begin{cases} 1 & \text{plant in position (i, j) is of the first variety} \\ 2 & \text{plant (i, j) is of the second variety} \\ k : k = 1 : n & \text{plant (i, j) is of the kth variety.} \end{cases} \tag{59}$$

Using the matrix A and the same Von Neumann neighbourhood as in the single variety model, we can subsequently employ our two mechanisms that may bring about a dilution effect.

3.2.1 Pathogen Transmission Approach

The first of these mechanisms is a reduction in the rate of pathogen transmission between sites of a different type. As a starting point, in a continuous time context with n varieties akin to the SIR model, this would be represented by $3n$ differential equations of the following form:

$$\frac{dS_k}{dt} = -S_k \sum_{i=1}^{n} \frac{\beta_{k,i} I_i}{N_i} + \xi_k R_k \tag{60}$$

$$\frac{dI_k}{dt} = S_k \sum_{i=1}^{n} \frac{\beta_{k,i} I_i}{N_i} - \gamma_k I_k \tag{61}$$

$$\frac{dR_k}{dt} = \gamma_k I_k - \xi_k R_k \tag{62}$$

for $k = 1 : n$. It should be noted, of course, that these equations can be condensed to $2n$ equations by using the fact that $S_k + I_k + R_k = N_k$, where N_k is constant in the absence of births and deaths. We must also adapt the system of differential equations to account for possible asymmetry in the dynamics of the various types present in the system. Rather than assuming that the rate of transmission β is uniform, we implement it in the form of an $n \times n$ matrix, whose elements $\beta_{i,j}$ describe the rate at which a susceptible of type i becomes infected as a result of contact with an infected individual of type j. Similarly, we consider that the recovery rate γ and rate of immunity loss ξ are potentially different for each type and implement them as vectors whose elements

γ_i and ξ_i describe the relevant parameters for infected individuals of type i.

Simpler models can be recovered from this construction through judicious paramter selections. For example, by setting $\xi_k = 0 \ \forall k$ we obtain the SIR model. The SIRS construction is chosen in part because it incorporates a wide range of possible dynamics.

We now consider how to adapt our cellular automaton so that it is analogous to the above system of equations. One way in which this is done is by changing the way the infection is passed between sites at each time step; specifically, by altering the form of the probability matrix $P_{i,j}(t)$ so that the rate of transmission β is no longer uniform for each configuration of sites. Instead, it takes the form of a $n \times n$ matrix $\beta_{k,l}$, describing the transmission rate for a susceptible site of variety k from an infected site of variety l. Assuming some normalisation, i.e. that a typical rate of transmission is given by $\beta_{k,l} = 1$, then a reduction in transmission is represented by $\beta_{k,l} < 1$, and an increase in transmission is represented by $\beta_{k,l} > 1$. This flexible approach can represent a range of scenarios, including uneven transmission rates between the two varieties in isolation, i.e. $\beta_{k,k} \neq \beta_{l,l}$ if $k \neq l$.

Factoring in $\beta_{k,l}$, the new form of the matrix $P_{i,j}(t)$ describing the probability of a susceptible site (i, j) becoming infected after time step t is

$$P_{i,j}(t) = \frac{1}{4}\left(\beta_N \delta(S_{i,j-1}(t)-1)+\beta_E \delta(S_{i+1,j}(t)-1)+\beta_S \delta(S_{i,j+1}(t)-1)+\beta_W \delta(S_{i-1,j}(t)-1)\right) \tag{63}$$

where β_N, β_E, β_S, and β_W describe the transmission rates from the adjacent sites in the respective directions North, East, South, and West. Explicitly, for a site (i, j):

$$\beta_N = \beta_{A_{i,j},\, A_{i,j-1}} \tag{64}$$

$$\beta_E = \beta_{A_{i,j},\, A_{i+1,j}} \tag{65}$$

$$\beta_S = \beta_{A_{i,j},\, A_{i,j+1}} \tag{66}$$

$$\beta_W = \beta_{A_{i,j},\, A_{i-1,j}}. \tag{67}$$

As with the probabilities of transmission, the recovery and loss of immunity times for the agent-based system must also be adapted. This is simply done by noting that, when evaluating $Q_{i,j}(t + 1)$ and $P_{i,j}(t + 1)$, that I_{max} and R_{max} are now dependent on $A_{i,j}$. A visual diagram for the difference in pathogen transmission is given in Figure 8.

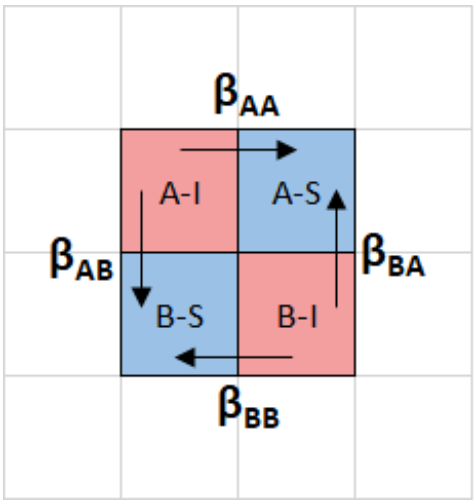

Figure 8: A transmission mechanism for the dilution effect. The infected plants in the top-left and bottom-right transmit infection to the susceptible plants in the top-right and bottom-left. They do so rates $\beta_{k,l}$, where k is the variety of the infected plant and l is the variety of the susceptible plant.

Keeping every other aspect of the model as given in the previous section, these mechanisms and procedures give the complete cellular automaton for the case of multiple varieties and a change in the transmission rate between them.

3.2.2 Ecological Approach

The second mechanism for explaining a possible dilution effect is rooted in the ecological concepts of complementarity and facilitation. Complementarity refers to a tendency for a community that is more richly (i.e. more species) or evenly (i.e. more balanced) biodiverse to be more productive, owing to differences in resource requirements between the species [44]. In agriculture, this would suggest that crops grown in a polyculture would perform better and produce a higher yield than crops grown in a monoculture, similar to the findings of Zhu et al. [154]. Facilitation is a mechanism by which a species assists another through altering its environment to make it better suited to its success. An example of this would be nitrogen-fixing in legumes facilitating other species; in one context, Dornbusch et al. found that an exotic nitrogen-fixing legume could facilitate the invasion of an exotic grass species [39]. These mechanisms are assumed to affect pathogen spread by increasing fitness within the community, conferring some resistance to disease.

Directly implementing this mechanism in a continuous time context is more difficult, as the SIR model and its derivatives assume homogeneous mixing, meaning that

interactions between species occur at rates relative to their prevalences. As a result, any reduction in disease prevalence and severity stemming from complementarity or facilitation would be reflected in the global transmission rate β and recovery rate γ. Heterogeneous mixing is a further advantage of the cellular automaton approach, as it can implement different configurations of the various species using the matrix $A_{i,j}$, and therefore easily capture any potential differences in dynamics stemming from inter-species interaction where homogeneous mixing does not apply.

We implement this mechanism in the cellular automaton by assuming that, among sites of the same type, the probability of disease spreading from one site to another is changed should the site be bordered by any number of sites of a different type. This probability is assumed to be proportionate to the number of sites of a different type around the site in question, i.e., the site receives the full effect should it be completely surrounded, or half of the effect should be half surrounded. Mathematically, the full effect is represented by a constant ϵ.We can then define a coefficient $C_{i,j}$ to be:

$$C_{i,j} = \epsilon + \frac{1}{4}(1-\epsilon)(\delta_{A_{i,j}\ A_{i,j-1}} + \delta_{A_{i,j}\ A_{i+1,j}} + \delta_{A_{i,j}\ A_{i,j+1}} + \delta_{A_{i,j}\ A_{i-1,j}}), \tag{68}$$

where δ_{ij} is the Kronecker delta function, given by

$$\delta_{ij} \coloneqq \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases} \tag{69}$$

This coefficient corresponds to the change in the probability of infection for the site as a function of the types of its neighbours. The factor in the last term represents the number of matching sites in the neighbourhood of the site (i, j). Inspecting the form of this coefficient, we see that in the case where there are no matching sites, $C_{i,j} = \epsilon$, and where there are four matching sites, $C_{i,j} = 1$, exactly as required. Using the coefficient $C_{i,j}$, then the probability of a susceptible site (i, j) becoming infected in time step t is given by:

$$P_{i,j}(t) = \frac{\beta C_{i,j}}{4}\ \big(\delta(S_{i,j-1}(t)-1)+\delta(S_{i+1,j}(t)-1)+\delta(S_{i,j+1}(t)-1)+\delta(S_{i-1,j}(t)-1)\big). \tag{70}$$

In this equation, β is the same for each variety, unlike in the previous section. As when using the pathogen transmission approach, we note that the rates of recovery and loss of immunity I_{max} and R_{max} for each site can now depend on $A_{i,j}$. In addition to having different base rates for the various types of sites, we also have the option of assuming that ecological effects result in different recovery times or immunity

durations, introducing a dependence on the surrounding entries of A. Figure 9 shows this mechanism visually.

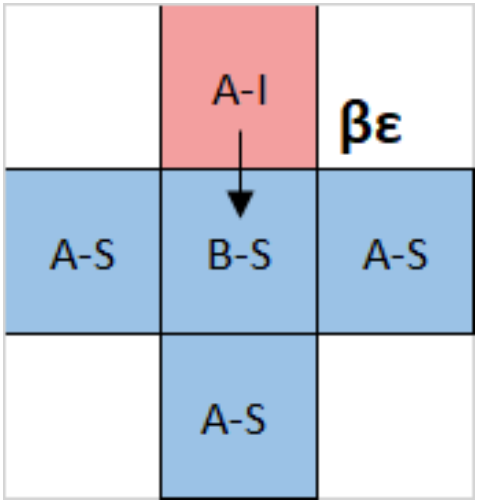

Figure 9: An ecological mechanism for the dilution effect. The susceptible plant of variety B is surrounded by the other variety A, so the probability of transmission is reduced through multiplication by ϵ, as described by Equation 68.

Having produced these two agent-based models with a mechanism implementing a dilution effect, we now explore their dynamics.

3.3 Results and Discussion

To compare the two approaches, we consider the same system for each. We use:

- A 50×50 array comprised of two types of site
- Periodic boundary conditions
- 1000 time steps
- An SIR analogue, i.e. $R_{max} > 1000$ and recovered individuals always retain immunity for the remainder of the simulation
- $I_{max} = 2$ for both types of site
- 1000 realisations

For the initial conditions, the initially infected sites are those within a radius r of the centre of the array, while the others are all initially susceptible. Mathematically:

$$S_{i,j}(0) = \begin{cases} 1 & \text{if } \sqrt{(N/2-i)^2 + (N/2-j)^2} \leq r \\ 0 & \text{otherwise} \end{cases} \tag{71}$$

To examine the dilution effect, we consider a layout of diagonal bands of each variety interspersed at various ratios, i.e.

$$A = \begin{bmatrix} 2 & 1 & 1 & 2 & 1 & \dots \\ 1 & 1 & 2 & 1 & 1 & \dots \\ 1 & 2 & 1 & 1 & 2 & \dots \\ 2 & 1 & 1 & 2 & 1 & \dots \\ 1 & 1 & 2 & 1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{72}$$

In the coming results, we refer to the ratio of these diagonal bands, e.g a ratio of 1:2 where variety 2 is planted in every third band. For infection risk, we consider the final size of the outbreak simulated (i.e. the number of recovered individuals when no infected individuals remain after a sufficient number of iterations < 1000) for variously even distributions (frequencies of the diagonal bands) and values of β and ϵ, depending on the model. For the pathogen transmission approach, we set β to be symmetrical in each site type and the transmission rate between different types to be less than or equal to the transmission rate between the same type, i.e.

$$\begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix}, \tag{73}$$

where $0 \leq b \leq 1$. For the ecological approach, we simply consider $0 \leq \epsilon \leq 1$. Using intervals of 0.02 for the parameters, these simulations show that increasing biodiversity (i.e. evenness) reduces the reduction in transmission from β and ϵ, respectively, required to make an outbreak less likely to occur, and the final size of an outbreak when they do occur. This is indeed a dilution effect, taking on a similar form despite the different mechanisms that bring it about.

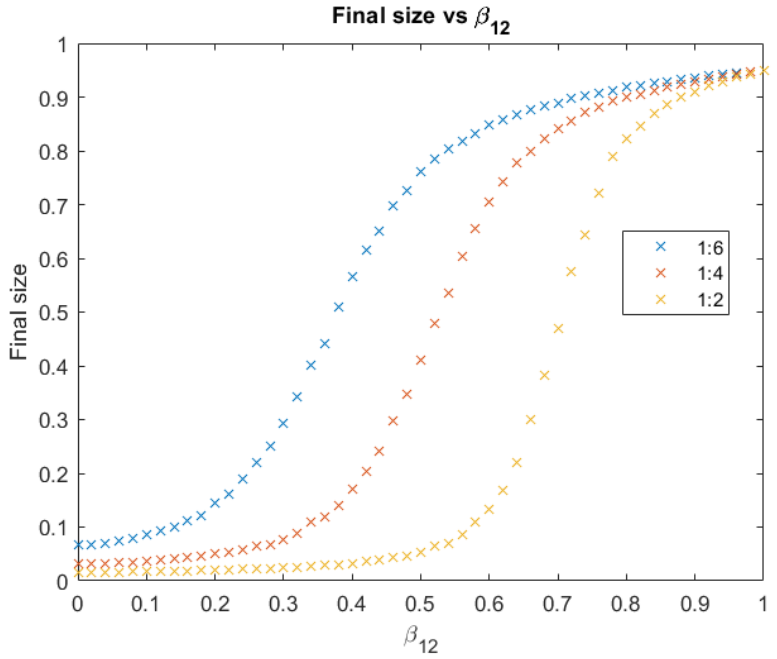

Figure 10: Final sizes vs. the cross-species transmission rate $\beta_{12} = b$ for various spacing ratios of the diagonal bands of each variety.

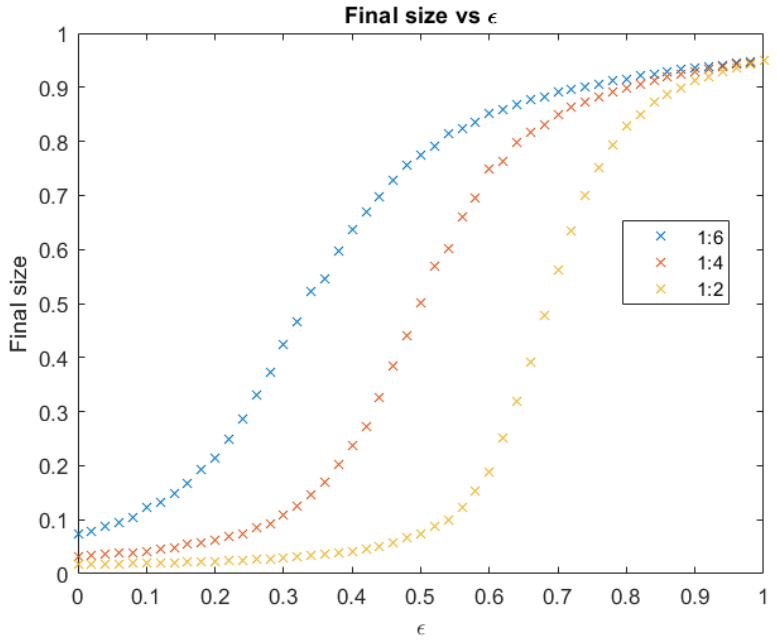

Figure 11: Final sizes vs. the reduction in transmission from fitness ϵ for various spacing ratios of the diagonal bands of each variety.

We can compare results with our continuous time system for the pathogen transmission approach and see analogous findings, allowing for expected differences between deterministic and stochastic methods (e.g. bifurcation at a critical point vs a gradual shift in outbreak probability) [34]. This comparison suggests that a dilution effect can be observed on various scales, from an agent-based setting describing a small community to a continuous time setting describing a big community.

We also note that these models can be adapted to show an amplification effect just as easily. Cases where cross-type transmission is more likely ($b > 1$) would produce an amplification effect, as would cases of ecological competition, where the presence of one type inhibits the ability of the other to fight off disease by some mechanism, like a reduction in fitness. ($\epsilon > 1$). This possibility suggests that the presence of a dilution/amplification effect is dependent on the system in question; different species interact differently, and with many ecological and epidemiological factors at play, it is clear that the dilution effect is not a universal truth but rather, a property of specific ecosystems under certain circumstances.

In terms of the significance of these results to the wider dilution effect debate, while it is impossible to comment on generality from one example, it is clear that dilution is a consequence of numerous interactions rather than its own phenomenon. This result is an expected one; that dilution can arise through many mechanisms is widely agreed upon whether the authors argue for [28, 104] or against [81, 119] a generalised dilution effect, and in specific case studies [75]. In the next chapter, we extend the lattice model to include interspecies interactions in animals to consider the dilution effect in a two-species predator-prey system, before discussing the advantages and disadvantages of the approach compared to other models.

4 Animals and Demography on a Lattice

We can use the model for disease transmission in plants we discussed in the previous section as a basis for a similarly agent-based model for animals. We accomplish this by making several changes, relaxing certain assumptions, and changing certain mechanisms to be more appropriate in animal ecology. As will become apparent in this section, this approach allows for creating a complex, flexible model that can be readily adapted to account for various ecological and epidemiological phenomena to a greater extent than a typical system of differential equations.

A key difference between the agricultural context of the previous section and the animal models of this section is the presence of births and deaths among animals. As such, the continuous-time SIR model to which we compare must be adjusted, though we retain the use of the lattice as an underpinning structure.

4.1 A Predator-Prey Model

To provide a basis for our epidemiological model, we begin with a predator-prey automaton based on that established by Arashiro and Tomé [5], retaining the 2D lattice structure of the plant-based model. In this model, our matrix A has three possible states, which we denote X, Y and Z. A site occupied by prey is denoted by the character X, while a site occupied by a predator is denoted Y, and Z denotes an empty site. A more detailed discussion of their model is found in Section 2.3.1.

To adequately model a predator-prey system, three processes must be accounted for. Firstly, predators must consume prey, with the availability of food corresponding to increased predator fecundity. Secondly, prey must reproduce with a positive growth rate in order to survive while being predated. Thirdly, predators must die due to having a negative population growth rate in the absence of prey. These processes should occur stochastically depending on the von Neumann neighbourhood of the site in question. In particular, we seek to have a predation rate that increases with the number of surrounding prey, with a similar effect in place for prey reproduction. For this purpose we define three probability parameters, $p_{X\mapsto Y}$, $p_{Y\mapsto Z}$ and $p_{Z\mapsto X}$, each of which is on the interval [0, 1]. An important distinction is that the second of these probabilities is the chance of a predator dying at any given time-step, while the first and last refer to the *maximum* probability of prey and predator reproduction.

We define the probability of a site i, j changing occupation states from state U to state V by $P_{i,j}^{U\mapsto V}$. Then, for a given time step, the probability of each state change

occurring is given by

$$P_{i,j}^{X\mapsto Y} = \frac{1}{4} p_{X\mapsto Y} \left(\delta_{Y,A_{i+1,j}} + \delta_{Y,A_{i,j+1}} + \delta_{Y,A_{i-1,j}} + \delta_{Y,A_{i,j-1}}\right) \tag{74}$$

$$P_{i,j}^{Y\mapsto Z} = p_{Y\mapsto Z} \tag{75}$$

$$P_{i,j}^{Z\mapsto X} = \frac{1}{4} p_{Z\mapsto X} \left(\delta_{X,A_{i+1,j}} + \delta_{X,A_{i,j+1}} + \delta_{X,A_{i-1,j}} + \delta_{X,A_{i,j-1}}\right), \tag{76}$$

where δ_{xy} is the Kronecker delta function. We use periodic boundary conditions for the system, so the subscripts should be taken with modulo N arithmetic, adjusted for matrix entries beginning at 1 rather than 0. The factor of $1/4$ in the first and third equations is implemented to normalise the probabilities to account for the total number of neighbours in the von Neumann neighbourhood. At every time step, the sites of the occupation matrix A are updated according to the following rules:

$$X \mapsto Y \quad \text{if } p < P^{X\mapsto Y} \tag{77}$$

$$Y \mapsto Z \quad \text{if } p < P^{Y\mapsto Z} \tag{78}$$

$$Z \mapsto X \quad \text{if } p < P^{Z\mapsto X}, \tag{79}$$

where $0 \leq p \leq 1$ is a uniformly distributed random number, independently computed for each time. Examples of the calculation of each probability in practice for sites X, Y, and Z are shown in Figure 12.

The system is Markovian and updates synchronously, meaning that at each time step, a matrix $A(t+1)$ is computed using only the values of the prior matrix $A(t)$. The automaton is akin to the Lotka-Volterra predator-prey models. This model is very similar to the model of Arashiro and Tomé [5], who opted to impose a limit on the parameters that $p_{X\mapsto Y} + p_{Y\mapsto Z} + p_{Z\mapsto X} = 1$, which allows for the elimination of one parameter by effectively absorbing it into the time-scale of the simulation. We must relax such assumptions, as we will seek to further this model by adding an infection in the population, adding more parameters and making the assumption unjustified.

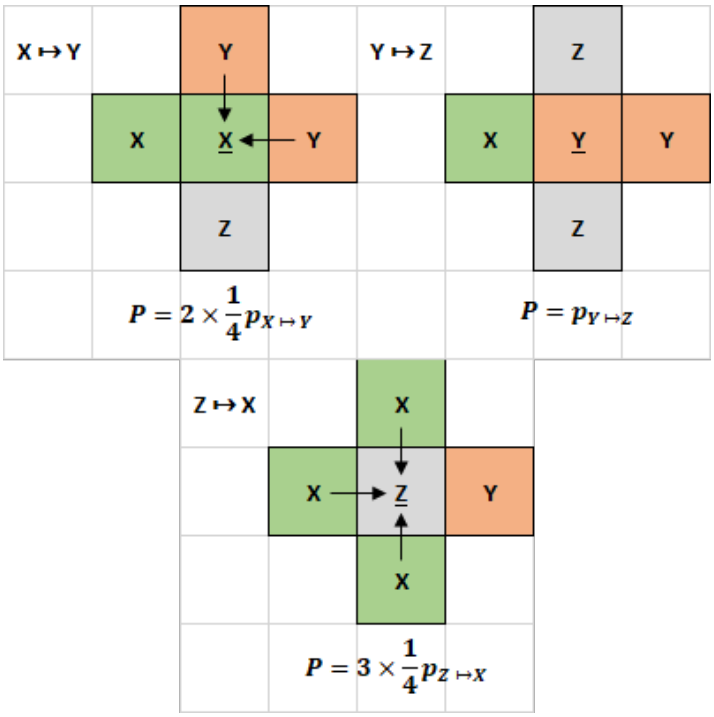

Figure 12: Probabilities P for the underlined prey (X), predator (Y), and empty (Z) site, highlighting their relationships to the configuration of the neighbouring cells.

We can advance this model by adding a further parameter, if desired, to prevent prey from occupying every cell in the absence of predators. This improves the realism of the model in that case, as it is equivalent to having a logistic intra-species competition term in a Lotka-Volterra ODE system. However, including two prey transitions would render the model asynchronous without judicious selection of parameters, and logistic prey death is largely unnecessary where predators reduce the prey population at a much faster rate than intraspecific competition. As such, the total occupancy of A by prey is considered to be a proxy for the predator-free equilibrium.

4.2 Adding Infection

Having detailed the predator-prey dynamics, it now stands to introduce infection to the model. As with the plant-based model, we describe the infectious state of each site using a separate infection matrix, denoted $\mathcal{S}_{i,j}$, and employ an SIRS framework. The elements of $\mathcal{S}$ are S, I, R, and E, the first three corresponding to the susceptible, infected and recovered states and the last being the null state where the site is empty. We seek to implement an infection that spreads through contact among prey and from prey to predator through the consumption of infected prey. A zoonotic infection transmitted

this way is bovine tuberculosis caused by the *Mycobacterium bovis* bacterium, which circulates among white-tailed deer (*Odocoileus virginianus*) in the wild, and spreads to cougars (*Canis latrans*) via predation. This fact allows for the use of coyotes as sentinels for monitoring prevalence in deer, and thus the risk of zoonotic transmission to deer hunters [7].

Three processes are associated exclusively with the infection matrix, without any change in the occupation matrix. These are the production of newly infected prey from adjacent infected prey, the recovery of infected prey and predators, and the loss of immunity among recovered prey and predators. The first of these occurs at a maximum probability $p_{S_X \mapsto I}$ based on neighbouring cells, while the other two occur at probabilities $p_{I \mapsto R}$ and $p_{R \mapsto S}$. The rates of recovery and loss of immunity are assumed to be equal for each species.

$$P_{i,j}^{S_X \mapsto I} = \frac{1}{4} p_{S_X \mapsto I} \left(\delta_{X,A_{i+1,j}} \delta_{I,S_{i+1,j}} + \delta_{X,A_{i,j+1}} \delta_{I,S_{i,j+1}} + \delta_{X,A_{i-1,j}} \delta_{I,S_{i-1,j}} + \delta_{X,A_{i,j-1}} \delta_{I,S_{i,j-1}} \right) \tag{80}$$

$$P_{i,j}^{I \mapsto R} = p_{I \mapsto R} \tag{81}$$

$$P_{i,j}^{R \mapsto S} = p_{R \mapsto S}. \tag{82}$$

The product of two Kronecker delta functions in the expression for $P_{S_x \mapsto I}$ are designed to return 1 for only cases where the neighbour is both infected and prey and 0 otherwise. The X subscript specifies that this probability applies only to prey, the X state. The equation for $P_{S_x \mapsto I}$ is similar to Equation 63, where the interspecies transmission rate is set to 0.

Additionally, there are three additional state changes in $\mathcal{S}$ associated with changes to the occupation matrix A. Whenever a predator dies and a cell moves to the empty state Z, the corresponding infection matrix entry becomes E. Whenever a new prey is born, the infection matrix entry also changes from E to S, reflecting an assumption of no vertical transmission in prey. Additionally, when a predator is born, there is a probability of infection in the predator based on the number of surrounding infected prey, denoted $P_{S_Y \mapsto I}$, and given by

$$P_{i,j}^{S_Y \mapsto I} = \frac{1}{4} p_{S_Y \mapsto I} \left(\delta_{X,A_{i+1,j}} \delta_{I,S_{i+1,j}} + \delta_{X,A_{i,j+1}} \delta_{I,S_{i,j+1}} + \delta_{X,A_{i-1,j}} \delta_{I,S_{i-1,j}} + \delta_{X,A_{i,j-1}} \delta_{I,S_{i,j-1}} \right).$$

Otherwise, the demographic processes of the original predator-prey model are unchanged, equivalent to assuming no differences in birth and death probabilities between susceptible, infected and recovered animals. The entire model operates

synchronously, with any demographic change processes occurring before infection changes. The code for the model is provided in the Appendices for review if needed.

Throughout the simulation, we track the numbers of predators and prey and the susceptible, infected and recovered populations at any point in time using a three-dimensional array. Parameter choices can depend on the infectious disease in question, but in general, the parameter space where demography probabilities are significantly lower than the infection probabilities is of most interest. That choice is justified given that the typical duration of infection is less than a lifespan and the expected durations of infection and lifespan are equal to the reciprocal of the recovery and death rates for predator and prey. An expected prey lifespan is more difficult to calculate and depends on predator density and distribution patterns.

The model is versatile, with the SIRS implementation able to be reduced to SIR and SI models through judicious choices in parameters. The SIS model can also be made with slight adjustments, sending individuals to the S state upon recovery instead of the R state. Its steady-state behaviour for the SIRS case can be qualitatively observed to include endemic and non-endemic states for the prey absorbing state and the state where predator and prey coexist, as shown in Figure 13.

As can be observed in the bottom row of Figure 13, there is substantial spatial heterogeneity in the distribution of infections, in contrast to the assumption of homogeneous mixing inherent in ODE models. There is a certain amount of stochastic variation in both infection status and species populations once the system equilibrates, driven both by the stochastic randomness of the model and the travelling fronts formed by both each species and infection.

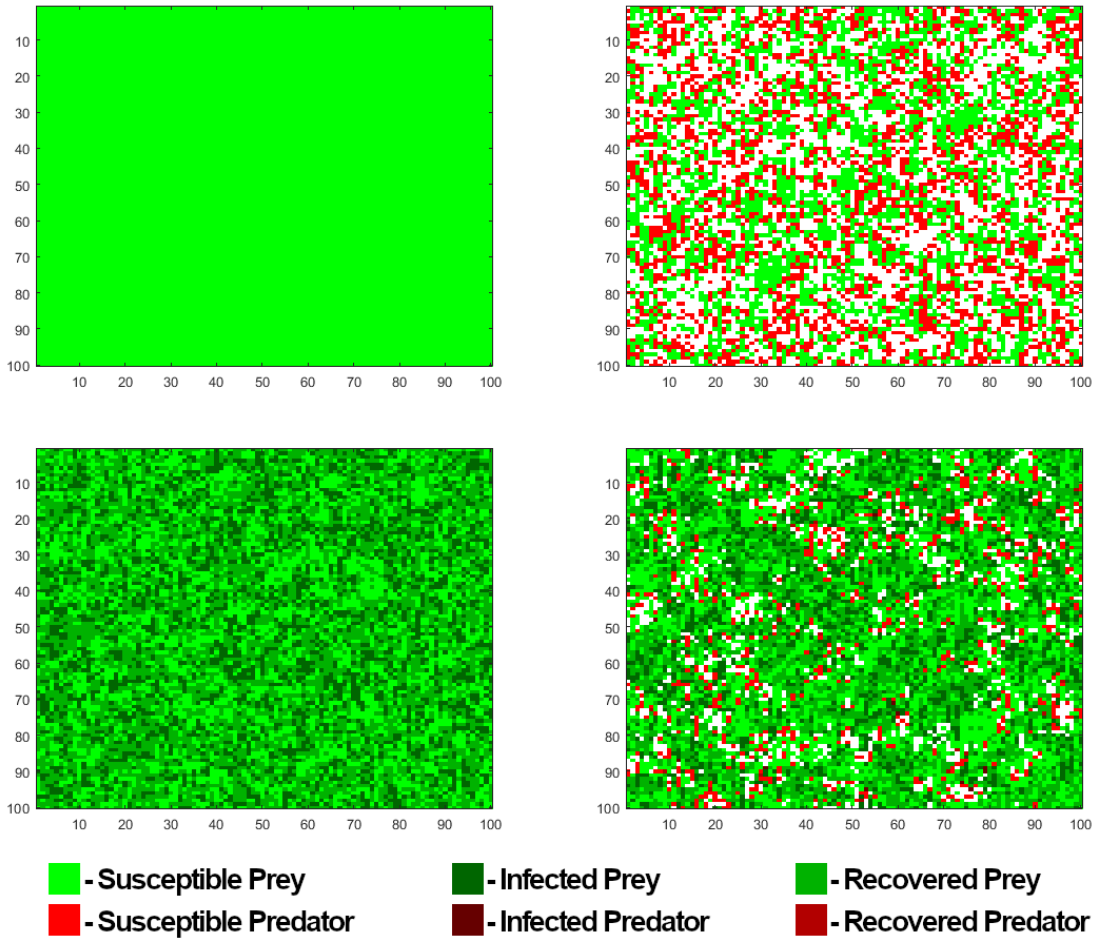

Figure 13: The long-term dynamics of the predator-prey infection agent-based model with $N = 100$, meaning there are a total of 10000 sites. Without logistic effects or a random death probability, prey do not die in the absence of predators and are constrained only by space. Clockwise from top-left, these simulations show the non-endemic prey-only, non-endemic coexistence, endemic coexistence, and endemic prey-only steady states.

4.3 Relationship Between Biodiversity and Infection

With only two species and easily accessible infection data from the construction of the model, discussing the dilution effect for the model is relatively simple but enlightening. We can narrow down the parameter space through certain restrictions. In a typical food web, the feeding links between predator and prey involve a larger predator consuming smaller prey around 90% of the time, when mean body sizes are measured by either weight or length [31]. Likewise, body size is generally negatively correlated with

population density [63]. While it is not true in all situations, we are most interested in the case where prey outnumber predators, and that should inform our choice of demography parameters. Since the typical infection lasts less than the total length of one life in the case of the SIRS model, which allows agents to be reinfected, we also require that the demographic processes take place on a significantly slower time scale than the epidemiological processes.

These place a significant restriction on the appropriate parameter values. An example of a system that gives the desired behaviour is shown in Table 3.

Parameter	Value	Description
$p_{X \mapsto Y}$	0.08	Predator consumes prey, reproduces
$p_{Y \mapsto Z}$	0.04	Predator dies
$p_{Z \mapsto X}$	0.06	Prey reproduces
$p_{S_{X \mapsto I}}$	0.5	Prey infect prey
$p_{S_{Y \mapsto I}}$	1	Prey infect predator
$p_{I_{\mapsto R}}$	0.1	Infected recover
$p_{R_{\mapsto S}}$	0.05	Recovered lose protection

Table 3: Parameter values for the predator-prey SIRS model on a lattice

An example simulation with $N = 100$ over 1000 time-steps is shown in Figure 14. The simulation initially sees a rise in prey, followed by a rise in infections among both species before remaining somewhat stable at equilibrium after $t = 200$.

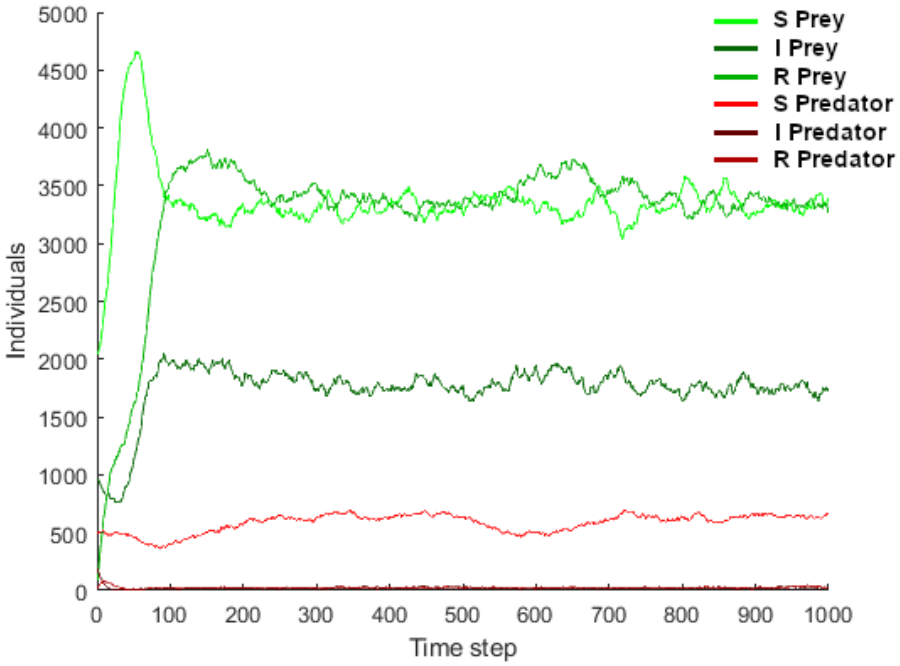

Figure 14: One simulation of the predator-prey cellular automaton. There is a low but non-zero number of infected predators for all time steps. Initially, 30% of cells contain prey, and 10% contain predators, with the rest empty. Of the initially occupied cells, 30% are infected; the rest are susceptible.

Since the agent-based model is a stochastic process, there is a statistical uncertainty associated with both the final equilibrium states arising from random noise and the simulation itself. Uncertainty requires the use of statistical techniques in our analysis. To calculate confidence intervals, we must first establish whether the data points for an individual simulation at equilibrium are normally distributed. The alternative to this is to calculate the mean over many simulations, a time-consuming process though a normal distribution would be assured by the central limit theorem. To check the data distribution, we administer the Shapiro-Wilk test and confirm this graphically. The test statistic W for the Shapiro-Wilk test on a sample $x_1, \ldots, x_n$ is given by

$$W = \frac{\left(\sum_{i=1}^{k} a_i x_{(i)}^2\right)^2}{\sum_{i=1}^{k} (x_i - \bar{x})^2}, \tag{83}$$

where $\bar{x}$ is the mean of the sample and $x(i)$ is the i-th lowest value in the sample [124]. The coefficients a_i relate to the properties of the normal distribution. Examining Figure 14, we see that it is stable after approximately 200 time-steps, so our sample for the

Shapiro-Wilk test is the number of each species in each state.

Performing the test on the number of susceptible prey for time-steps $t \geq 200$ gives $W = 0.9911$, corresponding to a p-value of $9.3 * 10^{-5}$. This result indicates a statistically significant deviation from the normal distribution. However, calculating the effect size from a quantile-quantile (Q-Q) plot reveals that the deviation size is effectively minimal, meaning that we can treat the data as normally distributed for practical purposes. This is the expected behaviour for a very large sample size, so we do not need to aggregate the results of many simulations in our analysis. We can perform the same process for each of the other combinations and states, though we omit the full data for brevity. Notably, the proportions of infected and recovered predators are also normally distributed despite their low number.

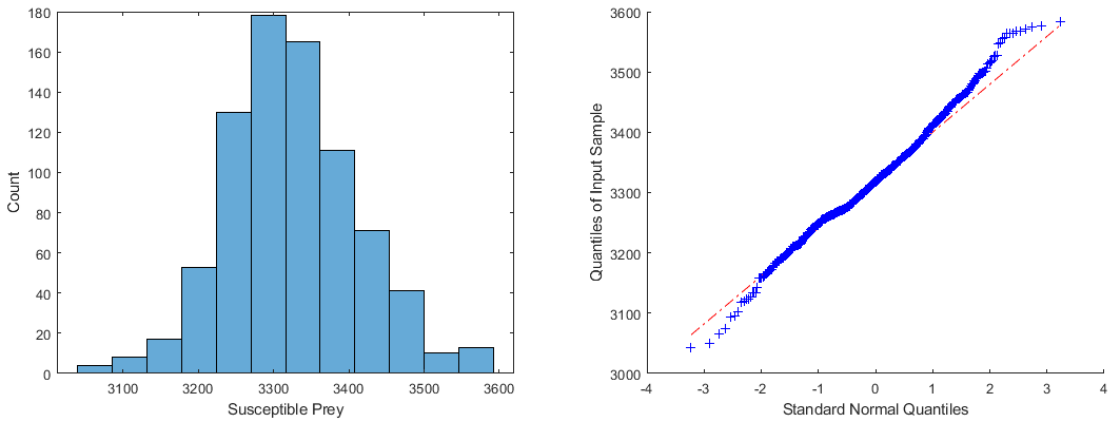

Figure 15: A histogram and Q-Q plot for susceptible prey. A skew is apparent, but the distribution is close to normal.

Having established the validity of statistical methods, we can now alter parameters and use the resulting changes in biodiversity and infection risk to observe a dilution effect for the system. Selecting a measure of biodiversity is simple, as the Simpson and Shannon indices will exhibit similar behaviour for the two-species case in that they will both have critical points where predator and prey populations are equal. We opt for the Gini-Simpson index for simplicity and to avoid the redundant use of multiple metrics to achieve the same result.

Selecting a measure of infection risk is a more complicated proposition and should be informed by context. If we are concerned with infection in the system overall, we can consider prevalence data among all hosts or an individual target host. Similarly,

we can count the total number of infected hosts, both overall and in a target host. A host might be targeted on the basis of its contact with humans, where it might be a potential cause of a spillover event that precipitates the emergence of a novel infectious disease. Prevalence data might be used where the contact is frequency-dependent, while total infected data might be used where contact and transmission are thought to be density-dependent. For thoroughness, we used several different measures of infection risk in our analysis and compared the results for each.

In our model, we retain the use of the parameter values in Table 3, with the exception of the predator death probability, $p_{Y\mapsto Z}$. We run simulations with values of $p_{Y\mapsto Z}$ on evenly spaced intervals of 0.001 between the values 0.02 and 0.06. For each simulation, we use 10000 sites and simulate for 2000 time steps to allow sufficient time for the simulation to reach equilibrium. We calculate seven quantities over the last 1000 time steps, six measures of infection risk, and the Gini-Simpson index $1 - \lambda$. The measures of infection risk encompass three infection counts and their associated prevalences. Four of these are the total infected population and prevalence of each species I_X, P_X, I_Y and P_Y. The remaining two are the counts for the entire system $I_X + I_Y$ and $(I_X + I_Y)/(N_X + N_Y)$. We also compute the 95% confidence interval for each quantity to ensure measurement accuracy.

The range for the parameter $p_{Y\mapsto X}$ encompasses three separate equilibria. At the lower extreme of the range, the higher numbers of the predator host result in the infection dying out due to the absence of predator-to-predator transmission, while at the higher end, the predators die off at a sufficiently fast rate such that the system reaches the prey-absorbing state. In between, the predator and prey coexist while the infection remains endemic. Figure 16 illustrates this, with the infection dying out below $p_{Y\mapsto Z} \approx 0.025$, and the predator being absent above $p_{Y\mapsto X} \approx 0.048$.

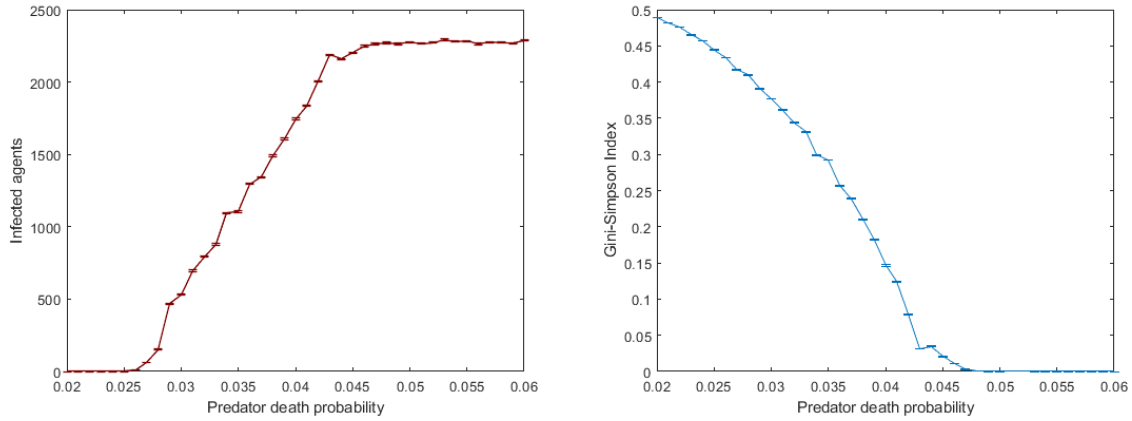

Figure 16: The values for both the number of infectious agents (left) and the Gini-Simpson biodiversity index (right) for different predator death probabilities. Error bars show a 95% confidence interval for each data point, demonstrating largely consistent equilibrium behaviour for the simulations.

Further examination to uncover a closer estimate of the critical values is possible but not relevant beyond a qualitative observation that an increase in biodiversity through the introduction of the predator can result in the instability of the endemic equilibrium, constituting a dilution effect. An increase or decrease in predator death probability can correspond to a real-world culling program, a ban on hunting, or a similar control strategy.

Examining other metrics more closely can reveal a more complex picture of the relationship between infection risk and biodiversity. Importantly, Figure 16 shows that biodiversity decreases with an increase in the predator death rate, and given that all quantities involved are strictly positive, this means that a dilution effect occurs where the measure of infection risk decreases, and an amplification effect occurs where it increases.

For both prey infection and prevalence and overall infection and prevalence, we see an increase in the infection risk with the predator death probability, and hence a dilution effect. However, if the target host is the predator, we see different behaviour. The prevalence of infection decreases with the predator death probability, but the competing effects of increasing the predator population while adding less competent hosts mean that the number of infected predators increases with biodiversity for higher values of the death probability, as shown in Figure 17. In a real-world setting, this transition between amplification and dilution would require careful consideration in guiding control strategies.

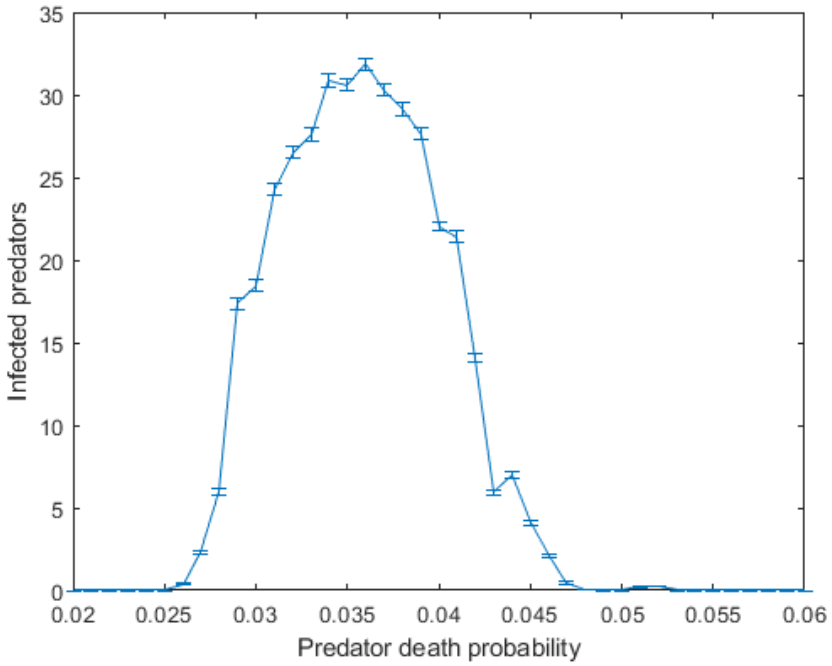

Figure 17: The number of infected predators at equilibrium with respect to predator death probability. The curve is less smooth than expected, indicative of stochastic variation on an even longer time scale.

While examining the slope of each curve can be used to make a binary determination of dilution or amplification, the dilution effect can be quantified using Equation 3. Given the level of variance observed and the error associated with the data beyond just the noise for one simulation, numerical differentiation using forward, central or backward difference equations is impractical. Instead, it is wise to employ interpolation, approximating a curve using splines and then differentiating analytically. Splines are preferable to polynomial interpolation, as they avoid Runge's phenomenon while maintaining continuity in the first and second derivatives of the piecewise interpolating functions, ensuring they are differentiable.

Our data are noisy owing to the use of only one simulation at each parameter value. We can employ smoothing before interpolation, using the surrounding simulations with parameters that differ only slightly to filter out the noise instead of many more computationally expensive simulations. For this, we can use MATLAB's smoothdata function with an appropriate choice of method described in the software documentation. We choose quadratic interpolation ("rloess" in MATLAB). Where data are approximately constant (and ill-suited for polynomial interpolation), we use a moving mean. We subsequently use the spline function to obtain a differentiable

piecewise polynomial expression and substitute the splines and their derivatives into the following equation:

$$\mathcal{D}_{I_Y} = \frac{1-\lambda}{I_Y} \frac{dI_Y}{dp_{Y\mapsto Z}} \frac{dp_{Y\mapsto Z}}{d(1-\lambda)}. \tag{84}$$

This process produces the graph seen in Figure 18; notably, using cubic splines ensures continuity in the elasticity. In the region where the predator is extinct, the Gini-Simpson index is constant, and the elasticity produces erratic behaviour since the effect of noise is dominant; over the course of the interpolation, it is likely that a random event that varies one of the means even slightly bleeds into surrounding data points. Additionally, when there are no infected predators, the expression for elasticity becomes undefined as $I_Y = 0$. We hence define that the elasticity outside of the range where both predator and infection persist to be 0.

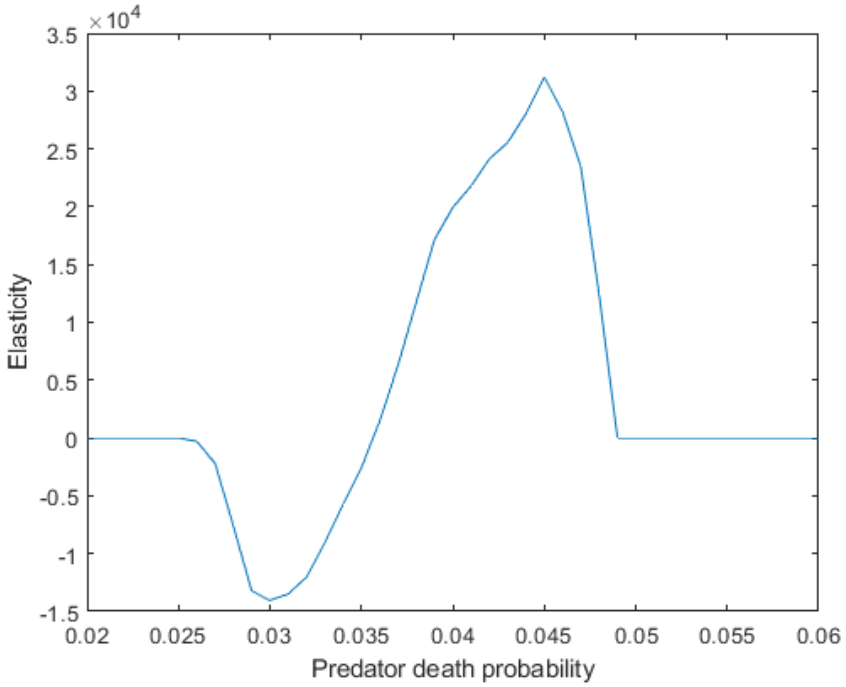

Figure 18: Elasticity of infected predators with respect to the Gini-Simpson index of the ecosystem. A dilution effect occurs below $p_{Y\mapsto Z} \approx 0.036$, and an amplification effect occurs above that threshold.

The predator-prey agent-based model with infection demonstrates the need for diligence in selecting a measure for infection risk and how numerical techniques can be applied to analyse the behaviour of a stochastic ecoepidemiological system. The sensitivity to measure choice and parameter space supports the arguments made by

Salkeld et al. in their meta-analysis, highlighting the heterogeneous relationship between host ecology and infection risk, depending on the pathogen and ecosystem in question [119]. Modelling from Roberts and Heesterbeek describing a two-species ODE system also found different effects depending on the measure of infection risk chosen [116].

The different shapes of the curves in Figures 16 and 17 have relevance to one argument made by Lafferty and Wood, when discussing the asymptotic extremes of zero biodiversity. They posit that the necessary condition of zero biodiversity coinciding with zero infection risk must mean that for dilution to occur, there must exist at least one critical point where amplification transitions to dilution [81, 149]. Such a critical point can be observed in Figure 18. It is important to note that a Gini-Simpson index of $1 - \lambda = 0$ where the species associated with spillover is extinct is a different concept of zero biodiversity, but the underlying principle remains the same. Applying this idea, our result suggests that introducting a predator may result in amplification, but that increasing predator population in an otherwise similar ecosystem where the predator is already established may give rise to a dilution effect.

4.4 Model Discussion

While we have only examined a predator-prey ecology with our lattice-based model, the approach is flexible, and other ecologies can be used to create a cellular automaton. For example, it would be possible to use the von Neumann framework introduced here to model competition. A two-species competition model would use a quadratic probability for the neighbourhood birth rate while maintaining a constant death rate. The advantage of this approach would be the ability to have different spatial-temporal dynamics for deaths and births, which is impossible under the assumption of homogeneous mixing in the standard ODE model. Depending on the desired behaviour, logistic effects can be assigned to births or deaths, as either a suppression of birth rate or increased deaths through overcrowding.

An advantage to logistic effects in this context is that the reproductive rate varies based on agent density, meaning there is potential for an Alee effect, where the birth rate is depressed compared to what logistic growth predicts, which can be either strong (resulting in local stability of extinction) or weak [41]. A lattice-based model could be used to investigate the interaction between infection dynamics that affect demography and the observation of an Alee effect.

Our implementation of the model featured only one competent host and a secondary host infected by consuming the other. It is easily possible to amend the transmission mechanism to introduce different infection dynamics, such as different contributions to the infection probability from each species, vertical transmission, or incorporating additional infectious states, such as E, for exposed but not yet infectious individuals. A further extension to the model is changing the choice of neighbourhood and the boundary conditions to reflect the mixing of the species or the physical properties of the ecosystem.

A limitation of the model is its approach to mixing. An infectious agent is confined to its neighbourhood for the entirety of its lifespan and does not move, posing a challenge for modelling disease outbreaks on a short time scale. The model is most suited to longer time scales since it incorporates demography effectively. Attempting to add movement is possible if desired; a possible implementation is provided in the appendices.

5 Going Off-Grid: A Spatially Explicit Model

Introducing movement also paves the way for agent-based models that abandon the grid-based structure of the models we have already discussed. We refer to this class of models as spatially explicit agent-based models. These models independently track the location and velocity of each animal, which confers them several advantages. With states no longer discretised, occupancy of sites no longer constrains movement, allowing for more elaborate behaviour to be implemented in exchange for more computational complexity in processing interactions.

5.1 A Single-Species Model

A single-species basic model must be established before incorporating multiple species or other systems. Working in two dimensions, we use a correlated symmetric bordering random walk for this purpose. We take a given step-size h and use a square box of side length L as our domain; one can be eliminated through an appropriate rescaling if desired. This choice means the x and y coordinates will always be bound by $0 \leq x, y \leq L$. With a given number of initial agents, we randomly assign their initial positions and angles corresponding to their movement and store them in two arrays. The rows of the matrix X, $X_i = [x_i, y_i]$ store the horizontal and vertical position of agent i, while the same agent's direction of movement is stored in the entries of the column matrix θ.

We then iterate a loop to determine the movements of each agent. At the beginning of each time step, an $n \times 2$ matrix P of random numbers $0 < p_{i,j} < 1$ is generated. The positions of the agents are altered in a random order in each iteration. Their direction of movement is maintained between time steps with a predetermined correlation probability c; should $p_{i,1} > c$, a new angle is chosen with a uniform, random distribution. This is done using $p_{i,2}$; specifically $\theta_i = 2\pi p_{i,2}$. With this complete, we update the locations of each agent according to the following:

$$x_{new} = x_{old} + h\cos(\theta) \tag{85}$$

$$y_{new} = y_{old} + h\sin(\theta). \tag{86}$$

Since h is fixed, animals move at constant speed.The movement portion of the time step is complete once this is done for each animal. Of course, with the above formula, it is easily possible for animals to escape the confines of our box. In contrast to the lattice-based approach, we choose to use reflecting boundary conditions, which are equivalent

to Neumann boundary conditions in an ODE context, meaning that no animals should pass through the boundaries of the box. Physically, this could model a body of water or an island that confines the animals to that ecosystem. We deal with this by concluding the movement portion of an iteration with the following:

$$X'_{ij} = \frac{N}{\pi} \arccos(\cos(\frac{\pi X_{ij}}{N})). \tag{87}$$

This equation conveniently adjusts any out-of-bounds positions to the correct location within the box; for any changed locations, we also reflect the angle θ, depending on the relevant boundary or boundaries. This formula can relatively easily be adjusted for removal boundary conditions (corresponding to Dirichlet conditions) and periodic boundary conditions if needed; these have the advantage of not requiring any angle shifts.

5.2 Adding Infection

There are many possible infection implementations in a spatially explicit model, as describing positions in continuous space enables a more detailed approach to transmission. We discuss two possible methods in this section. The first is the traditional contact-based transmission, while the second is the environmental transmission introduced in the modelling section.

5.2.1 Contact-Based Transmission

Once the movement is complete for a particular iteration, we deal with any changes in infection status synchronously. We retain the matrix elements S_i to denote the infection status of individual i, with 0, 1, and 2 corresponding to susceptible, infected and recovered. In place of the probabilistic approach to recovery and loss of immunity used for the lattice-based model, we can implement a timed approach. For our spatially explicit model, a timer matrix counts the number of iterations an animal remains in those states and advances them to the recovered or susceptible states after I_{max} and R_{max} iterations, respectively. Using a timer this way forces all infected individuals to remain infected for the mean duration, as opposed to the exponential distribution created by having a set probability of recovery at every time step. Neither of these choices is an entirely realistic depiction of infection, but the timed approach is a closer fit for many infections. A probability distribution function of recovery times can be specified if a probabilistic approach is desired.

Since we are not working on a lattice in the spatially explicit model, animals do not have neighbours prescribed by an array, so we must change our mechanism for infection. We do this by having interactions occur whenever a susceptible animal and one or more infected animals are within a certain radius, r. As such, generating a matrix with entries d_{ij} representing the distance between animals i and j is prudent. Such a matrix will always be symmetric and have zeroes along its main diagonal. The elements of this matrix are given by

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \tag{88}$$

recalling that x and y are the columns of the position matrix X. This distance matrix will prove useful in the future, but we also need to account for infection status; after all, we are only interested in the distance between pairs of infected and susceptible agents. We do this by setting the distances to zero for any element where the row corresponds to a non-infected animal. Zeroing such rows also has the benefit of removing any duplicate entries corresponding to distances between non-infected and infected animals, which arise from the symmetry of d. We do this by taking the following element-wise Hadamard product (symbolised by $\otimes$), which we denote C:

$$C = d \otimes (S_{mod\,2}[1 \quad 1 \quad \dots \quad 1 \quad 1]), \tag{89}$$

where the subscript of the infection matrix S denotes that the entries are each taken in modulo 2 arithmetic. In our context, this has the effect of setting 2 (recovered) to 0. Since S is a column matrix, we multiply it by an $n \times 1$ row matrix to produce an $n \times n$ matrix with rows of 1 for infected animals and 0 for all others. We then take the Hadamard product of this matrix with d. Having found the relevant distances and sorted by infection status, we can now test for any infections. We retain the use of β as the probability of infection for a particular interaction between a susceptible and an infected individual; we choose a constant β and to only have interactions occur within a certain radius, though it is possible to have β be an exponential or polynomial function if desired as well, for as long as they satisfy that $\beta(r) \to 0$ for large distances r from the susceptible.

With our infection mechanism determined, we have a complete picture of how to update the infection status for each time step. For any susceptible animal k, we count the number of entries in C satisfying $0 < C_{i,k} \leq r$. This count is the number of potential sources of infection for the susceptible animal, which we will denote as n, so the probability of remaining susceptible is therefore given by $(1 - \beta)^n$. We generate a uniformly distributed random number $0 < q < 1$, and if $q > (1 - \beta)^n$, then they

become infected, i.e. $S_i = 1$. An inherent assumption in the process is that the outcomes of separate contacts with each infected agent are mutually independent for a given susceptible agent. That assumption is generally untrue, as infectiousness and immunocompetence vary in a population; we consider such effects negligible under our assumption.

We use the timer-based mechanism for recovery and loss of immunity, so for infected and recovered animals, in each time step, we add one to the relevant element of the timer matrix and update the infection status if they exceed I_{max} or R_{max} by changing the relevant entry of S.

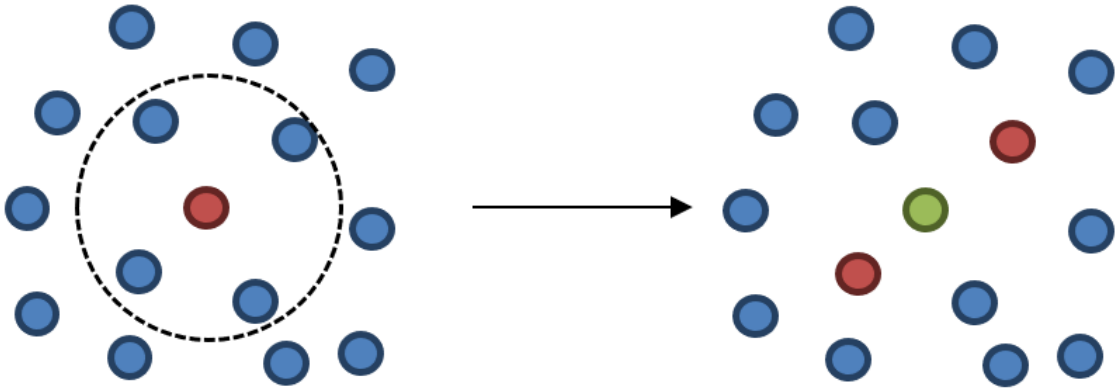

Figure 19: A visualisation of an infection state update, where the infected agent is about to recover, and can transmit infection within a radius r. A blue circle is susceptible, a red circle is infected, and a green circle is recovered.

After updating the infection status for each animal as in Figure 19, the time step is complete. After a prescribed number of time steps, the simulation concludes. This model with one species will underpin the basic model we use when extending to multiple species. However, it is also possible to use different infection dynamics; we will detail the implementation of transmission as we did for the lattice-based model in the coming sections.

5.2.2 Environmental Transmission

As with an ODE-based model, we can adapt the mechanism for transmission to occur via the environment. We introduce another matrix to represent infection in the environment, denoted E as in Equations 21 to 24 and as distinct from the empty state in the previous chapter. The matrix E gives the presence and location of pathogens in

the environment. Where there are N individuals present, E is an $(\rho \times N, 2)$ matrix; ρ here is the number of times during their infection period in which infected individuals deposit pathogens into the environment. In our model, these deposits occur at time steps corresponding to evenly spaced intervals within the (fixed) infection duration after rounding. These intervals exclude the initial and final time steps. For example, if $\rho = 4$ and $I_{max} = 10$, pathogens would be deposited in the second, fourth, sixth and eighth time steps after infection. When the i-th deposit occurs from the k-th individual, the pathogens are assigned the location of the infected individual at the time, occupying the $\rho \times (k-1) + i$-th row of E.

These pathogens remain static until they decay, requiring another timer matrix Q^E, with dimensions $(\rho \times N, 1)$. Where the pathogen in the i-th row of E occupies a site, $Q_i^E = Q_i^E + 1$. In our model, decay occurs uniformly after η time steps (i.e. when the relevant entry of $Q_E = \eta$), though this is not necessarily the case and can be amended to fit any desired distribution; alternatives may be polynomial or exponential. Where an individual has not deposited pathogens into the environment, those entries of E are set to $-L$, well outside of the region of the box between $0 < x, y < L$. This choice signifies the absence of pathogens without interfering with the rest of the model and saving the computation time associated with dynamicall adjusting array sizes.

Having established how transmission into the environment is handled, we now detail how pathogens in the environment transmit infection into the susceptible population. We use a similar process to the model with standard infection. However, instead of using a matrix of distances between all combinations of individuals, we must use a matrix of distances between all combinations of pathogen sites and individuals. The matrix C once again is used to give only the distances between pathogens in the environment and susceptible individuals; since only susceptible individuals can become infected, these are the only distances we will require to determine new infections. The method for calculating C differs since E will typically be larger than the position matrix X. Here, it takes the form

$$C_{ij} = \sqrt{(x_i - E_{x,j})^2 + (y_i - E_{y,j})^2}, \tag{90}$$

where $E_{x,j}$ and $E_{y,j}$ refer to the x- and y-coordinates of the j-th pathogen site in the environment. In this case, C is now an $N \times N_E$ matrix, where N_E is the number of pathogen sites in the environment. How the probability of infection is calculated from C remains the same; we use ω to represent this probability to remain in alignment with the corresponding system of ODEs, in which we have chosen to label the rate of infection from the environment the same way to make the distinction between the two infection

rates. It is important to note that ω here (and β in the standard transmission case) are functionally and mechanically identical in our agent-based models. The difference between them is an artefact of the chosen ODE nomenclature.

Having laid out a single-species model, we now detail how we extend to multiple species as with the lattice-based models earlier.

5.3 Multiple Species

Adapting the model for multiple species is a process similar to what was followed for the other agent-based models. We track species in the matrix A. Since agent positions are stored in the matrix X, in our spatially explicit model, A is a row vector with elements i corresponding to the species of the agent. Unlike the lattice-based model, there is no empty state.

We follow a similar process to our other models in defining other parameters. Where there are n species, the transmission rate becomes an $n \times n$ matrix β_{ij}, describing the rate of transmission from infected individuals of species i to susceptible individuals of species j. The recovery time and time for loss of immunity I_{max} and R_{max} can also be vectorised to be different for each species. We treat step sizes and correlations for the correlated random walks similarly; they become h_i and c_i for the i-th species. For simplicity, we assume that the radius of infectivity for an infected individual remains constant.

Introducing the effects of demography for a spatially explicit model is significantly more convoluted, though we outline some potential implementations in detail in the appendices. The principal difficulty encountered is handling births and deaths without an explicit empty state, as in the lattice-based models. Placing newborns in the grid randomly is unrealistic, and defining a neighbourhood and birth probability causes further burden on computation time. As such, we omit demography in our use of this model and consider biodiversity as a model input via our selection of initial conditions.

5.4 Estimating R_0

A property of the spatially explicit model is that it takes many time steps for an outbreak to progress. Since infection duration is preset, it is impossible for an atypically quick recovery of one initially infected agent to avert an outbreak entirely. It is possible to use the data from a spatially explicit model to estimate R_0, a quantity omitted from our analysis for the lattice-based model. We employ a method outlined in the book by

Diekmann et al. that is typically utilised in the context of real-world infection data, as developed by Roberts et al. [34, 114].

In the initial phases of an outbreak, where the population is almost entirely susceptible, the number of newly infected individuals grows approximately exponentially. Mathematically, it is described by an exponential function, $i(t) \approx e^{rt}$, where r is a parameter called the exponential growth rate. Over time periods insufficient for recovery or with chronic infection absence, we can equate $i(t)$ with the number of infected individuals $I(t)$, depending on model specifications. Advancing the initial outbreak by an interval Δt, then we can observe that

$$\frac{I(t+\delta t)}{I(t)} = e^{r\Delta t}$$

for small t, provided that Δt is small enough that the influence of acquired immunity within the population and recovery of infected individuals does not significantly affect the approximation. We can set $\Delta t = T_G$, the generation interval representing the expected time between a primary case and a secondary infection, and $I(t) = 1$ to represent one initial case. Recalling the definition of the basic reproduction number R_0 as the expected number of secondary cases from one primary case, then we can obtain the approximation

$$R_0 \approx e^{rT_G}$$

during the initial phases of an outbreak or epidemic. In our agent-based models, where we can observe or explicitly set r and T_G, this approximation can be an invaluable tool for estimating R_0.

In reality, the generation interval T_G is not just a particular number we set for our model but a mean interval dependent on a particular probability density function of time. We denote this time $f(t)$ and note using the formula for the mean value of a continuous variable that

$$T_G = \int_0^\infty t f(t) dt,$$

where f(t) is appropriately normalised. Our models use time steps where the probability of infections occurring can be more readily analysed than T_G, so we seek an expression based on $f(t)$ rather than T_G. In a simple environment in a closed, fully susceptible population where demography is either absent or slow enough to neglect, we can describe the incidence $i(t)$ by the formula

$$i(t) = I_0\delta(t) + \frac{S(t)}{N}\int_0^\infty A(\tau)i(i-\tau)d\tau,$$

where δt is the Dirac delta function, representing the introduction of initial infections. The kernel $A(t)$ is the infectivity of an infected individual at time t, which is equal to $R_0 f(t)$, such that integrating over 0 to ∞ would give R_0, the expected number of secondary cases. In the initial phases of an outbreak, $i(t) \approx e^{rt}$, and $S(t) \approx N$. If t is small and non-zero, then this is simplified to

$$R_0 \int_0^\infty e^{-rt} f(t)dt = 1. \tag{91}$$

These equations assume an SIR context where infected individuals are directly infectious. Some of our agent-based models do not use this context but feature infection via pathogens introduced into the environment. Because each infected individual sheds pathogens at set times that are directly attributable to each infected individual rather than the global number of infected individuals, we maintain that the only change to the derivation above would be in the precise definition of the kernel. Instead of infectivity of an individual, it would be infectivity in the environment stemming from an individual.

Our models use discrete time, and that presents a dilemma for employing this technique. We have two options: discretise the formula using the model-appropriate discrete $f(t)$, or fit a continuous curve to the model-appropriate $f(t)$ and proceed that way. Both of these methods can be viable, and comparing the two can be a worthwhile approach to exploring the utility and limitations of this estimator.

5.4.1 An Example of Inference

Here, we present the method for inferring R_0 alongside an example calculation. This example scenario will involve a single species and a pathogen spread via environmental transmission, where it may prove more challenging to count the secondary cases from one primary case in a simulation. The initial condition in this scenario should be set to the desired non-endemic equilibrium to mimic the introduction of an infectious disease to a typical population and avoid any influence on the infection dynamics from the population equilibrating.

First, we take the prevalence data, or the number of newly infected individuals at each timestep from a simulation. We can use prevalence in place of incidence since at the beginning of an outbreak, $I \approx 0$, and the timescales for prevalence I and pathogen levels in the environment E mean that the impact of recovery and decay is negligible.

For our example simulation, our chosen parameters are shown in Table 4:

Parameter	Value	Description
c	0.9	Correlation
h	0.0125	Step-size
L	1	Length
r	0.03	Infection radius
β	0.05	Infection probability
I_{max}	400	Recovery time

Table 4: Parameter values for the agent-based model with environmental transmission used to estimate R_0

In choosing these parameters, it was paramount to consider the timescales over which each process operates. We are interested in the case where the movement of individuals to new "neighbourhoods" occurs faster than the spread of infection, which in turn occurs much faster than the births and deaths, and our parameter selections reflect this. For example, should infection reach local saturation faster than an individual can move, then local prevalence is near-total, and infection dynamics are approximately linear in line with the diffusion of infected individuals from the position of the initially infected individuals.

We then seek to fit an exponential curve to the incidence data since we expect the initial phase of an outbreak to be characterised by exponential growth. We do this by taking the natural logarithm of each data point and performing a linear regression on the results. Supposing the regression result is $rx + c$, then $A = e^c$ and therefore $i(t) = Ae^{rt}$. Of course, due to the stochastic nature of the simulation, output will vary greatly between simulations. The period over which the exponential growth occurs in each simulation differs, so a thorough protocol is required to find the best-fitting curve for each simulation.

Fitting an exponential curve to the data requires isolating the domain over which the exponential increase occurs. To establish a domain over which to search, we start from $t = 1$ and check whether the prevalence is at a local maximum at each time step. If the prevalence does not exceed the last local maximum for more than a certain threshold

of time steps, we cut off the data at that point. In our example, this threshold is 200 timesteps. What remains over this domain is used for curve fitting.

Having established the domain over which to fit the exponential curve, we use linear regression on the natural logarithm of the prevalence data there. We then calculate the coefficient of determination R^2 of the regression line over the domain. This process is then repeated, shaving off the last timestep of data each time, gradually reducing the size of the domain. The process stops once the domain size reaches a predetermined minimum number of points (100 in the example).

We also consider the intercept of the exponential function. This intercept represents the initial number of infected individuals and should be approximately equal to the number used for the simulations. If this number is above or below a specific tolerance factor (3 in our example) of the known number, we do not consider that line, irrespective of its R^2. The line fitting this criterion and producing the highest R^2 value over its domain, and therefore the best fit of exponential behaviour is used to give the final approximation for that simulation.

Since this is a stochastic process, it is best to repeat it a number of times to get a statistically significant result. In our example, we performed 1000 simulations.

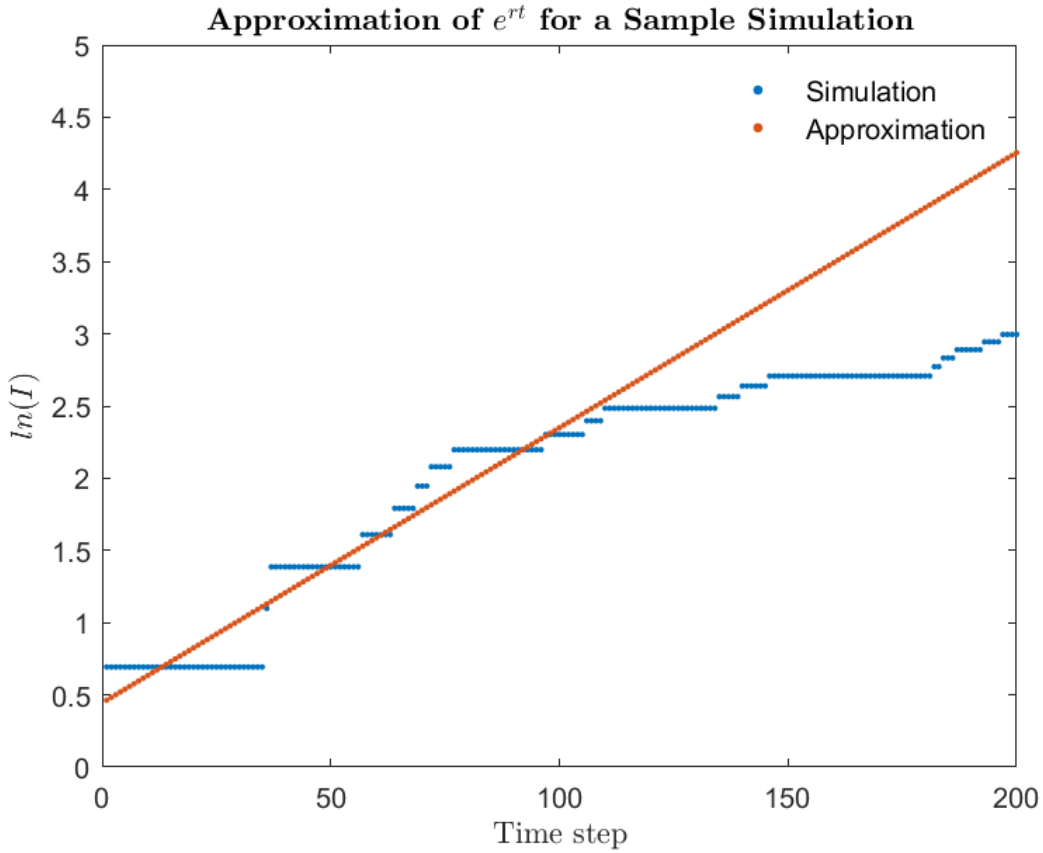

Figure 20: An example simulation with r around the median. Despite the approximation assuming small t, it gives a worse fit for very small values, owing to a high degree of stochastic randomness with low infection prevalence.

Of course, a stochastic simulation will not always yield the expected exponential increase. Even where $R_0 > 1$, the randomness underpinning the model can mean that the outbreak fails to gain a foothold in the population immediately or even at all. In these circumstances, we may not see an immediate exponential increase. To account for this, we use a combination of measures to isolate the exponential increase within the dataset and to ensure that the result is sensible.

One measure is checking the value of r. A negative value for r corresponds to an exponential decrease rather than an increase, so these are discarded. In epidemiological terms, this is where the infection would fail to spread significantly from its initial seeding in the population. Another measure is checking the value of the coefficient of determination R^2 for the best regression of each simulation. If this value is too low, it indicates dynamics that do not reflect the desired immediate exponential increase, and we discard these as well.

Both of the previous measures mentioned relate to the parameter output of the model. However, there may also be circumstances where the local maxima and the various thresholds do not permit a sensible regression. These runs are also discarded. It is important to note that discarding these data does not reflect any unintended behaviour in the model. Doing so is essential because we are interested in the case analogous to deterministic ODE model behaviour when $R_0 > 1$ is a point of comparison. As such, any different behaviour should be excluded.

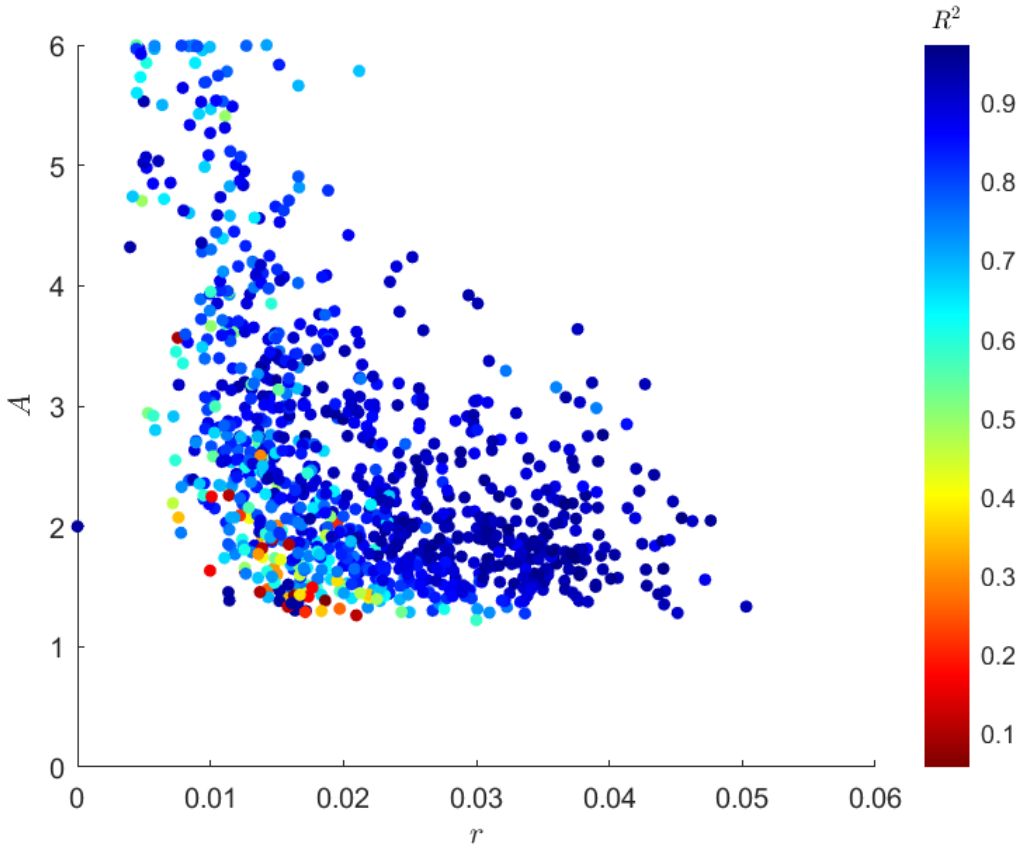

Figure 21: Data for the intercept A and growth rate r from 994 simulations. Six simulations with $r < 0$ were removed. The colour of each point represents the highest possible R^2 value using the procedure outlined previously. The best fits are mostly found around the middle values of r, while the worst fits mostly occur around low values of A and r.

From the 1000 example simulations, 943 remain after discarding those with unsuitable dynamics: 6 where $r < 0$, and 49 where $R^2 < 0.5$. A histogram of the

remaining r values shows that they are normally distributed about a positive value, as expected. We use the median value of r to calculate R_0 using Equation 91. Being a constant, the value of A is absorbed into R_0. The data are shown in Figure 21.

Having found a value of r, we must now estimate $f(t)$ for our model. The simulations used in the example employed environmental transmission. Infectious individuals would shed pathogens into a site two times while infected: once at the beginning of their infection and again at the conclusion of their infection. These sites can infect susceptible individuals equally at each timestep until they disappear; i.e. their infectivity does not decay but is constant until they are removed from the simulation after a fixed number of timesteps. Assuming so is unrealistic, but this simple implementation will suffice in the context of the example.

This transmission approach means that we can take $f(t)$ to be the sum of the distributions for the two different sites, denoted $f_1(t)$ and $f_2(t)$, and given by

$$f_1(t) = \sum_{i=1}^{E_{max}} \delta(t-i) \tag{92}$$

$$f_2(t) = \sum_{i=I_{max}+1}^{I_{max}+E_{max}} \delta(t-I_{max}-i), \tag{93}$$

where $\delta(t)$ is the Dirac delta function.

In Equations 92 and 93, we assume that the probability of producing a secondary infection is the same at each timestep. Though the infectivity is constant for each site, this is somewhat unrealistic because each time step is not mutually independent; if there are no individuals around the radius of infectivity for a site, it is likely to remain that way in subsequent timesteps. Accordingly, it is expected that $f(t)$ would decline over time somewhat, but assuming that $f_1(t)$ and $f_2(t)$ are constant provides a useful lower bound estimate on R_0. We denote the lower bound estimate R_0^-. Substituting $f(t)$ into Equation 91, we find that

$$\frac{1}{R_0} = \int_0^\infty e^{-rt}\left(f_1(t) + f_2(t)\right) dt \tag{94}$$

$$\implies \frac{2E_{max}}{R_0^-} = \sum_{i=1}^{E_{max}} e^{-rt} + \sum_{i=I_{max}+1}^{I_{max}+E_{max}} e^{-rt}.$$

This result is derived using the sifting property of the Dirac delta function and the linearity of integration over the summation. The factor of $2E_{max}$ ensures the normalisation of $f(t)$.

Using Equation 94 and our initial choices of parameters, all that is needed to find R_0^- is the growth rate r from our simulations. From the sample of 943 simulations, we find that the mean r value $\bar{r} = 0.0212$. Substituting the relevant parameters and evaluating the sum gives the final $R_0^- \approx 8.69$.

As demonstrated here, this technique can infer R_0 from simulations. However, there are circumstances where the expected exponential growth does not occur even accounting for randomness, so this method must be applied carefully. For example, the exponential increase may not be observed for parameter combinations where the infection is locally rapid, reaching saturation in the immediate area quickly before spreading out. Instead, the dominant driver of increased infection becomes the diffusion of infected individuals from the random walk, and the new cases appear more slowly. The value obtained from this method can also be compared to that obtained from simply counting the number of secondary infections from one "patient zero", though that method also has its drawbacks, mainly where the number of agents is low enough to violate the assumption of a totally susceptible population over the course of the infection for patient zero.

5.5 A Three-Host System

At equilibrium, the relative abundances of the species will remain constant. Equivalently, we can also assume that the time scale of the outbreak is sufficiently short that any effect of births and deaths would be minimal. We use our spatially explicit model with contact-based transmission with two species at equilibrium and consider the introduction of a third species. More specifically, we are interested in knowing when the increased competence of the third host is sufficient to counteract the diluting effect from adding another species when interspecies transmission is assumed to be much less likely than intraspecies transmission.

We approach this by considering a system where two species A and B exist at equilibrium in a square ecosystem of length $L = 10$. An infection circulates that is modelled with an SIR approach. The species are assumed to be identical regarding their movement and transmission and have equal population densities, with a total of 1200 agents. We take the correlation probability to be $c = 0.04$ and the step size to be $h = 0.05$. In general, we assume that the transmission rates are highest between

members of the same species and lower between members of different species. We set the transmission rates to be identical for the two species with infection duration of $I_{max} = 200$ time steps. The radius of contagiousness is taken to be $r = 0.04$. Initially, 2.5% of the population is infected. The introduced species C is identical to the other two, except for its transmission rate $\beta_{i,C}$, which differs by a factor compared to the rates for other species. The transmission rate matrix β is

$$\beta = \begin{bmatrix} 0.3 & 0.05 & 0.05 \\ 0.05 & 0.3 & 0.05 \\ 0.1 & 0.6 & 0.1 \end{bmatrix}$$

such that an infected agent of the introduced species has twice the probability of infecting any susceptible agent compared to the other two species. We vary the amount of the introduced species $0 \leq N_C \leq 800$ on intervals of 20 but maintain the total number of agents $N = 1200$. The introduction affects the other two species equally, and their populations are given by $N_A = N_B = (1200 - N_C)/2$. We perform 10 realisations of the model for each N_C, producing a 410 data point-sample.

Our chosen measure of infection risk is the final size of the outbreak in each simulation, R_∞. We opt to compare two measures of biodiversity by using both the Shannon index H and the Gini-Simpson index $1 - \lambda$. The risk of infection is calculated as the mean of the 10 simulations for any given N_C. The biodiversity measures can be evaluated analytically in terms of one variable, N_C:

$$(1 - \lambda) = 1 - \left(\frac{N_C}{1200}\right)^2 - 2\left(\frac{1200 - N_C}{2400}\right)^2 \tag{95}$$

$$H = -\frac{N_C}{1200}\log\frac{N_C}{1200} - 2\frac{1200 - N_C}{1200}\log\frac{1200 - N_C}{1200}. \tag{96}$$

Since we seek to calculate an elasticity eventually, we are also interested in their derivatives, given by

$$\frac{d(1 - \lambda)}{dN_C} = \frac{400 - x}{480000} \tag{97}$$

$$\frac{dH}{dN_C} = \frac{1}{1200}\left(\log 600 - N_C/2 - \log x\right). \tag{98}$$

Both derivates are zero at $N_C = 400$, where all three species are at equal abundance, as expected. We seek to calculate the elasticity that quantifies the dilution effect, which is

$$\mathcal{D}_{1-\lambda} = \frac{1-\lambda}{R_\infty} \frac{dR_\infty}{dN_C} \frac{dN_C}{d(1-\lambda)} \tag{99}$$

$$\mathcal{D}_H = \frac{H}{R_\infty} \frac{dR_\infty}{dN_C} \frac{dN_C}{dH}. \tag{100}$$

The raw final size data with the confidence intervals are shown in Figure 22.

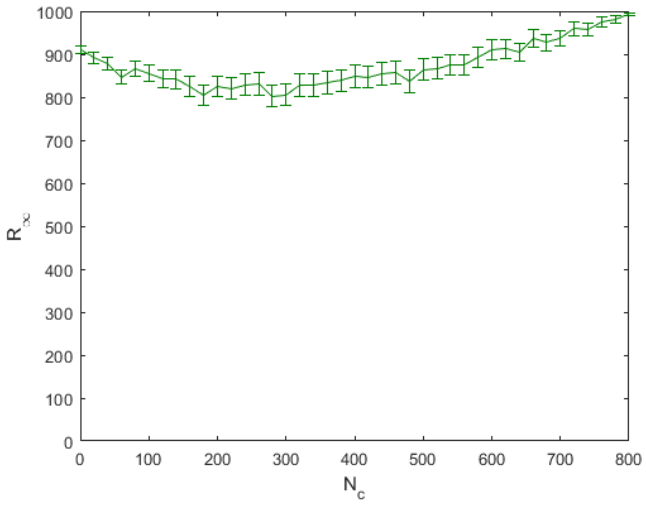

Figure 22: Mean final sizes from 10 simulations with different values of N_C. The error bars signify a 95% confidence interval.

Compared to the lattice-based model shown in Figure 16, the error is much greater. The higher margin for error is owed to the need to manage computation time alongside statistical validity, as it is easier to increase sample size (and decrease error) with a more lightweight model. The high error where the three species coexist in more even numbers reflects a high dependence on low probability cross-species transmission events. We rely on a degree of correlation between final sizes for adjacent values of N_C to handle outliers and smooth the data. We then obtain a differentiable expression for the elasticity calculation using cubic spline interpolation. The noisier data presents a hurdle for the smoothing process, as shown in Figure 23.

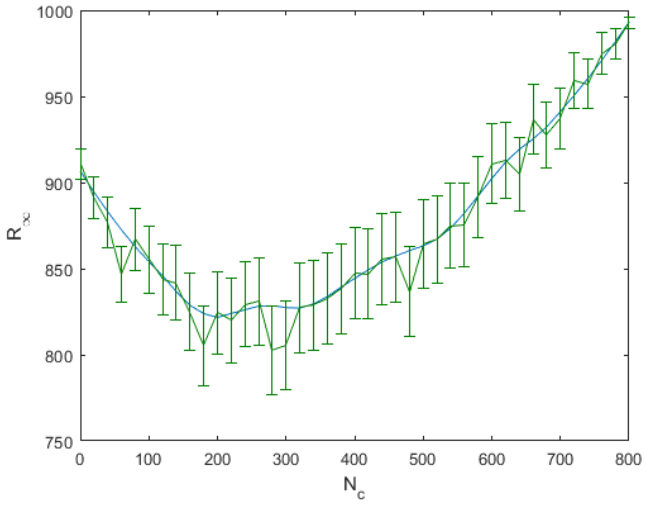

Figure 23: Smoothed data for mean final sizes R_∞ for each N_C. The error bars signify a 95% confidence interval. Note the axis scaling compared to Figure 22

Despite this, we can calculate the elasticity using both biodiversity metrics as shown in Figure 24.

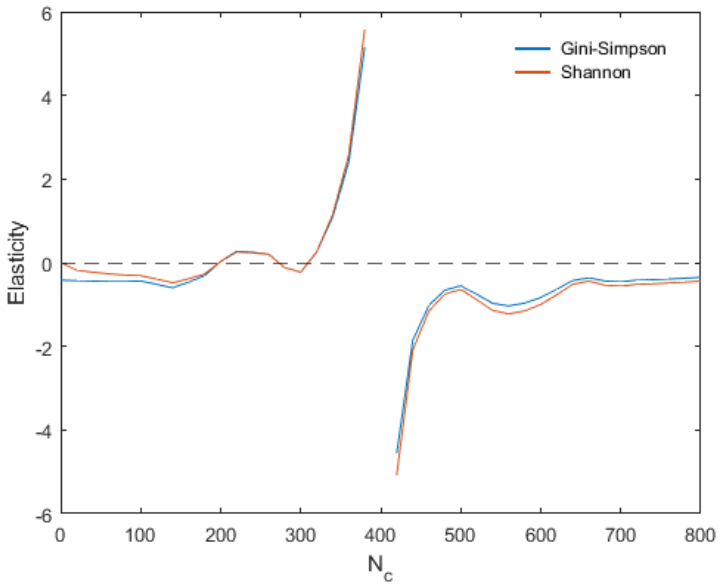

Figure 24: Elasticities calculated using final size and both the Gini-Simpson and Shannon indices, for various values of N_C. The dotted line along the x-axis is the threshold for amplification (i.e. elasticity greater than 0) and dilution (i.e. elasticity less than 0).

We produce nearly identical results for both of the biodiversity metrics. There is a critical point at $N_C = 400$ where the derivatives of the biodiversity indices are 0, and we see asymptotic behaviour on either side. A dilution occurs where $N_C > 400$, as introducing a more competent host in that region decreases biodiversity and increases final size R_∞. The amplification from the introduced species is apparent in a limited region where $N_C < 400$. The exact critical point is unclear owing to the error in the data, which caused the increase in elasticity not to be monotonic.

Despite the amplification effect in that region, at no point before maximum biodiversity at $N_C = 400$ does the final size ever exceed the final size for $N_C = 0$ The region in which the final size decreases with N_C would decrease with the transmission rate of the more competent host. An upper bound for the phenomenon would be when the entries of β for interspecies transmission of the more competent host are equal to the intraspecies transmission entries for the other hosts.

In essence, this result is evidence of a dilution effect in that introducing a more competent host still decreased the final size of the outbreak until the introduced host threatened biodiversity by outcompeting the other species. However, the assumptions that underpin that conclusion do not hold for many infections, where transmission between species occurs more often than within species or where a vector transmits the infection. We also assumed that the resource competition between the species affected each species equally, resulting in the same total number of agents no matter the relative abundance of each species. That does not hold in general, and the subsequent change in population density stemming from an increased total of agents may result in a different observation of dilution or amplification.

The significance of this result is that in the literature, a number of examples given significant weight in discussion of the dilution effect focus on settings where loss of biodiversity predominantly impacts low quality hosts [64, 71]. In this setting, with strong interspecific competition and where intraspecific transmission is most common, dilution is a trivial consequence. This modelling approach presents a converse scenario, where the population of a more competent host is driving biodiversity, leading to dilution outside of a small region near a critical point of equal populations, as seen in Figure 24. Our results suggest a regime where the dilution effect can be observed more reliably, but still not assuredly, concurring with a similar observation made by Keesing and Ostfeld [70].

5.6 Model Discussion

The spatially explicit model is more computationally expensive than the lattice-based model, with many calculations required to generate the distances between hosts and many more processes to execute in a given timestep, such as host movement and infection progress. In exchange for this, the model can better demonstrate the spatial component of an emerging infection. In particular, the matrix-based process of calculating distances involves $n(n-1)$ distance calculations since we duplicate them in a symmetric matrix. This computation is an $\mathcal{O}(n^2)$ operation, while the matrix multiplications involved in computing the matrix d are an even higher-order process. In contrast, each process in the lattice-based model requires $\mathcal{O}(n)$ operations, and therefore it can run significantly faster.

A further limitation of the spatially explicit model is the numerous matrices required to handle all the information needed. This level of detail renders explicitly handling demography an involved and expensive process, as the lattice structure provides an easy system for identifying hosts. A potential implementation of this using system-wide population data is provided in the appendices, though it is noted that including demography can slow the model down substantially. Omitting demography requires biodiversity to be handled directly through the initial conditions. It limits the scope of the model to investigate the influence of biodiversity on infection to the exclusion of the converse scenario. Further work is needed to investigate the most effective implementation of demography.

A strength of the spatially explicit model is how its boundary conditions can directly relate to the boundary conditions for the diffusion equation. Unlike other models, handling movement explicitly allows for the implementation of Neumann, Dirichlet, and periodic boundary conditions corresponding to many kinds of biomes. Specific boundary conditions can be used to model control measures such as trapping and releasing animals that encroach on settlements, which are not handled as well by ODEs or lattices.

Since the spatially explicit model captures fine spatial dynamics and is most elegant without demography, it is best suited to modelling systems on shorter time scales than the lattice-based model, where the population can be assumed to be constant. Simpler systems also mitigate the increased computation time. Similarly, the ability to capture more dynamics with its boundary conditions makes the models well-suited to simulating smaller communities with detail in their environment.

6 Modelling Toxoplasmosis

Having discussed both continuous-time and discrete-time approaches for describing the dynamics of infectious diseases with agent-based models, we now turn our attention to a pair of case studies with more complex dynamics that are better suited to modelling with ODEs. The first of these case studies is toxoplasmosis. Toxoplasmosis is caused by the *Toxoplasma gondii* protozoan, with members of the family Felidae acting as its definitive hosts, within which it can reproduce sexually. The protozoa are spread from felids via the shedding of oocysts in their faeces, from which they can be ingested by other felids, as well as various possible intermediate hosts, including mice, birds, livestock and humans. Reproducing asexually, the protozoa form cysts in the tissues of these intermediate hosts, resulting in further spread of infection should these hosts be consumed by cats or other predators [22].

Toxoplasmosis is notable for the behavioural shifts it brings about in its hosts, which have been the subject of numerous scientific studies [13, 45, 146]. Berdoy, Webster and MacDonald discovered that infection with *T. gondii* in some cases caused their natural aversion to the scent of cats to be overridden with a "fatal attraction" that would increase their risk of predation, resulting in higher levels of transmission of the pathogen between the intermediate and definitive hosts, constituting an intriguing relationship between host ecology and transmission [13]. A study by Ingram et al. suggested that this loss of aversion may be permanent, or at least persisting well after acute infection [62].

We seek to use our models to describe toxoplasmosis in a population with cats and some intermediate hosts to analyse how changing the various parameters and number of hosts in the model affects the equilibrium states of the system. We place particular emphasis on factors pertaining to the dilution effect, such as evenness in steady-state populations of hosts, the basic reproduction number R_0, the number of infected cats, and the prevalence of toxoplasma protozoa in the environment.

6.1 A Base Model

We begin by considering a system of ODEs to model toxoplasmosis. Initially, we consider the case with cats and prey species, which we will call mice from this point forward. In constructing our model, we refer to cats by assigning the subscript c to our variables and parameters, and to mice with the subscript m. We use an SIR framework for cats and mice, in which recovered individuals retain immunity until death, neglecting

the possibility of vertical transmission. The model treats the shedding of oocysts by infected cats as an environmental contamination, with E representing the number of oocysts present at time t. In the absence of cats, we assume competition for finite resources (i.e. logistic growth) in mice. This results in a system of seven ODEs, and is structurally similar to the single-species model presented by Kelting [72] discussed in Section 2.1.3.

$$\frac{dS_c}{dt} = -\omega_c E S_c + \mu_c N_c - \alpha_c S_c + \phi_{cm} N_c (S_m + R_m) + \delta \phi_{cm} (I_c + R_c) I_m \tag{101}$$

$$\frac{dI_c}{dt} = \omega_c E S_c - \gamma_c I_c - \alpha_c I_c + \delta \phi_{cm} S_c I_m \tag{102}$$

$$\frac{dR_c}{dt} = \gamma_c I_c - \alpha_c R_c \tag{103}$$

$$\frac{dS_m}{dt} = -\omega_m E S_m + \mu_m N_m - \alpha_m S_m - \phi_{mm} N_m S_m - \phi_{mc} N_c S_m \tag{104}$$

$$\frac{dI_m}{dt} = \omega_m E S_m - \gamma_m I_m - \alpha_m I_m - \phi_{mm} N_m I_m - \delta \phi_{mc} N_c I_m \tag{105}$$

$$\frac{dR_m}{dt} = \gamma_m I_m - \alpha_m R_m - \phi_{mm} N_m R_m - \phi_{mc} N_c R_m \tag{106}$$

$$\frac{dE}{dt} = \lambda I_c - \eta E. \tag{107}$$

Once again, we have used $N_i = S_i + I_i + R_i$ for each species for ease of interpretation. The parameters within these equations remain as introduced earlier in Equations 21 to 24.

The structure of these equations allows for both cats and mice to become infected through contact with oocysts in the environment, at rates ω_m and ω_c per susceptible individual per unit oocyst. Infected cats shed oocysts into the environment at rate λ per infected cat, while infected mice do not shed oocysts at all. Infected mice become more likely to be consumed by cats, represented by a factor $\delta \geq 1$. In lieu of a direct means by which susceptible cats consume infected mice and become infected, we consider transmission through the predator-prey route to be equivalent to the increased fecundity in cats as a result of infected mice, via the $\delta \phi_{cm} S_c I_m$ term in Equation 102. This is not generally true, but it produces a good approximation of the overall effect, since the model is not age-structured. The remaining increase in fecundity through predation results in susceptible cats, accounting for the last two terms in Equation 101.

6.1.1 Steady-State Analysis

We now consider the possible steady states of this model, setting the left-hand sides of each equation to be equal to 0. First, we find the possible infection-free steady states by setting $I_c = I_m = R_c = R_m = E = 0$. The total extinction steady state occurs when $S_c^* = S_m^* = 0$. Two other steady states exist, the first of which is the cat-free state, with $S_c^* = 0$ and

$$S_m^* = \frac{\mu_m - \alpha_m}{\phi_{mm}}. \tag{108}$$

This expression is the steady state of the logistic equation, which aligns with the construction of the system outlined previously. For this steady state to exist, it must hold that $\mu_m > \alpha_m$, since $S_m^* \geq 0$. The second steady state is one in which cats and mice coexist, given by

$$S_c^* = \frac{(\mu_c - \alpha_c)\phi_{mm} + (\mu_m - \alpha_m)\phi_{cm}}{\phi_{cm}\phi_{mc}} \tag{109}$$

$$S_m^* = \frac{\alpha_c - \mu_c}{\phi_{cm}}. \tag{110}$$

This steady state exists only where $\alpha_c > \mu_c$, which, in the context of the problem being modelled, corresponds to the case where cats depend on the increase in fecundity from their access to prey. A further requirement then is that $(\mu_m - \alpha_m)\phi_{cm} > (\alpha_c - \mu_c)\phi_{mm}$. This requirement, in context, represents the need for sufficient mice to generate a sufficient increase in fecundity in cats. Notably, the existence of this steady state is not dependent on the reduction of the mouse population as a result of predation by cats. Since we are interested in the cases where each species survives, we restrict ourselves to the parameter space where $\alpha_c > \mu_c$ and $\mu_m > \alpha_m$ so that this fixed point may be stable in the absence of infection. We also note the lack of a steady state in which cats exist without mice; this is a consequence of the construction of the equations. Should the mouse population be zero, then the cat ODEs are reduced to the basic system for exponential population growth, resulting in a population that either grows boundlessly or asymptotically approaches 0.

We analyse the stability of these fixed points by constructing a Jacobian J and considering its eigenvalues. The Jacobian is a 7×7 matrix with the following form whose entries J_{ij} are the partial derivatives of the right-hand side of the differential equation for the i-th variable with respect to variable j.

Each steady state is stable where every eigenvalue λ of the Jacobian evaluated at the fixed point has a negative real part, i.e. $\forall i = 1 : 7, Re(\lambda_i) < 0$. The Jacobian

evaluated at the total extinction steady state is given by

$$J = \begin{bmatrix} r_c & \mu_c & \mu_c & 0 & 0 & 0 & 0 \\ 0 & -\alpha_c - \gamma_c & 0 & 0 & 0 & 0 & \\ 0 & \gamma_c & -\alpha_c & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & r_m & \mu_m & \mu_m & 0 \\ 0 & 0 & 0 & 0 & -\gamma_m - \alpha_m & 0 & 0 \\ 0 & 0 & 0 & 0 & \gamma_m & -\alpha_m & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & -\eta \end{bmatrix}, \tag{111}$$

where we have used the intrinsic growth rate $r_i = \mu_i - \alpha_i$ to simplify certain expressions. After finding the eigenvalues of this matrix, we see that five are always negative, while the other two are given by $\lambda_6 = \mu_c - \alpha_c$ and $\lambda_7 = \mu_m - \alpha_m$. Hence, the total extinction steady state is stable only if $\mu_c < \alpha_c \wedge \mu_m < \alpha_m$.

We repeat the same process for the mouse-only infection-free steady state, with a Jacobian given by

$$J = \begin{bmatrix} r_c + \frac{r_m \phi_{cm}}{\phi_{mm}} & \mu_c + \frac{r_m \phi_{cm}}{\phi_{mm}} & \mu_c + \frac{r_m \phi_{cm}}{\phi_{mm}} & 0 & 0 & 0 & 0 \\ 0 & -\alpha_c - \gamma_c & 0 & 0 & 0 & 0 & \\ 0 & \gamma_c & -\alpha_c & 0 & 0 & 0 & 0 \\ \frac{-r_m \phi_{mc}}{\phi_{mm}} & \frac{-r_m \phi_{mc}}{\phi_{mm}} & \frac{-r_m \phi_{mc}}{\phi_{mm}} & -r_m & \alpha_m & \alpha_m & \frac{-r_c \omega_m}{\phi_{mm}} \\ 0 & 0 & 0 & 0 & -\gamma_m - \mu_m & 0 & \frac{r_c \omega_m}{\phi_{mm}} \\ 0 & 0 & 0 & 0 & \gamma_m & -\mu_m & 0 \\ 0 & \lambda & 0 & 0 & 0 & 0 & -\eta \end{bmatrix}. \tag{112}$$

We again see five eigenvalues that are always negative and a sixth, which follows from the existence requirement, $\lambda_6 = \mu_m - \alpha_m$. The seventh eigenvalue is given by

$$\lambda_7 = \frac{\phi_{cm}(\mu_m - \alpha_m) + \phi_{mm}(\mu_c - \alpha_c)}{\phi_{mm}}, \tag{113}$$

which is negative provided that $\phi_{cm}(\mu_m - \alpha_m) < \phi_{mm}(\alpha_c - \mu_c)$. This condition is the reverse of the existence condition for the infection-free steady state with both species

present; it is important to note that this can only hold when cats depend on the increase in fecundity from mice to survive.

Attempting the same process for the infection-free steady state with both species proves challenging. The Jacobian is given by

$$J = \begin{bmatrix} 0 & \alpha_c & \alpha_c & \frac{r_m\phi_{cm}+r_c\phi_{mm}}{\phi_{mc}} \\ 0 & -\alpha_c-\gamma_c & 0 & 0 \\ 0 & \gamma_c & -\alpha_c & 0 \\ \frac{r_c\phi_{mc}}{\phi_{cm}} & \frac{r_c\phi_{mc}}{\phi_{cm}} & \frac{r_c\phi_{mc}}{\phi_{cm}} & \frac{r_c\phi_{mm}}{\phi_{cm}} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & \frac{r_m\phi_{cm}+r_c\phi_{mm}}{\phi_{mc}} & \frac{-(r_m\phi_{cm}+r_c\phi_{mm})\omega_c}{\phi_{mc}\phi_{cm}} \\ \frac{\delta(r_m\phi_{cm}+r_c\phi_{mm})}{\phi_{mc}} & 0 & \frac{(r_m\phi_{cm}+r_c\phi_{mm})\omega_c}{\phi_{mc}\phi_{cm}} \\ 0 & 0 & 0 \\ \mu_m+\frac{r_c\phi_{mm}}{\phi_{cm}} & \mu_m+\frac{r_c\phi_{mm}}{\phi_{cm}} & \frac{r_c\omega_m}{\phi_{cm}} \\ \frac{(1-\delta)r_c\phi_{mm}-(\alpha_m+\gamma_m+\delta r_m)\phi_{cm}}{\phi_{cm}} & \frac{(1-\delta)r_c}{\phi_{cm}} & \frac{-r_c\omega_m}{\phi_{cm}} \\ \gamma_m & -\mu_m & 0 \\ 0 & 0 & -\eta \end{bmatrix}. \quad (114)$$

We obtain two eigenvalues which are always negative ($\lambda_1 = -\alpha_c$, $\lambda_2 = -\mu_m$), and a further two from a solution to a quadratic equation. These are given by

$$\lambda_{3,4} = \frac{\phi_{mm}(\mu_c-\alpha_c) \pm \sqrt{\alpha_c-\mu_c}\sqrt{(\phi_{mm}(4\phi_{cm}+\phi_{mm})(\alpha_c-\mu_c)-4\phi_{cm}^2)(\mu_m-\alpha_m)}}{2\phi_{cm}}. \quad (115)$$

It can be seen that the first term of the numerator is negative in the regime where the fixed point exists and that the argument of the first square root of the product in the second term will be negative as well. Should the second square root be positive, then the second term in the numerator will be purely imaginary, and the eigenvalues will have negative real parts. If this is not the case, then the absolute value of the

second term must be less than the first. Evaluating these two conditions recovers an inequality identical to the existence requirement, so these two eigenvalues will always have negative real parts where the steady state exists. These eigenvalues do not depend on the parameters that describe infection, and reflect the dynamics of the underlying Lotka-Volterra predator-prey system with logistic growth in the prey.

The remaining three eigenvalues are the roots of a cubic polynomial with coefficients of many terms; we omit their exact forms due to their complexity, and note that solving them is challenging. We can use the Routh-Hurwitz stability criterion to narrow this down. We denote the coefficients of the λ^n term of the polynomial to be a_n. The Routh-Hurwitz criteria state that, for a cubic polynomial to have all of its roots have negative real parts, the coefficients must satisfy

- $a_0 > 0$
- $a_2 > 0$
- $a_2 a_1 > a_0$.

We find that $a_2 > 0$ where the fixed point exists. We also find that $a_0 > 0$ is satisfied depending on two other conditions. Firstly, there exists a requirement on the ratio λ/η in terms of the infection rate of cats, ω_c, given by

$$\frac{\lambda\omega_c}{\eta} < \frac{(\alpha_c + \gamma_c)\phi_{cm}\phi_{mc}}{(\mu_m - \alpha_m)\phi_{cm} - (\alpha_c - \mu_c)\phi_{mm}} \tag{116}$$

$$\implies \frac{\lambda\omega_c}{\eta} < \frac{\alpha_c + \gamma_c}{S_c^*}. \tag{117}$$

This requirement is that the number of new infections in cats is insufficient to sustain toxoplasmosis in the environment, addressing the route of infection in which cats become infected directly. The second requirement addresses the other route of infection by which cats become infected by predating infected mice. Explicitly, this requirement specifies a minimum rate for the recovery of mice γ_m, in terms of the infection rate of mice ω_m and the increase in the rate of predation of infected mice, δ. Mathematically, this is given by

$$\gamma_m + \mu_m > \frac{\delta\lambda\omega_m\phi_{cm}S_c^*S_m^*}{(\alpha_c + \gamma_c)\eta - \lambda\omega_c S_c^*} - (\delta - 1)\phi_{mc}S_c^*. \tag{118}$$

The form of this inequality leads us to draw several conclusions. As expected, we can observe that the right-hand side increases with ω_m and the ratio λ/η. The role

of δ is more complex. Differentiating the threshold with respect to δ reveals that the right-hand side increases with δ provided that

$$(\alpha_c + \gamma_c)\eta\phi_{mc} < \lambda\phi_{mc}\omega_c S_c^* + \lambda\phi_{cm}\omega_m S_m^*. \tag{119}$$

We find that the third of the Routh-Hurwitz criteria is satisfied as a consequence of the first two, i.e. that $a_0 > 0 \wedge a_2 > 0 \implies a_2 a_1 > a_0$. It follows that the two previous inequalities are the conditions for the disease-free equilibrium to be stable when combined with the existing criteria previously discussed.

Having found conditions for the disease-free equilibria, we now focus on the case where infection persists. Knowing when the infection-free equilibria are stable is insufficient for drawing conclusions on the behaviour of the system more broadly, so we turn our attention to investigating possible conditions for when the parasite will persist. Finding fixed points with infection for the system is impractical because of its algebraic complexity, so we use the next-generation matrix method to find this condition.

6.1.2 The Next-Generation Matrix

In order to determine a condition for toxoplasmosis to be endemic in our model, we aim to establish where $R_0 > 1$ using the next-generation method.

We construct the next-generation matrix by considering the three infected compartments of our toxoplasmosis models, as outlined by Diekmann and Heesterbeek [35], and applied in an ecological context by Roberts and Heesterbeek [115]. Specifically, we consider the equations for the time-derivatives of I_c, I_m, and E. We evaluate a submatrix of the Jacobian J considering only these compartments and their corresponding variables I_c, I_m, and E. We denote this matrix by H; in the case of our model, it takes the form of the following 3×3 matrix:

$$H = \begin{bmatrix} \frac{\partial f_{I_c}}{\partial I_c} & \frac{\partial f_{I_c}}{\partial I_m} & \frac{\partial f_{I_c}}{\partial E} \\ \frac{\partial f_{I_m}}{\partial I_c} & \frac{\partial f_{I_m}}{\partial I_m} & \frac{\partial f_{I_m}}{\partial E} \\ \frac{\partial f_E}{\partial I_c} & \frac{\partial f_E}{\partial I_m} & \frac{\partial f_E}{\partial E} \end{bmatrix}. \tag{120}$$

To calculate R_0, we must decompose H into two further matrices, the transmission matrix T and the transition matrix Σ, which satisfy $H = T + \Sigma$ and are evaluated at the non-endemic equilibrium. The matrix T contains the terms corresponding to the emergence of new infections from the environment or via prey consumption, as well as the "transmission" of oocysts into the environment via shedding. The matrix Σ contains terms corresponding to the transfer of individuals out of the infected compartment by

all means, including natural deaths, death by predation or competition, recovery from infection, or the expiration of oocysts. For our system, these matrices are

$$T = \begin{bmatrix} 0 & \delta\phi_{cm}S_c^* & \omega_c S_c^* \\ 0 & 0 & \omega_m S_m^* \\ \lambda & 0 & 0 \end{bmatrix} \tag{121}$$

$$\Sigma = \begin{bmatrix} -\alpha_c - \gamma_c & 0 & 0 \\ 0 & -\alpha_m - \gamma_m - \delta\phi_{mc}S_c^* - \phi_{mm}S_m^* & 0 \\ 0 & 0 & -\eta \end{bmatrix}. \tag{122}$$

We now have everything we need to calculate the next-generation matrix of large domain, which is defined by Diekmann et al. [36] as $K_L = -T\Sigma^{-1}$. The largest eigenvalue of this matrix, known as the spectral radius ρ, gives the basic reproduction number R_0 that we seek. In the case of our model, K_L is given by

$$K_L = \begin{bmatrix} 0 & \frac{\delta\phi_{cm}S_c^*}{\alpha_m+\gamma_m+\delta\phi_{mc}S_c^*+\phi_{mm}S_m^*} & \frac{\omega_c S_c^*}{\eta} \\ 0 & 0 & \frac{\omega_m S_m^*}{\eta} \\ \frac{\lambda}{\alpha_c+\gamma_c} & 0 & 0 \end{bmatrix}. \tag{123}$$

We note that each element of the next-generation matrix of large domain is positive, which will prove helpful in determining R_0. The eigenvalues of this matrix are the solutions λ_i to the cubic equation

$$\lambda_i^3 - \frac{\lambda\omega_c S_c^*}{(\alpha_c+\gamma_c)\eta}\lambda_i - \frac{\delta\phi_{cm}\lambda\omega_m S_c^* S_m^*}{(\alpha_c+\gamma_c)\eta(\alpha_m+\gamma_m+\delta\phi_{mc}S_c^*+\phi_{mm}S_m^*)} = 0. \tag{124}$$

This equation has no quadratic term and two strictly negative coefficients, meaning that this is a depressed cubic with solutions that can be directly computed using the Cardano formula. When $\lambda_i = 0$, we note that the left-hand side of the equation and its derivative will be negative, while we can see that as λ_i approaches ∞, the left-hand side will grow in the positive direction without bound. These facts guarantee the existence of a single positive and real root; should other real roots exist, they must be

negative. The Cardano formula states that a real solution of an equation of the form $x^3 + px + q = 0$ is

$$x = \sqrt[3]{\frac{-q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{\frac{-q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}, \tag{125}$$

where we take the real cube root should its argument be real or the principal value with the maximum real part; otherwise, using the standard branch cut along the negative real axis. We know that this solution corresponds to our spectral radius (and therefore R_0) since we have deduced that it is the only positive root of the equation. We use this formula and perform some manipulation to obtain our required value for R_0, which is given by

$$R_0 = B^{\frac{1}{3}}\Big(\big(1 + \sqrt{1 - A}\big)^{\frac{1}{3}} + \big(1 - \sqrt{1 - A}\big)^{\frac{1}{3}}\Big), \tag{126}$$

where A and B are defined as functions of the various parameters and the non-endemic equilibria S_c^* and S_m^* as

$$A = \frac{4\lambda\omega_c^3(\alpha_m + \gamma_m + \delta\phi_{mc}S_c^* + \phi_{mm}S_m^*)^2 S_c^*}{27\eta(\alpha_c + \gamma_c)\delta^2\phi_{cm}^2\omega_m^2 S_m^{*2}} \tag{127}$$

$$B = \frac{\delta\lambda\phi_{cm}\omega_m S_c^* S_m^*}{2\eta(\alpha_c + \gamma_c)(\alpha_m + \gamma_m + \delta\phi_{mc}S_c^* + \phi_{mm}S_m^*)}. \tag{128}$$

To analyse the regions where $R_0 > 1$, we can set $R_0 = 1$ and plot B with respect to A, finding the critical curve which separates the $R_0 < 1$ and $R_0 > 1$ regions. As Figure 25 shows, should $B > 1/2$, it is assured that $R_0 > 1$, which is a useful threshold.

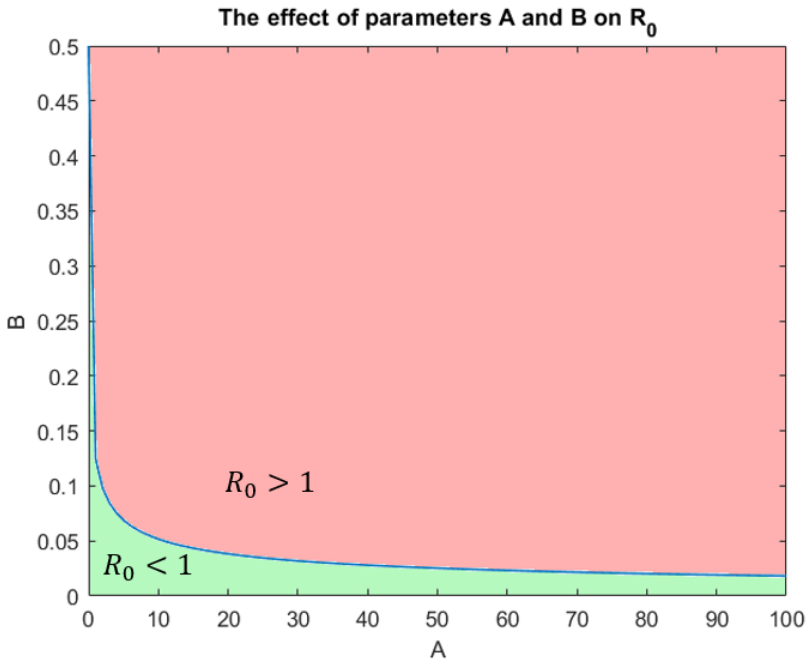

Figure 25: Parameters A and B where $R_0 = 1$, the threshold for a sustained outbreak. The shaded areas indicate regions of stability for the endemic and non-endemic equilibria.

From this graphical and analytical evidence, it is always the case that $R_0 > 1$ where

$$\delta\lambda\phi_{cm}\omega_m S_c^* S_m^* > \eta(\alpha_c + \gamma_c)(\alpha_m + \gamma_m + \delta\phi_{mc}S_c^* + \phi_{mm}S_m^*). \tag{129}$$

A similar condition with an additional term can be obtained by observing that $x^3 + px + q = 0$ will have a positive root $R_0 > 1$ where $-(p + q) > 1$, producing the following:

$$\delta\lambda\phi_{cm}\omega_m S_c^* S_m^* > \left(\eta(\alpha_c + \gamma_c) - \lambda\omega_c S_c^*\right)(\alpha_m + \gamma_m + \delta\phi_{mc}S_c^* + \phi_{mm}S_m^*). \tag{130}$$

However, it is not assured that the system converges to a steady-state even when $R_0 > 1$, and certain parameter values lead to oscillatory behaviour in both the predator and prey populations and the infection dynamics. Hence, while we have we determined an expression for the basic reproduction number R_0, we have only found thresholds for the *possible* stability of the endemic equilibrium in conjunction with our findings for the non-endemic equilibrium, all of which we will use in our upcoming analysis of the dilution effect.

6.2 Model Behaviour and the Dilution Effect

This model has a relatively rich range of possible behaviours. As alluded to in our steady-state analysis, we can observe various possible non-endemic steady states, including those with just mice, both mice and cats and total extinction. We also numerically find at least one non-endemic steady state, a state of exponential increase in cats, and a state of stable oscillations in cats and mice in the presence of infection, a particularly compelling result in the context of the dilution effect as the corresponding predator-prey system without infection does not feature oscillations. In this case, the presence of an infectious disease fuels oscillatory behaviour; variation in parameters relating to the spread of disease can alter the amplitude and period of these oscillations, which has interesting implications in terms of biodiversity not traditionally encountered in discussions of the dilution effect.

We now aim to apply our model to quantifying the dilution effect where the endemic steady state is present. Our model is similar to those considered by Roberts et al. [116], in which the dilution effect of a two-species SIS model is quantified with respect to the degree of direct, cross-species transmission. Our model differs by transmission mode, featuring environmental transmission and transmission by predation. Like Roberts et al., we will consider how the degree of cross-species transmission affects the biodiversity of the system. In our case, we do this by varying the factor by which infected mice are more likely to be predated by cats, δ, resulting in more infection in cats.

In order to evaluate the dilution effect, we must first define how we measure both "biodiversity" and "risk of infection". With only two species, there is no need to use biodiversity indices that account for richness. So, we consider biodiversity in the form of evenness between the two species present, in the form of the ratio N_c/N_m. The closer this ratio is to 1, the more biodiverse our system is. We consider "risk of infection" in three ways. These are the basic reproduction number R_0, the number of oocysts in the environment, E, and the prevalence of infection in cats, $Y_c = I_c/N_c$, all measured at equilibrium. It may be tempting to consider the number of infected cats directly. However, we can observe from the form of the ODEs that at equilibrium, E is proportional to I_c, rendering this measure redundant.

As for the metric for measuring the dilution effect quantitatively, we use Roberts and Heesterbeek's method detailed in [116]; that is, we consider the *elasticity* of biodiversity and the measures of risk of infection; a non-dimensionalisation of the derivative of biodiversity with respect to risk of infection. Mathematically, this is given by

$$\mathcal{D}y = \frac{x}{y}\frac{dy}{dx}. \tag{131}$$

Since we want to consider the effect of biodiversity on the risk of infection, we aim to evaluate three quantities: $\mathcal{D}R_0$, $\mathcal{D}E$, and $\mathcal{D}Y_c$. We evaluate the derivatives of the risk of infection with respect to biodiversity using the chain rule and numerical calculus. Since we are varying δ, we can find the derivative of biodiversity and risk of infection with respect to that parameter using finite difference methods, except in the case of R_0, which can be differentiated analytically. We then employ the chain rule to obtain our required derivative and, hence, the elasticity we seek. For example, the equation for the elasticity of the disease prevalence in cats Y_c is given by

$$\mathcal{D}Y = \frac{dY_c}{d(N_c/N_m)}\frac{N_c}{N_m Y_c} \tag{132}$$

$$= \frac{dY_c}{d\delta}\frac{d\delta}{d(N_c/N_m)}\frac{N_c}{N_m Y_c}. \tag{133}$$

We must note that we interpret the results differently based on the equilibrium values for N_c and N_m. In the regime $N_c < N_m$, a positive value for the elasticity indicates an increase in biodiversity with an increase in the chosen metric, and a negative value indicates a decrease. In the regime $N_c > N_m$, the opposite is true. This is because the most biodiverse ecosystem, according to our chosen definition, occurs where $N_c = N_m$, so an increase in biodiversity requires the ratio N_c/N_m to be closer to 1.

All that remains before obtaining results is for us to select sensible values of the other parameters in the model, keeping in mind our requirements for the non-endemic equilibrium to exist and the endemic equilibrium to be stable, as well as our context of toxoplasmosis in cats and mice. Our parameters were chosen as shown in Table 5, aiming to maintain consistency with ecological principles and making inferences on other parameters where required to give sensible behaviour. In particular, we consider that mouse density at equilibrium should be greater than cat density, and ensure that cats have a high degree of dependence on the mouse population for survival. Calibrating the model with experimental and observational data is scope for future work.

Parameter	Value	Dimension
μ_c	0.1	Cats/Area/Time
α_c	1	Cats/Area/Time
ω_c	0.1	Cats/Area/(Oocysts Time)
γ_c	0.5	Cats/Area/Time
ϕ_{cm}	0.01	Cats/Area/(Mice Time)
μ_m	50	Mice/Area/Time
α_m	0.5	Mice/Area/Time
ω_m	0.1	Mice/Area/(Oocysts Time)
γ_m	0.5	Mice/Area/Time
ϕ_{mc}	0.5	Cats/Area/(Mice Time)
ϕ_{mm}	1	Mice/Area/(Mice Time)
λ	100	Oocysts/Area/Cat
η	10	Oocysts/Area/Time

Table 5: Parameters used for the toxoplasmosis model. Unit dimensions are provided in lieu of specific calibration.

Using the parameters shown in Table 5, we obtain a basic reproduction number $R_0 = 1.74$, and non-endemic equilibrium values $S_c^* = 4.5$ and $S_m^* = 90$, which align with what would be considered qualitatively reasonable. The density of oocysts would be significantly larger in practice, but this can be handled through an arbitrary choice of scaling factor if and when required. Using these values, we find elasticity values for many increasing values of $\delta > 1$, which conveniently all converge to an endemic equilibrium, producing the result seen in Figure 26.

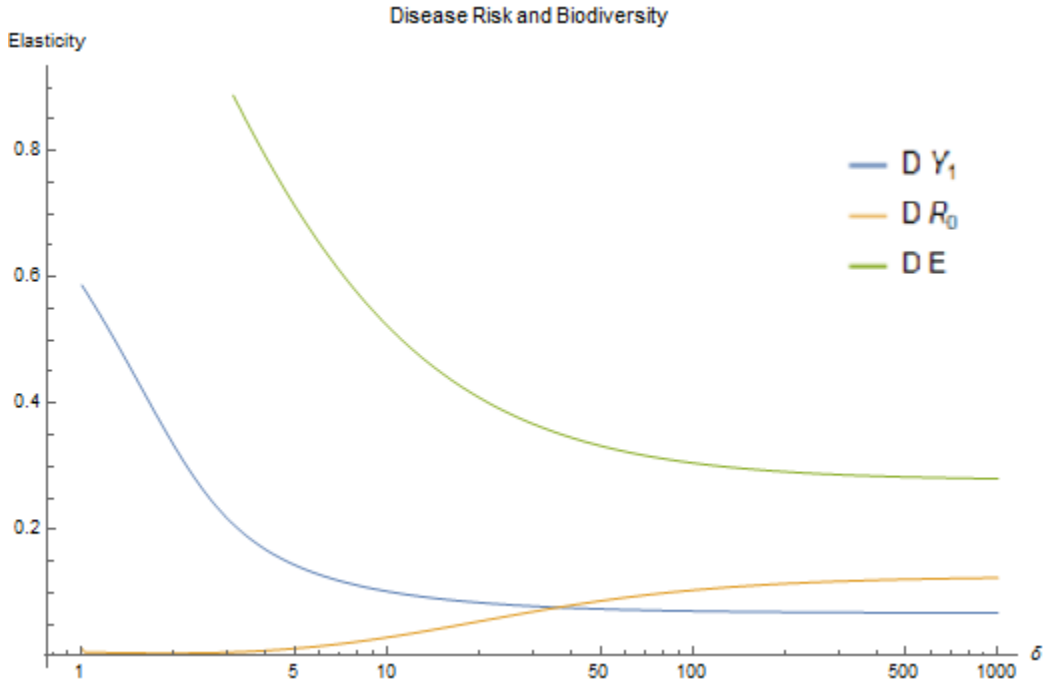

Figure 26: The elasticity with respect to δ for three different measures of infection risk.

We see that for all metrics, we observe an amplification effect rather than a dilution effect. This effect grows stronger as δ increases in terms of R_0 but grows less pronounced in the cases of the number of oocysts in the environment E and the relative prevalence of infected cats Y_c. This result contradicts the traditional idea of the dilution effect, where increased biodiversity decreases disease risk. It is important to note that if we make different choices for our parameters, the result does not hold, meaning we cannot generalise our results to all predator-prey systems. Exploring the parameter space more generally, we find cases where our metrics can give different results, changing from dilution to amplification or vice-versa as δ increases. This observation of dependence on metric is also made in the review by Salkeld et al., and in a recent modelling study by Gómez-Hernández et al. [49, 119].

Overall, the analysis for toxoplasmosis shows that where pathogens evolve beneficial traits that affect host behaviour, we can also see an amplification effect in a two-species system. The elasticity is positive for all metrics, showing that the result is insensitive to measurement choice. Consequently, such an evolution, that results in more food for predators outnumbered by prey and also more infection, can be considered to generally result in amplification. A crucial question regarding dilution and amplification is whether they are observed in naturally occurring systems [71], and this result shows a simple example of how natural processes can result in amplification.

While toxoplasmosis is not itself an emerging infectious disease, it is a relevant zoonosis in terms of public health. It is important to note that the emergence of an infectious disease is not predictable and that even transmission dynamics that do not correspond to recent epidemics may have future utility in describing unknown infections. Having examined toxoplasmosis, we move on to a vector-borne pathogen central to the dilution effect debate in Lyme disease.

7 Modelling Lyme Disease

Lyme disease, a tick-borne illness caused by bacteria of the genus *Borrelia*, is an important example in the debate surrounding the dilution effect and its generalisability. Throughout the 21st century, the tick species *Ixodes scapularis* has expanded its range on the east coast of North America, bringing with it *Borrelia burgdorferi*. Accompanying the ticks is an increase in infections emerging in new areas (such as Canada) that is projected to grow further with ongoing climate change [38, 100]. Owing to its public health relevance and relatively recent emergence, Lyme disease has been at the forefront of the discourse surrounding the dilution effect, and it is frequently cited as an example of the dilution effect in conservation biology.

Before examining views on the dilution effect as they relate to Lyme disease, we provide a brief breakdown of how it is transmitted. Infection is spread from ticks to vertebrate species through their bites, mainly in spring and summer. The tick vectors of Lyme disease exhibit a lifecycle with three stages: larva, nymph and adult. Larvae and nymphs prefer to feed on small species such as the white-footed mouse *Peromyscus leucopus*. Adult ticks prefer larger hosts such as the white-tailed deer *Odocoileus virginianus*. Nymphs cause the majority of human infections. In the absence of vertical transmission, younger larvae are uninfected until they feed on their first host before moulting, while adult ticks are both less likely to feed on humans and more likely to be detected and removed quickly when they do, owing to their substantially larger size.

The three-stage lifecycle of ticks and their different preferred hosts and infection rates create a complex ecoepidemiological system. Consequently, an age-structured agent-based approach for a system with a vector is impractical. Therefore, Lyme disease is ideally suited to an ODE-based model that can accommodate the complex infection dynamics.

7.1 A Base Model

As a basis for our analysis of the dilution effect as it applies to Lyme disease, we use the system of equations used by Nguyen et al. [97]. Their model uses three species and nine compartments, striking a desirable balance between detail in describing infection and demographic dynamics and ease of computation and mathematical analysis. The compartments in the model are

- L_S, the population of tick larvae; in the absence of vertical transmission [17], it is assumed all are susceptible
- N_S and N_I, the population of susceptible and infected nymphs

- A_S and A_I, the population of susceptible and infected adult ticks
- M_S and M_I, the population of susceptible and infected mice
- D_S and D_I, the population of susceptible and infected deer

Their model assumes logistic growth in mice and deer, with respective growth rates and carrying capacities r_M, r_D, K_M, and K_D. Tick larvae also have a growth rate and carrying capacity r_L and K_L. The growth rate accounts for the proportion of adults that obtain a blood meal and produce eggs and then the proportion of those eggs that go on to hatch into larvae, while the carrying capacity accounts for the supply of blood meals from the deer and mouse populations. As the ticks progress through their lifecycle, larvae and nymphs feed on mice and deer at contact rates β_L and β_N, respectively. It should be noted that this β, while still a contact rate, does not apply specifically to infection. This choice of nomenclature reflects the shared parasitic nature of the tick-host and Lyme-host relationships. There is also intraspecies competition between larvae to become nymphs, with a corresponding rate ϵ_L. Nymphs and adults have natural death rates δ_N and δ_A.

Infection dynamics use a simple SI system. There is no recovery, as untreated Lyme borreliosis persists in studies of rodent hosts [11, 94]. Infection in mice and deer as a result of their contact with infected nymphs or adults occurs at contact rates β_{NM}, β_{AM}, β_{ND}, and β_{AD}. Of course, larvae, nymphs and adults have different feeding patterns, with nymphs preferring mice and adults preferring deer, so scaling factors b_L and b_N account for the specific rates of interaction between host species at the relevant stage of the tick lifecycle. The parameters ψ_N and ψ_A describe the proportion of susceptible larvae and nymphs that become infected due to feeding on an infected host. For simplicity and in accordance with Nguyen et al.'s assessment of the epidemiological evidence, we assume hereafter that $\psi_N = \psi_A = 1$, i.e. all susceptible larvae and nymphs become infected after contact with an infected host. We retain the parameters for completeness.

$$L_S'(t) = r_L(A_S+A_I)\left(M_S + M_I + b_A(D_S + D_I)\right)\left(1 - \frac{A_S + A_I}{K_L\left(M_S + M_I + b_A(D_S + D_I)\right)}\right) - \beta_L L_S\left(M_S + M_I + b_L(D_S + D_I)\right) - \epsilon_L L_S^2 \quad (134)$$

$$N_S'(t) = \beta_L l_S\left(M_S + b_L D_S\right) + (1-\psi_N)\beta_L L_S\left(M_I + b_L D_I\right) - \beta_N N_S\left(M_S + M_I + b_N(D_S + D_I)\right) - \delta_N N_S \quad (135)$$

$$N_I'(t) = \psi_N \beta_L L_S(t) \left(M_I + b_L D_I\right) - \beta_N N_I \left(M_S + M_I + b_N(D_S + D_I)\right) - \delta_N N_I \quad (136)$$

$$A_S'(t) = \beta_N N_S \left(M_S + b_N D_S\right) + (1 - \psi_A)\beta_N N_S \left(M_I + b_N D_I\right) - \delta_A A_S \quad (137)$$

$$A_I'(t) = \psi_A \beta_N N_S \left(M_I + b_N D_I\right) + \beta_N N_I \left((M_S + M_I) + b_N(D_S + D_I)\right) - \delta_A A_I \quad (138)$$

$$M_S'(t) = r_M(M_S + M_I) \left(1 - \frac{M_S}{K_M}\right) - \beta_{NM} M_S N_I - \beta_{AM} M_S A_I \quad (139)$$

$$M_I'(t) = \beta_{NM} M_S N_I + \beta_{AM} M_S A_I - r_M(M_S + M_I)\frac{M_I}{K_M} \quad (140)$$

$$D_S'(t) = r_D(D_S + D_I) \left(1 - \frac{D_S}{K_D}\right) - \beta_{ND} D_S N_I - \beta_{AD} D_S A_I \quad (141)$$

$$D_I'(t) = \beta_{ND} D_S N_I + \beta_{AD} D_S A_I - r_D(D_S + D_I)\frac{D_I}{K_D}. \quad (142)$$

This model can be easily modified by adding extra species or by the variation of its parameters to determine the effect of changing biodiversity on infection risk. Nguyen et al. also provide values of each parameter, making it an ideal starting point for our purposes. Other ODE Lyme models exist, for example those of Wallace et al. used to assess the role of temperature on infection risk [108, 144]. These models use many more compartments by explicitly representing questing and engorged ticks and consider seasonality through periodic time-dependence of their parameters, making it difficult to consider measures of infection risk such as R_0.

For the various parameters in the ODE system, Nguyen et al. also set out values incorporating ecological data [97]. Their parameter choices are highlighted in Table 6 and were directly sourced or computed from modelling and observational studies [12, 33, 70, 84, 88, 142].

With our model and parameters established, we now consider its behaviour before extending it and considering the relationship between biodiversity and infection risk.

Parameter	Value	Units
b_N	0.003	Dimensionless
b_A	3.982	Dimensionless
r_L	0.0012	Larvae/ha/Day
r_M	0.05	Mice/ha/Day
r_D	0.1	Deer/ha/Day
K_L	1.15×10^5	Larvae/ha
K_M	50	Mice/ha
K_D	0.25	Deer/ha
β_L	6.55×10^{-6}	Larvae/ha/(Hosts Day)
β_N	1.36×10^{-5}	Nymphs/ha/(Hosts Day)
ϵ_L	1.24×10^{-8}	Larvae/ha/Day
β_{NM}	2.25×10^{-6}	Nymphs/ha/(Hosts Day)
β_{AM}	9.03×10^{-6}	Adults/ha/(Hosts Day)
β_{ND}	3.04×10^{-7}	Nymphs/ha/(Hosts Day)
β_{AD}	3.38×10^{-7}	Adults/ha/(Hosts Day)
δ_N	0.0022	Nymphs/ha/Day
δ_A	0.0027	Deer/ha/Day

Table 6: Parameters used for the Lyme disease model of Nguyen et al. [97].

7.1.1 Model Analysis

As we are working with a previously studied model as our basis, we briefly summarise its dynamics here in lieu of a detailed breakdown of steady states and their stability. A disease-free and endemic equilibrium can be found, each of which is well-defined and stable under the selected parameters. With this model, calculating the Jacobian gives a transmission matrix T and a transition matrix Σ given by the following:

$$T = \begin{bmatrix} 0 & 0 & \beta_L L_{S0} & \beta_L b_L L_{S0} \\ \beta_N(K_M + b_N K_D) & 0 & \beta_N N_{S0} & \beta_N b_N N_{S0} \\ \beta_{NM} K_M & \beta_{AM} K_M & 0 & 0 \\ \beta_{ND} K_D & \beta_{AD} K_D & 0 & 0 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \beta_N(K_M + b_N K_D) + \delta_N & 0 & 0 & 0 \\ 0 & \delta_A & 0 & 0 \\ 0 & 0 & r_M & 0 \\ 0 & 0 & 0 & r_D \end{bmatrix},$$

where the subscript S_0 denotes the population at the non-endemic (i.e. disease-free) equilibrium, calculated algebraically by letting $L_I = N_I = A_I = M_I = D_I = 0$. The non-endemic equilibria for non-ticks are equal to K_M and K_D and have been substituted into the matrices. Evaluating the next-generation matrix gives

$$T\Sigma^{-1} = \begin{bmatrix} 0 & 0 & \frac{\beta_L L_{S0}}{r_M} & \frac{\beta_L b_L L_{S0}}{r_D} \\ \frac{\beta_N(K_M + b_N K_D)}{\beta_N(K_M + b_N K_D) + \delta_N} & 0 & \frac{\beta_L N_{S0}}{r_M} & \frac{\beta_L b_L N_{S0}}{r_D} \\ \frac{\beta_{NM} K_M}{\beta_N(K_M + b_N K_D) + \delta_N} & \frac{\beta_{AM} K_M}{\delta_A} & 0 & 0 \\ \frac{\beta_{ND} K_D}{\beta_N(K_M + b_N K_D) + \delta_N} & \frac{\beta_{AD} K_D}{\delta_A} & 0 & 0 \end{bmatrix}.$$

The eigenvalues of this matrix are used to calculate the basic reproduction number R_0. Substituting each parameter and computing the spectral radius allows us to reproduce $R_0 = 1.28$, the same result obtained by Nguyen et al. [97].

For biodiversity in this system, we have three species, each of radically different sizes and numbers in the population. Consequently, we propose measuring biodiversity by normalising for biomass. In the case of the Simpson index, this is equivalent to the probability of a randomly selected infinitesimal unit of biomass matching another, and likewise for the entropy prediction of the Shannon index. Motivating this proposal is the fact that the calculus of the indices over a small step size becomes exceedingly cumbersome where a heavier species such as deer has vastly lower population density than a species like a mouse, which in turn has vastly lower density than a tick

species. Such contexts feature changes becoming so small as to allow rounding errors to dominate. Scaling for biomass allows for convenient measurement of the changes to biodiversity, from small changes to model parameters.

For the Simpson index, a biomass-adjusted formula would be given by

$$\lambda_m = \sum_{i=1}^{R} \left(\frac{n_i m_i}{M} \right)^2 \tag{143}$$

where m_i is the expected mass of a member of species i and $M = \sum_{i=1}^{R} n_i m_i$ is the total biomass of the system. We label this adjusted-index λ_m. In our context, we take the mean mass of adult ticks, mice and deer; the true expected values would be slightly lower but similar for each species.

Before adding a new species, we consider the effect of varying one of the model's parameters on biodiversity and our chosen risk of infection R_0 through the next generation matrix. Our procedure for doing this is to substitute the parameters with one varying at each step and use a central difference approximation to calculate the derivatives of both biodiversity and R_0 with respect to the varied parameter. R_0 can be computed directly, but any measure of biodiversity requires the system to be solved at each timestep. We aim to establish an ideal step size and order of central difference approximation to minimise error while maintaining a reasonable computation time.

7.2 Numerical Methods

Before adding extra species, we consider the behaviour of the approximation with the base model for various step sizes and degrees of approximation for the central difference formulae. We retain the use of the parameters set out by Nguyen et al. [97] and vary β_N, the contact rate for nymphs, between the values of 0 and 10^{-4}, at 50 evenly spaced points, calculating biodiversity at each point. These biodiversities use R_0 as their measure of infection risk and the Gini-Simpson index as their measure of biodiversity. Since the convergence time for the ODE system is variable, we use a very high value of $t_{max} = 10^7$, or 10 million days in the time-scale of the parameters, to be sure of convergence.

We repeat this process for different values of h, the step size used on either side of each point, and examine the convergence of the resulting biodiversity with respect to R_0 as h decreases. Similarly, we can also increase the degree of the central difference

approximation and examine the convergence of biodiversity. For a central difference approximation of degree n, a first derivative can be approximated numerically using the following:

$$f'(x) \approx \sum_{p=-n}^{n} a_p f(x+ph) + \mathcal{O}(h^n) \tag{144}$$

where the coefficients a_p are given by

$$a_p = \begin{cases} 0 & p = 0 \\ \frac{(-1)^{p+1}(n!)^2}{p(n-p)!(n+p)!} & p \neq 0 \end{cases}. \tag{145}$$

We utilise the Gini-Simpson index as our biodiversity metric to ensure that a lower elasticity signifies a dilution effect. In Figure 27, we see a dilution effect from changing β_n and hence increasing R_0, though this effect is lessened as the value of β_n increases.

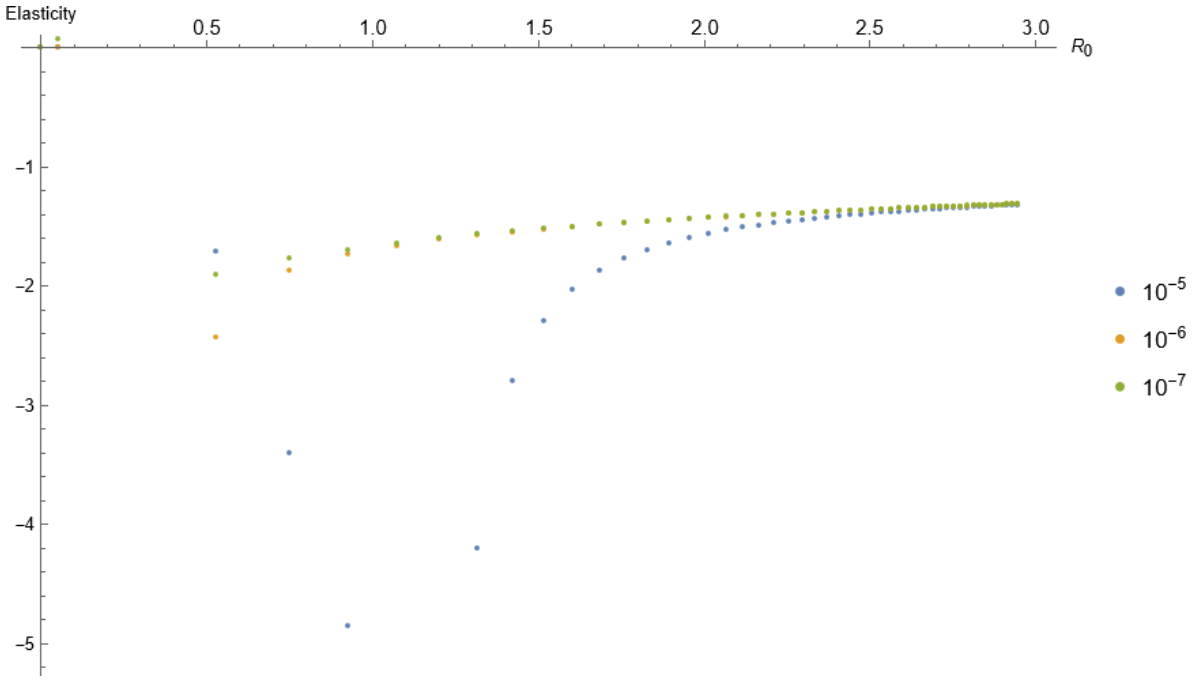

Figure 27: Elasticity with respect to R_0 when the nymph feeding contact rate β_n is varied, for various step sizes h, shown by the colour of each series of points. The central difference approximation used is first order.

We observe rapid convergence when reducing step sizes using a first-degree approximation. For all step sizes, the convergence is quicker where R_0 is higher, as seen

in the rapidly decreasing gaps between the points representing the different step sizes at equal R_0. Additionally, low step sizes include significant errors, particularly around $R_0 = 1$, representing the critical point between the endemic and non-endemic equilibria. As R_0 increases, we also note a decrease in the horizontal distance between points, indicative of the diminishing effect of an arithmetic increase of β_N on R_0. Lowering the step size reduces both the magnitude of this error and the range of R_0 values over which these errors are observed. For step sizes $h < 10^{-6}$, corresponding to one to two orders of magnitude lower than β_N over its range, we see highly stable results, with data points that appear indistinguishable and no visible compounding of errors relating to numerical methods or machine precision.

Increasing the degree of the central difference approximation also sees rapid convergence, similar to what is observed when varying the step size. We observe a dilution, as shown in Figure 28.

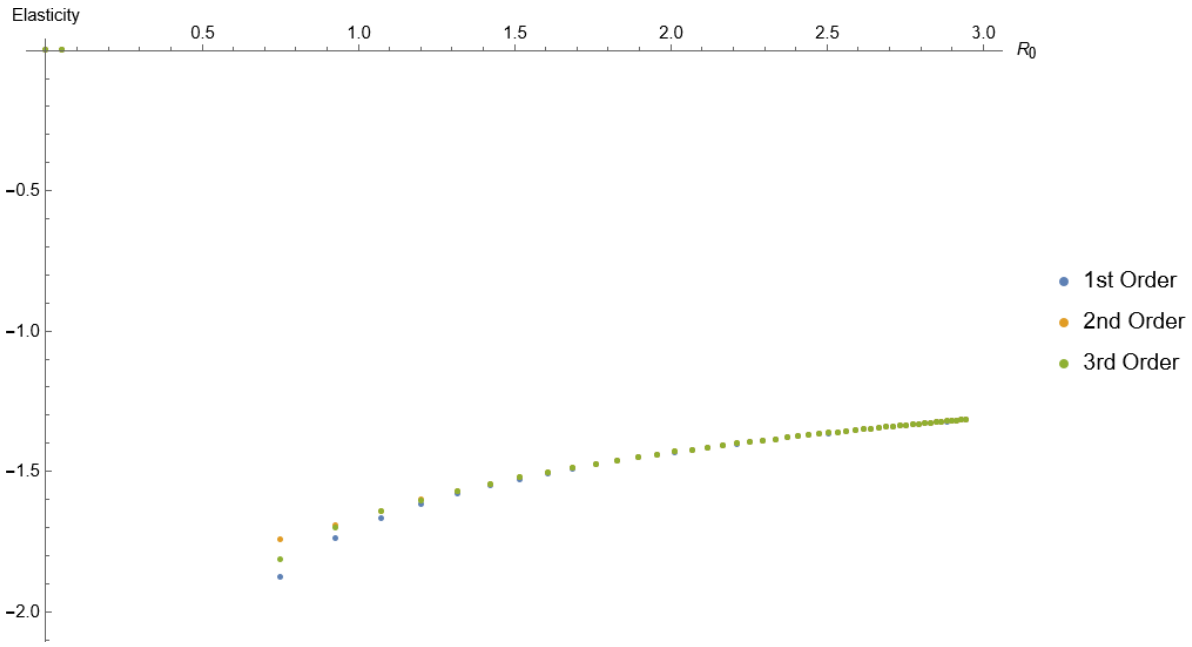

Figure 28: Elasticty with respect to R_0 when the nymph feeding contact rate β_N is varied, for 1st, 2nd and 3rd order central difference approximation, shown by the colour of each series of points. The step size for the central difference approximation is 10^{-6}.

With the step size held at 10^{-6}, an oscillatory behaviour can be observed, where even degrees appear to underestimate the magnitude of the true elasticity, while odd degrees appear to overestimate. This phenomenon can be attributed to the signs of the terms of the coefficients a_p, which alternate and begin on positive values for even degrees and negative values for positive degrees. Increasing the degree of approximation

also means incorporating a domain of β_N, meaning a greater buffer is needed above 0 for a reasonable approximation, as shown by the increased error close to $R_0 = 0$.

These trials give us a good idea of how our methods behave and inform our future choices of step size and approximation degree. They also provide an example of a dilution effect. However, it comes with the caveat of the overwhelmingly high ratio of ticks to mice and deer that trivialises the biodiversity indices. Considering biomass can aid in the pursuit of an alternative measure. Still, the primary purpose of the work in this section is to establish the behaviour and calibration of the central difference methods in calculating elasticities. As such, further discussion is withheld until future modelling provides a more pertinent setting.

7.3 The Effects of Interventions

Our analysis of the dilution thus far has concerned the steady states of ODE systems without regard to the initial conditions. In the real world, we can simulate the effects of human intervention to mitigate the risk of infection by setting up a set of parameters representing conditions before and after action. For example, we could simulate the culling of a species through an addition to the death rate, quarantine through the reduction of the contact rate, or environmental resource management by altering logistic terms.

Before expanding the Lyme disease model's scope, we can model the effect of intervention on deer or mice populations by culling rodents or deer. These oft-discussed measures have been hotly debated, particularly in relation to deer [80, 133]. Deer culling has been considered or taken by numerous jurisdictions on a regional level, including within Scotland, Ireland and the United States [52, 107, 138]. Despite significant discussion on the topic, the exact relationship between deer and tick abundances remains a key unknown in Lyme disease ecology [76]. In particular, we are interested in demonstrating how an action such as a deer cull that might be associated with a long-term reduction in infection risk might cause an increase in infection risk in the short term.

In the case of Lyme disease, infection risk is most closely associated with questing ticks, i.e., ticks lying in wait on plants for a passing host. Since all larvae in our model are considered uninfectious, the only potential routes to human infection are from nymphs and adults. Owing to their smaller size, bites from nymphs are more likely to go unnoticed, and therefore, nymphal infection prevalence (NIP) is a value of interest in assessing the risk of human infection from Lyme disease, though this is not

a definitively proven link [85]. Nymphal infection prevalence P_N is calculated by

$$P_N = \frac{N_I}{N_S + N_I}. \tag{146}$$

To model the culling of deer, we alter the model parameters laid out in Table 6 after the system reaches equilibrium. By lowering the carrying capacity of deer K_D, we can simulate a successful culling program that reduces the population to a target density. To lower the carrying capacity, we amend our model by substituting $K_D(t)$ where K_D appears, given by

$$K_D(t) = K_D - \Delta K_D H\left(t - \tau\right), \tag{147}$$

where the target density after culling begins at time τ is $K_D - \Delta K_D$, and H is the Heaviside step function, defined as

$$H(x) := \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \tag{148}$$

Altering K_D causes an eventual reduction in the deer population, but the short-term and long-term behaviour of infection risk can differ. Since the primary means of human infection is nymph bites, we consider two measures of infection risk: the nymphal infection prevalence and the population of infected nymphs. Each of these measures of infection risk is suggestive of a type of disease transmission. Considering nymphal infection prevalence ignores the effect of population density and addresses the risk of frequency-dependent transmission. In contrast, considering the total number of infected nymphs takes density into account and addresses the risk of density-dependent transmission.

We select $\Delta K_D = 0.15$, corresponding to a 60% decline in the deer population, at time $\tau = 10^5$. The long time until intervention is chosen for clarity in interpreting results in the future, rather than as a reflection of the time taken for the ODE solutions to converge. We normalise both infection risk measures by dividing each by their value at equilibrium before the cull. The effects of the deer cull on infection risk are shown in Figure 29. Immediately after the cull begins, no decline in NIP can be observed; a minute increase is apparent, suggesting that where frequency-based transmission dynamics are suspected, the anticipated decline in infection risk will take years to eventuate, preceded by a period of no significant change.

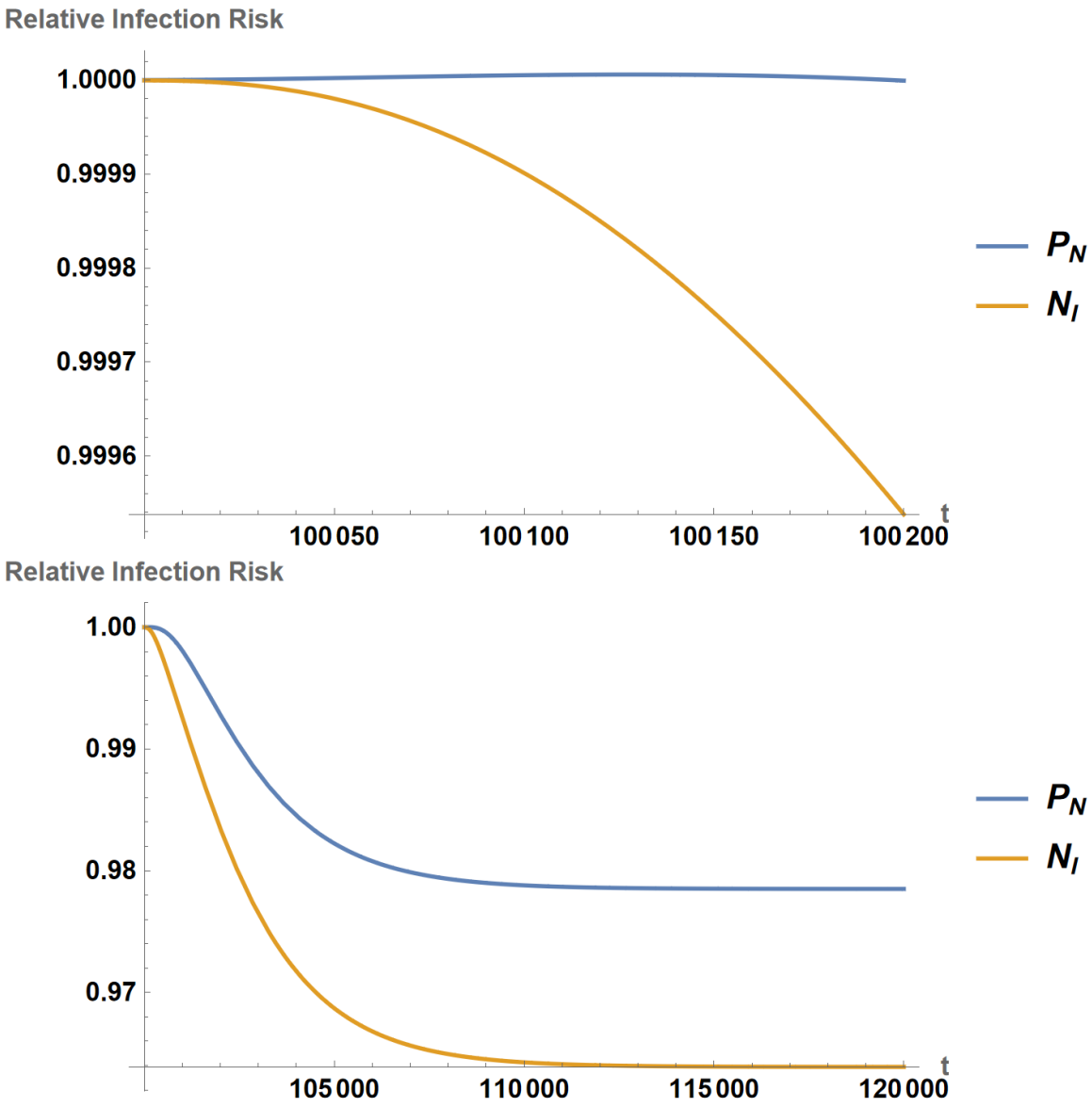

Figure 29: Nymphal infection prevalence P_N and total infected nymphs N_I after a cull reduces the deer carrying capacity by 60%. Two time scales are presented, one of 200 days and another of 20000.

Notably, the equilibrium decline in infection risks is no greater than 4% for either measure. This low dependence on the deer population suggests that mice alone can sustain a sizable tick population. Levi et al. [82] made similar findings, arguing with spatial data that deer recolonisation is not the dominant factor in the emergence of Lyme disease. It is essential to note that this result is not a conclusive argument against deer culls. The result is highly sensitive to model calibration. Specifically, it varies enormously with the ratio of the carrying capacities K_M and K_D and the scaling

factor representing the adult tick deer preference b_A. To illustrate, Figure 30 shows the impact of a 10^2-increase in b_A.

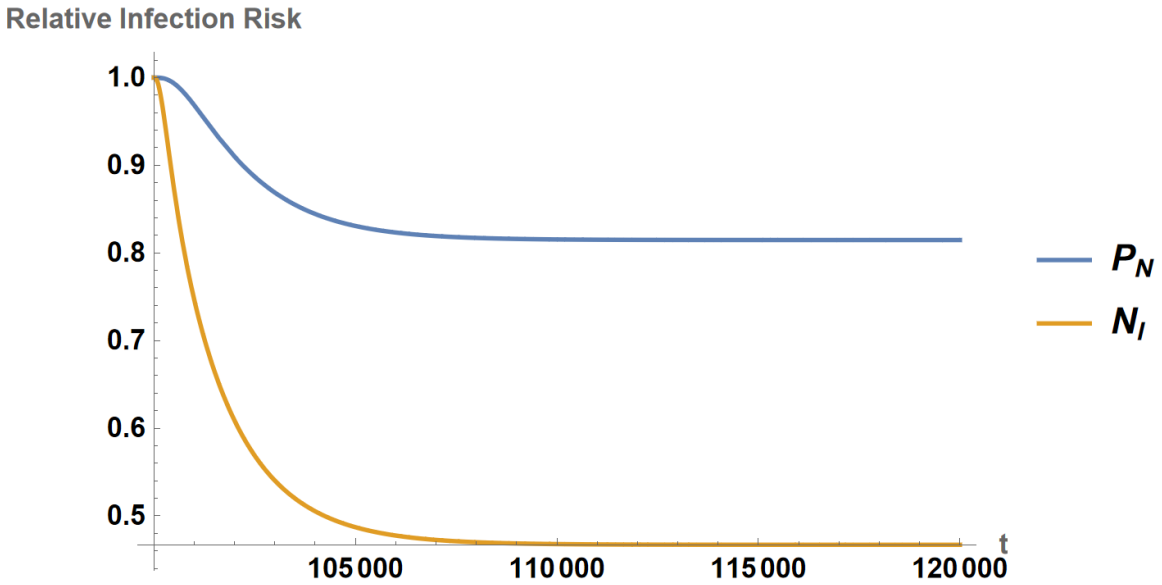

Figure 30: Nymphal infection prevalence P_N and total infected nymphs N_I after a cull reduces the deer carrying capacity by 60%, but where adult ticks more significantly prefer deer, with the change $b'_A = 100b_A$.

Where the adult tick deer preference is more pronounced, a deer cull is more effective in reducing disease risk, particularly in terms of the number of infected nymphs, which declines by more than 50%. The impact of the shift on NIP is comparatively muted but more substantial than what was seen for the lower b_A value. It is important to note that none of the parameter values are definitive; rather, they are the product of estimations applied to particular settings at particular times. Different locations and host species may have different parameter values that best describe them. Our results show the need to consider the specific ecosystem for policymakers aiming to cull deer to reduce Lyme disease risk.

As for biodiversity and the topic of dilution, a reduction in deer population precipitates a decrease in the tick population, meaning that the question of whether biodiversity increases is non-trivial. Since ticks and deer differ in size by many orders of magnitude, it is wise to consider alternative metrics that account for biomass rather than population. We use the masses in Table 7 when accounting for biomass.

Species	Mass (kg)	Source
White-footed Mouse	0.02	[8]
White-Tailed Deer	70	[65]
Black-legged tick (larva)	8×10^{-7}	*Estimated*
Black-legged tick (nymph)	4×10^{-6}	[58]
Black-legged tick (adult)	2×10^{-5}	[58]

Table 7: Masses of each species in our Lyme disease system. Sources are provided where available.

Adjusting for biomass creates a substantial change in results. Table 8 shows the Gini-Simpson indices before and after the cull.

Index	Before Cull	After Cull
$1 - \lambda$	$7.65 * 10^{-4}$	$7.74 * 10^{-4}$
$1 - \lambda_m$	0.844	0.247

Table 8: Biodiversity indices measured before and after culling 60% of deer. The indices used are the standard Gini-Simpson index $1-\lambda$, and the biomass-adjusted Gini-Simpson index $1 - \lambda_m$.

The table shows a marked difference. Where the Gini-Simpson index is used, the values before and after the cull are approximately equal to the minimum value of 0. There is a slight increase in the index, as the culling of deer affects the contribution to the index from the tick population more than that of the deer. A markedly different result is obtained when species mass is utilised, as the biomass-adjusted index decreases significantly. The values of the biomass-adjusted index are much higher, showing that the species each occupy a similar amount of the biomass in the ecosystem, highlighting the usefulness of the adjustment. As seen in Figure 31, the changes in both indices are monotonic.

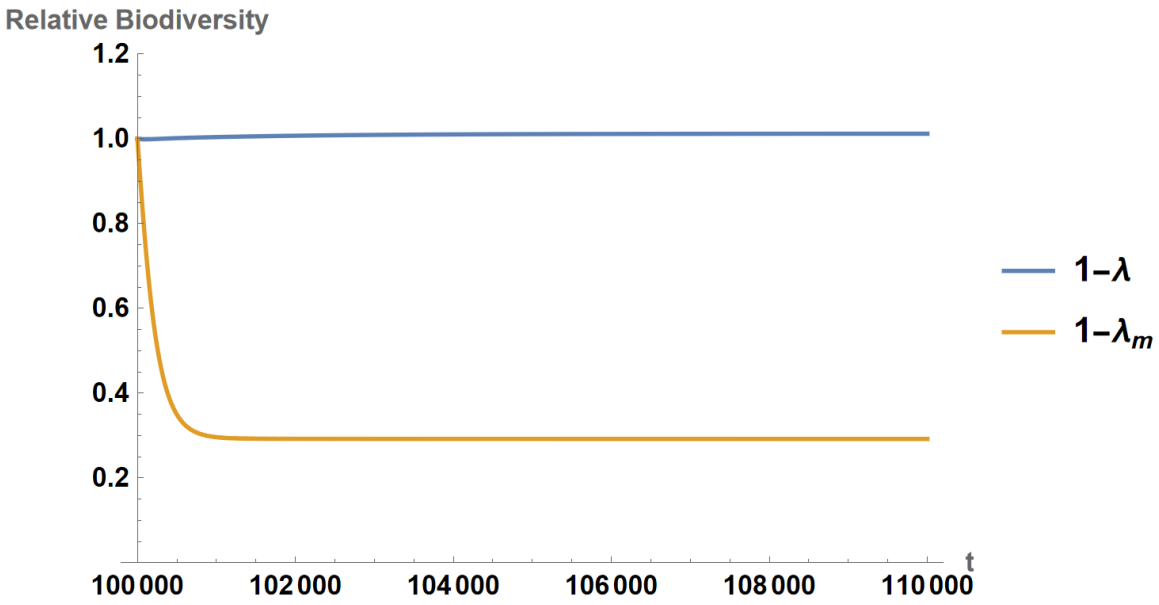

Figure 31: Relative biodiversity indices after the culling of 60% of deer.

In terms of dilution and amplification, we can note without further calculation that for a given number of infected nymphs N_I, there is a dilution effect without adjusting for species mass and an amplification effect when mass scaling occurs. The same conclusion holds for the NIP P_N, outside of the very earliest days at the beginning of the cull, where the reverse is true but with a relatively insignificant effect. The observation of a transition from dilution to amplification over time (or vice-versa) has been made in plant communities [131]. In conjunction with these studies, our results show that evolving dynamics over decades-long time scales should be an important consideration in studying diversity and disease relationships. We also note that observing a dilution and amplification effect is sensitive to whether species mass is considered, showing another way that dilution and amplification must be carefully defined.

Overall, our examination of the deer cull shows the role of the specific ecosystem carrying capacities in the efficacy of culling as a control strategy. Using the parameters set out by Nguyen et al. [97], the reduction of infection risk is so minor as to not justify the expense exclusively from a public health context regarding Lyme disease. However, there are substantial differences in communities where Lyme is prevalent, as evidenced through different results obtained in experiments in different locations [80].

Another interesting element of the results in Figures 29-31 is the role of time scale. While deer density reaches equilibrium in around two years once the carrying capacity is reduced, it takes around thirty years for the infection risk metrics to reach

equilibrium. This is well-aligned with the time scales of the relevant studies on deer and tick populations, though many of the longitudinal studies on Lyme incidence in these populations do not span three decades, suggesting a need for longer studies [67, 77, 130].

7.4 Introducing Competition

Having demonstrated the behaviour of our approximation methods in the context of our models, we now expand the scope of our modelling scenario beyond the basis model by introducing a fourth species. Since we are interested in investigating systems where two species compete for resources, this fourth species should compete with mice and/or deer. For Lyme Disease, a substantial body of literature compares the competencies of various rodents. In particular, Richter et al. demonstrate experimentally that *Borrelia spielmanii* infection manifests differently in various rodent species, with different incubation periods and levels of infectivity to susceptible nymphs exhibited over time [111]. We consider a case where two rodent species coexist, competing for food, with different competencies.

To examine the case of two competing rodents, we must amend our existing ODE model by introducing another species. We denote the second rodent species V (for vole, though the exact species remains hypothetical) and add two compartments, V_S and V_I, for a new total of 11 differential equations and variables.

$$L'_S(t) = r_L(A_S + A_I)\left(M_S + M_I + V_S + V_I + b_A(D_S + D_I)\right) \times \left(1 - \frac{A_S + A_I}{K_L\left(M_S + M_I + V_S + V_I + b_A(D_S + D_I)\right)}\right) - \beta_L L_S\left(M_S + M_I + V_S + V_I + b_L(D_S + D_I)\right) - \epsilon_L L_S^2 \quad (149)$$

$$N'_S(t) = \beta_L l_S\left(M_S + V_S + b_L D_S\right) + (1 - \psi_N)\beta_L L_S\left(M_I + V_I + b_L D_I\right) - \beta_N N_S\left((M_S + M_I) + b_N(D_S + D_I)\right) - \delta_N N_S \quad (150)$$

$$N'_I(t) = \psi_N \beta_L L_S(t)\left(M_I + V_I + b_L D_I\right) - \beta_N N_I\left(M_S + M_I + V_S + V_I + b_N(D_S + D_I)\right) - \delta_N N_I \quad (151)$$

$$A'_S(t) = \beta_N N_S\left(M_S + V_S + b_N D_S\right) + (1 - \psi_A)\beta_N N_S\left(M_I + V_I + b_N D_I\right) - \delta_A A_S \quad (152)$$

$$A_I'(t) = \psi_A \beta_N N_S \left(M_I + V_I + b_N D_I\right) + \beta_N N_I \left(M_S + M_I + V_S + V_I + b_N (D_S + D_I)\right) - \delta_A A_I \tag{153}$$

$$M_S'(t) = r_M \left(M_S + M_I - M_S \frac{M_S + M_I + \alpha_{MV}(V_S + V_I)}{K_M} \right) - \beta_{NM} M_S N_I - \beta_{AM} M_S A_I \tag{154}$$

$$M_I'(t) = \beta_{NM} M_S N_I + \beta_{AM} M_S A_I - r_M M_I \frac{(M_S + M_I) + \alpha_{MV}(V_S + V_I)}{K_M} \tag{155}$$

$$V_S'(t) = r_V \left(V_S + V_I - V_S \frac{V_S + V_I + \alpha_{VM}(M_S + M_I)}{K_V} \right) - \beta_{NV} V_S N_I - \beta_{AV} V_S A_I \tag{156}$$

$$V_I'(t) = \beta_{NV} V_S N_I + \beta_{AV} M_S A_I - r_V V_I \frac{(V_I + V_S) + \alpha_{VM}(M_S + M_I)}{K_V} \tag{157}$$

$$D_S'(t) = r_D (D_S + D_I) \left(1 - \frac{D_S}{K_D} \right) - \beta_{ND} D_S N_I - \beta_{AD} D_S A_I \tag{158}$$

$$D_I'(t) = \beta_{ND} D_S N_I + \beta_{AD} D_S A_I - r_D (D_S + D_I) \frac{D_I}{K_D}. \tag{159}$$

Here, α_{ij} is a constant describing the effect of competition of species j on the fecundity of species i. In this system, the first five differential equations describe tick dynamics, while the subsequent pairs of equations describe mice, the competing rodent and deer, in that order.

We make several assumptions in constructing a four-species model with age structure and logistic competition between mice and another rodent. We assume that ticks have no feeding preference between the rodents in all stages of their lifecycle, meaning that no further scaling factors b are required. Furthermore, in our analysis, we will assume that the two rodent species have identical birth rates and carrying capacities, i.e. $r_M = r_V$ and $K_M = K_V$. Additionally, we take the parameter $\alpha_{MV} = \alpha_{VM} = \alpha$, meaning that competition affects both species equally.

The two rodent species differ in their competence, i.e. their capacity to spread infection to ticks. We will assume that the transmission rates to a tick in stage i of its lifecycle $\beta_{iV} = C\beta_{iM}$, where $0 \leq C \leq 1$ is a scaling constant equal for all lifecycle stages. The upper bound of 1 is chosen since the competing rodent V is a less competent host by definition. These selections leave two free parameters: C and the relative effect of interspecies competition α.

7.4.1 Model Behaviour

Under these assumptions made in model construction, the subsystem of rodent competition can be assessed as a special case of the general two-species Lotka-Volterra system with competition. Owing to the lack of demographic interactions between rodents and the other species, we may consider these species in isolation. The coexistence equilibrium is given by $M = V = K/(\alpha + 1)$; the stability criterion can be derived from the linearisation of the system derived from the two eigenvalues, which are given by

$$\lambda_{1,2} = \frac{r_M \left(-1 \pm \alpha\right)}{\alpha + 1}. \tag{160}$$

Each eigenvalue is negative when $\alpha < 1$, but one becomes positive when $\alpha > 1$ when interspecies competition is stronger than the effect of intraspecies competition. Where $\alpha > 1$, both one-species equilibria are stable; the initial conditions determine the dominant species in that configuration. At $\alpha = 1$, the system asymptotically approaches a state given by $M + V = K$.

Figure 32 depicts phase portraits demonstrating these dynamics. Since we are interested in the stable coexistence of the rodents, in our analysis of this ODE system, we confine ourselves to the region of the parameter space where $0 < \alpha < 1$. The instability where $\alpha \geq 1$ is expected, owing to the equivalence to the two-species Lotka-Volterra system discussed in Section 2.2.3.

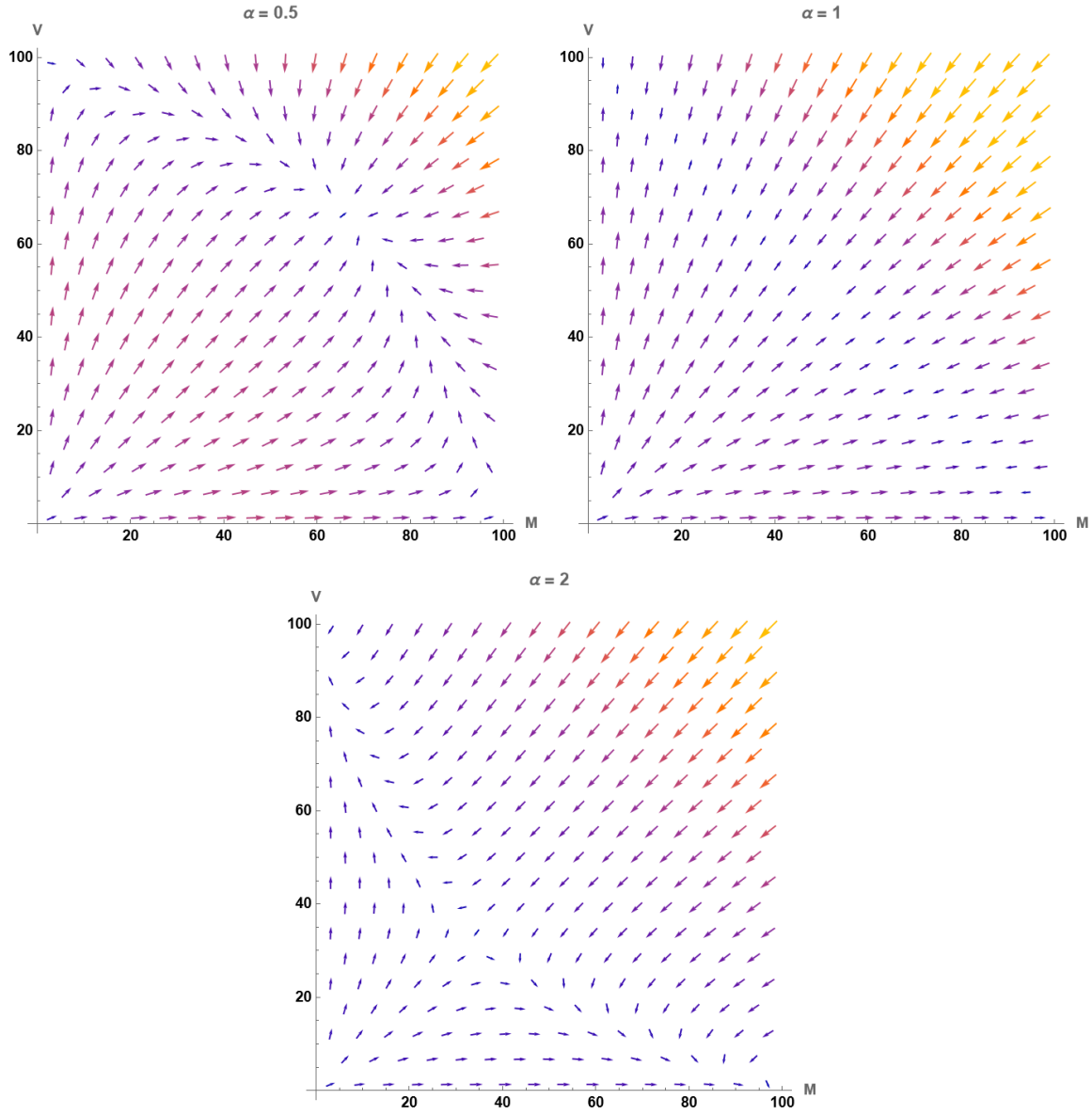

Figure 32: Phase portraits for two rodent competition with $r_M = r_V = 4$, $K_M = K_V = 100$. Note that the vector magnitudes (and colour scales, which are proportionate) are relative to the maximum for each plot.

We can compute the next-generation matrix to obtain an expression for the basic reproduction number. The non-endemic equilibrium for deer with all species present is K_D, while the tick population densities are satisfied by a system of simultaneous equations given by

$$L_{S0} = \frac{-c_L K_L \beta_L + \sqrt{\left(c_L K_L \beta_L\right)^2 + 4(1+\alpha) K_L R_L \epsilon_L A_{S0} \left(c_A K_L - (1+\alpha) A_{S0}\right)}}{2\left(1+\alpha\right) K_L \epsilon_L} \tag{161}$$

$$N_{S0} = \frac{c_L \beta_L L_{S0}}{c_N \beta_N + \left(1+\alpha\right) \delta_N} \tag{162}$$

$$A_{S0} = \frac{c_N \beta_N N_{S0}}{\left(1+\alpha\right) \delta_A}, \tag{163}$$

where the constants c_L, c_N and c_A are each given by

$$c_i = \left(1+\alpha\right) b_i K_D + 2K_M \tag{164}$$

for the lifecycle stage i. The constants describe the feeding populations of the other hosts at equilibrium, scaled by a factor of $1+\alpha$. With every parameter positive and real by definition, it follows that this equilibrium always exists and is positive. We produce equilibrium population densities for a given parameter set by solving these equations numerically.

The non-endemic equilibrium can be used to find the next-generation matrix. With the presence of competition, the Jacobian for the infectious states at the non-endemic equilibrium is given by

$$T + \Sigma = \begin{bmatrix} -\delta_N - c_N \beta_N \left(1+\alpha\right)^{-1} & 0 & \beta_L L_{S0} & \beta_L L_{S0} & b_L \beta_L L_{S0} \\ c_N \beta_N \left(1+\alpha\right)^{-1} & -\delta_A & \beta_N N_{S0} & \beta_N N_{S0} & b_N \beta_N N_{S0} \\ K_M \beta_{NM} \left(1+\alpha\right)^{-1} & K_M \beta_{AM} \left(1+\alpha\right)^{-1} & -r_M & 0 & 0 \\ C K_M \beta_{NM} \left(1+\alpha\right)^{-1} & C K_M \beta_{AM} \left(1+\alpha\right)^{-1} & 0 & -r_M & 0 \\ K_D \beta_{ND} & K_D \beta_{AD} & 0 & 0 & -r_D \end{bmatrix}. \tag{165}$$

The terms of the infectious-state Jacobian corresponding to interspecies competition are 0 at the non-endemic equilibrium. Therefore, the transition matrix $\Sigma = diag(T+\Sigma)$ is composed of the diagonal elements of $T + \Sigma$, i.e.

$$\Sigma_{ij} = \begin{cases} (T+\Sigma)_{ij} & i = j \\ 0 & i \neq j \end{cases} \tag{166}$$

Subtraction produces the transmission matrix T. Since Σ is a diagonal matrix, its inverse can be readily found, and the next-generation matrix is computed as $K = -T\Sigma^{-1}$.

$$K = \begin{bmatrix} 0 & 0 & \frac{\beta_L L_{S0}}{r_M} & \frac{\beta_L L_{S0}}{r_M} & \frac{b_L \beta_L L_{S0}}{r_D} \\ c_N \beta_N \left(1+\alpha\right)^{-1} & 0 & \frac{\beta_N N_{S0}}{r_M} & \frac{\beta_N N_{S0}}{r_M} & \frac{b_N \beta_N N_{S0}}{r_D} \\ \frac{K_M \beta_{NM}}{c_N \beta_N + (1+\alpha)\delta_N} & \frac{K_M \beta_{AM}}{(1+\alpha)\delta_A} & 0 & 0 & 0 \\ \frac{C K_M \beta_{NM}}{c_N \beta_N + (1+\alpha)\delta_N} & \frac{C K_M \beta_{AM}}{(1+\alpha)\delta_A} & 0 & 0 & 0 \\ \frac{K_D \beta_{ND}}{c_N \beta_N (1+\alpha) + \delta_N} & \frac{K_D \beta_{AD}}{\delta_A} & 0 & 0 & 0 \end{bmatrix}. \tag{167}$$

The basic reproduction number R_0 is equal to the spectral radius of K, i.e $R_0 = \rho(-T\Sigma^{-1})$, evaluated after computing the values of L_{S0} and N_{S0} after solving the simultaneous equations in Equations 161 to 163. With our analysis complete, we can now proceed to examining the relationship between biodiversity and infection risk.

7.4.2 Infection Risk and Biodiversity

We can compare infection risk and biodiversity metrics with and without the additional species to assess the impact of introducing a competing rodent species. Since $R_0 = 1.24$ for the three-species model, we can use the next-generation matrix to determine the region of the parameter space where R_0 is greater due to introducing a new species. With the knowledge that infection has no impact on the population dynamics of the species, we can also take the same approach for biodiversity measures. By examining the regions where R_0 and the biodiversity metrics change, we can consider a dilution or amplification brought about by the change in species richness, as shown in Figure 33.

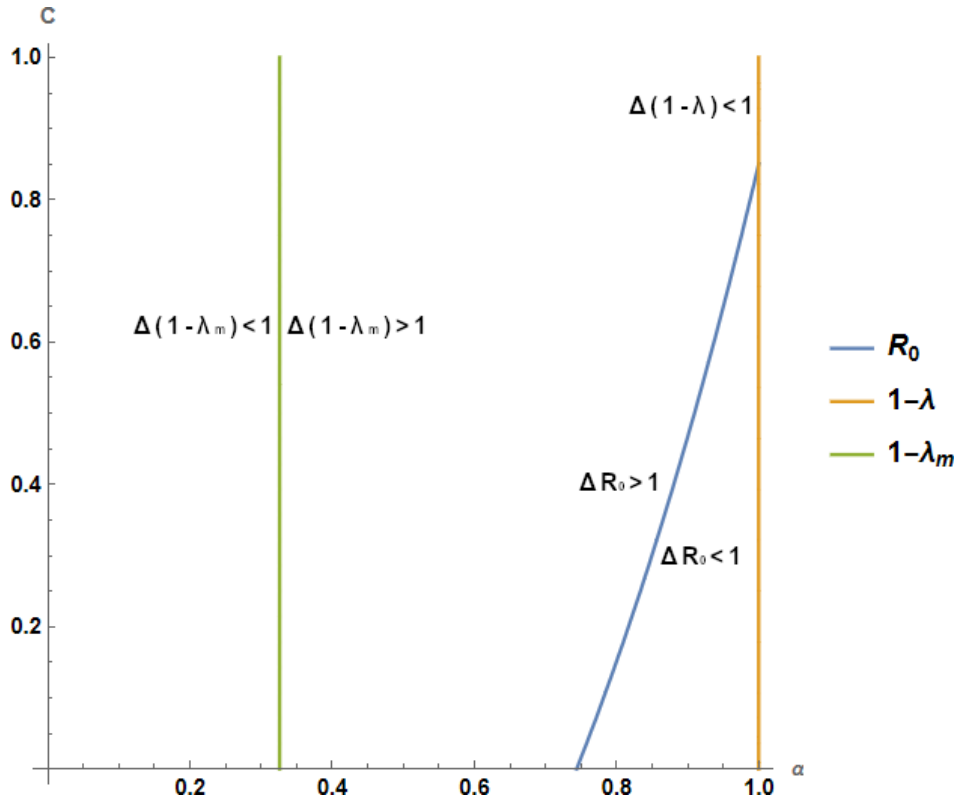

Figure 33: Contours in the α-C plane where R_0 and the Gini-Simpson indices are equal to the three-species case, assuming that $m_v = m_m$. The parameter α describes the relative effect of interspecies competition compared to intraspecies competition for both rodent species, while C describes the degree to which the competing rodent species is less competent than mice. The labels show regions where the respective quantity is increased ($\Delta > 0$) or decreased ($\Delta < 0$) with the addition of another rodent species. Note that for $1 - \lambda$ to be equal, α is not exactly equal to 1.

The labels on Figure 33 show that R_0 increases except in the small part of the region where α is sufficiently high and C sufficiently low. Notably, for sufficiently high C, there will always be an increase in R_0 from introducing a new species where the coexistence equilibrium is stable.

Whether these results are interpreted as dilution or amplification is sensitive to the choice of biodiversity metric. Where biomass is accounted for, the Gini-Simpson index $1 - \lambda_m$ increases for higher α levels to the right of the contour. Where this increase coincides with a reduction in R_0, we have dilution and amplification where R_0 is increased. To the left of the contour, biodiversity is decreased, and R_0 is increased, meaning dilution is observed. For the standard Gini-Simpson index $1 - \lambda$, biodiversity decreases to the left of the contour, showing an amplification where R_0 decreases and

a dilution where it increases. The opposite holds in the very small region to the right of the contour where $\alpha < 1$ still holds.

Since calculating exact solutions for all compartments with infection is impossible, we utilise numerical methods to consider NIP and the number of infected nymphs.

We consider all quantities within our parameter space for four different values of C, $\{10^i\}$, $i = \{-3, -2, -1, 0\}$. We use a maximum time of 10^7 days, a step size of $h = 10^{-5}$, and a second-degree approximation of the first derivative. For each value of C, we consider 100 equally spaced values of α from 0 to 0.99. The endpoint $\alpha = 1$ is excluded. Other parameters are identical to those presented in Table 6, with the parameters for the competing rodent species identical to that of the white-footed mouse.

To consider the effects of species richness, we can calculate NIP and the number of infected nymphs with one rodent species to establish a comparison to the system with two rodent species. Varying α for different values of C produces Figure 34. In all cases, the measure of infection risk decreases with α and C, with the measures of infection risk intersecting the value for the one-rodent case in the interval $0 < \alpha < 1$. This intersection value can be considered the transition between amplification, where adding a new species increases the chosen measure of disease risk, and dilution, where the opposite is true.

Under the given parameters, the equilibrium values for NIP and total infected nymphs are given by $Y_N^* = 0.360$ and $N_I^* = 4708$. These values provide the thresholds for dilution and amplification for the relevant measure of infection risk.

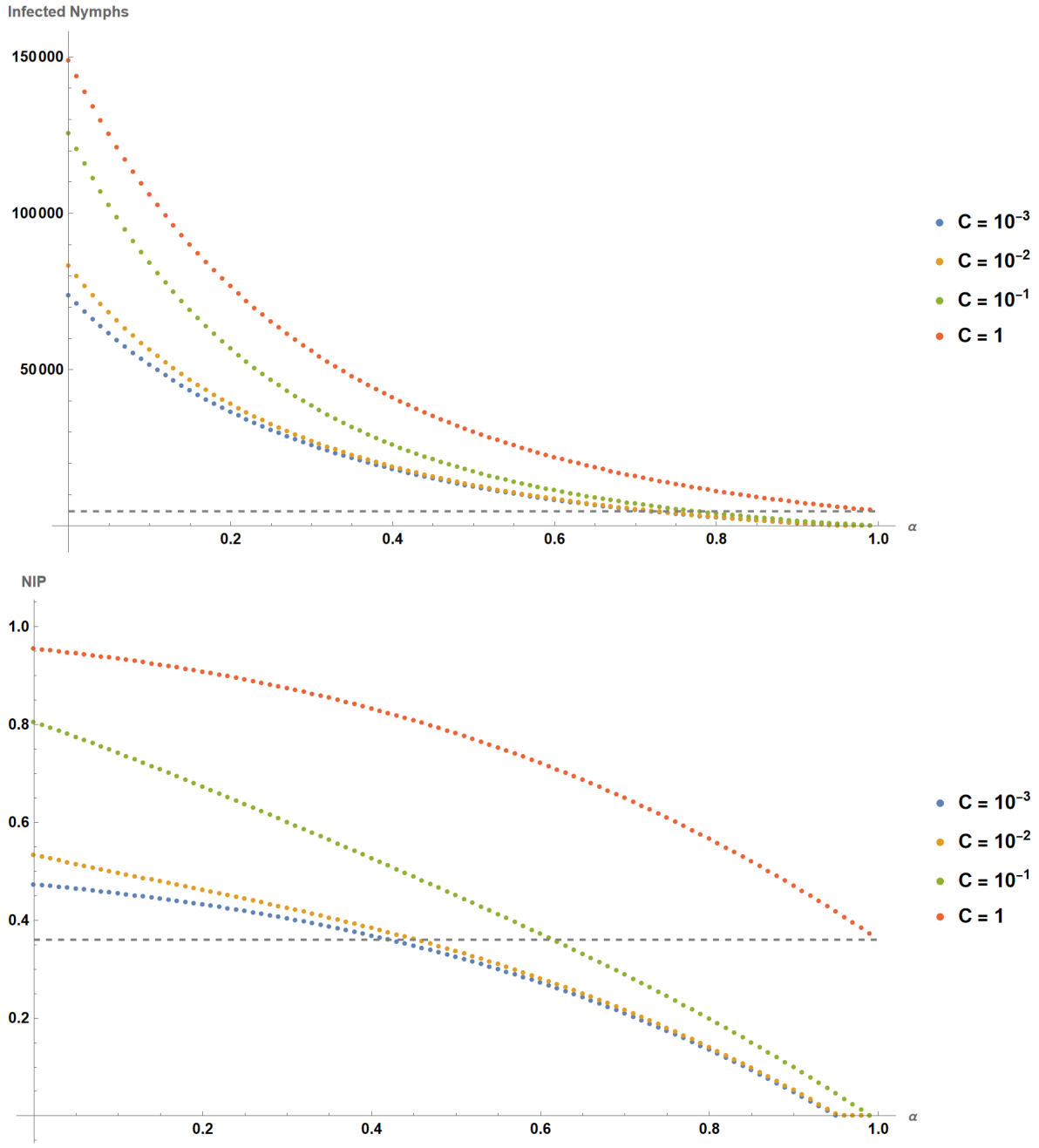

Figure 34: The measures of infection risk with respect to α for different values of C with two competing rodents. The measures of infection risk are total infected nymphs (above) and nymphal infection prevalence (below). The dotted line shows the value for each measure in the one rodent space using the parameters in Table 6.

Our parameter space spans a region where $R_0 < 1$, and Lyme disease is not endemic, as the number of infected nymphs goes to 0. This is observed where $C = 10^{-2}$ and

$C = 10^{-3}$.

Comparing NIP and the total number of infected nymphs, the values of α and C where amplification is observed are greater for R_0 and total infected nymphs than for NIP. In epidemiological terms, this observation implies that a dilution effect is more likely where Lyme disease transmission to humans is frequency-dependent, an observation also made by Keesing et al. [70]. Notably, given that the vector depends on the hosts for its reproduction, even adding a significantly less competent host will drive up infection risk when interspecies competition is not sufficiently robust.

This result is comparable to the spatially explicit agent-based model in the earlier chapter. However, we have yet to explicitly calculate biodiversity using an index, having employed only species richness. We can further examine the system by calculating a biodiversity index for a more complete picture. Since altering the scaling factor C impacts the infection dynamics and not the relative abundances of the species, the derivative with respect to C of any measure of biodiversity will be 0, resulting in an undefined elasticity. As a result, we find elasticities by varying α for different values of C, using the Gini-Simpson index as a measure of biodiversity.

The results of this process for both the NIP and the number of infected nymphs are presented in Figure 35. There is an amplification effect in all cases when the degree of interspecies competition α is increased. The parameter space captures values where $R_0 < 1$ and the infection dies out, as can be seen where lower values of C coincide with higher values of α. The elasticity in those regions appears to be 0 owing to errors of the magnitude of machine precision, where the measure of infection risk is effectively zero, but their derivatives vary. The primary reason behind the observation of an amplification effect is that any decrease in rodent populations will decrease the Gini-Simpson index, where biomass is not considered.

The change in elasticity is different for any given value of C and can have local maxima or a monotonic decrease depending on the infection metric and the value of C. Crucially, the elasticity is not a comparison that accounts for adding a new species. It merely describes the relationship between biodiversity and infection risk with respect to a shift in a particular parameter. Possible real-world interpretations are an infection control strategy that forces multiple pre-existing rodent hosts into greater competition or conflict for territory or feeding sites without affecting intraspecies competition. This premise differs from introducing a competing species as previously examined.

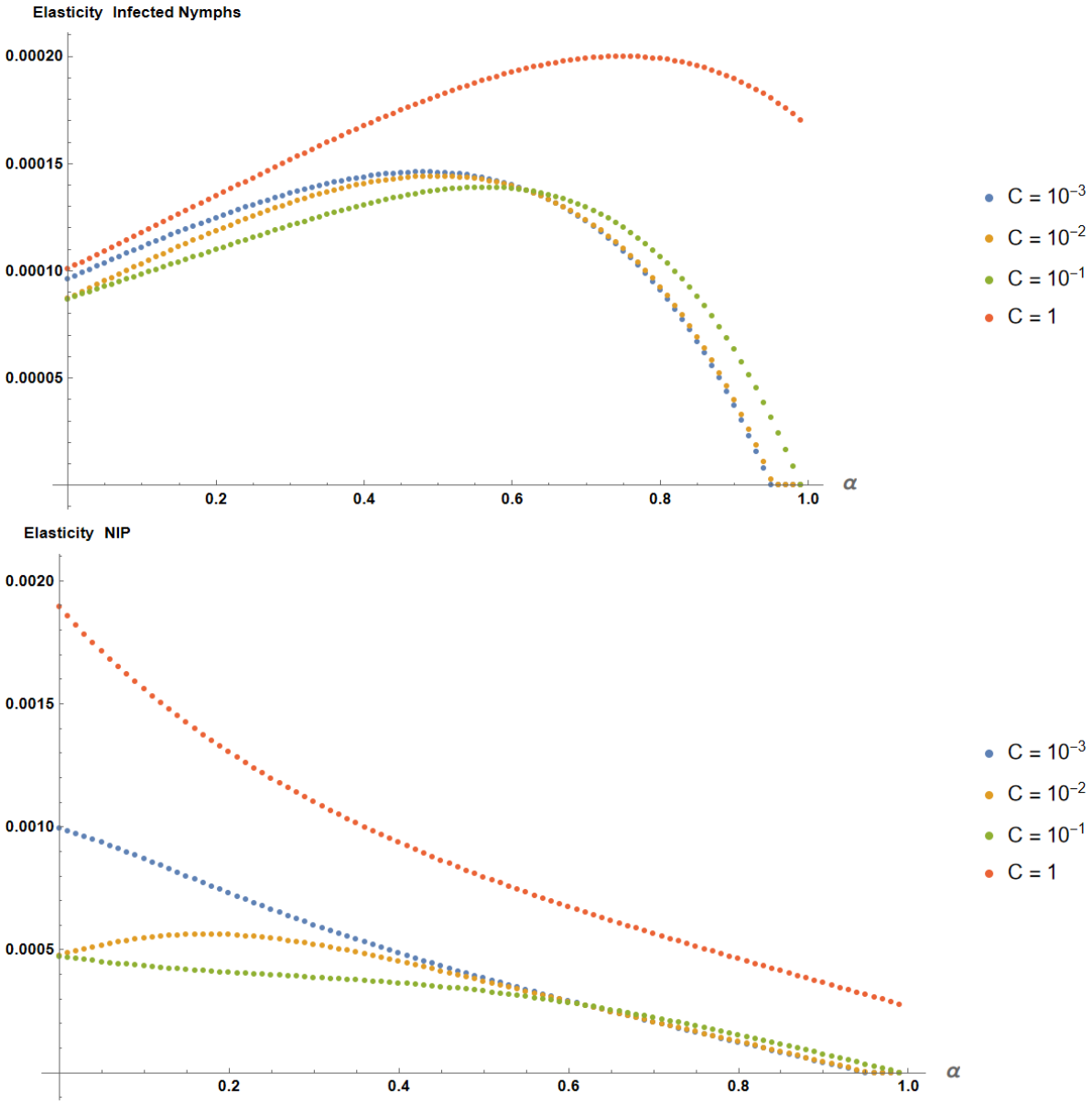

Figure 35: Elasticities of NIP and infected nymphs with varying interspecies competition α for different values of the infection scaling factor C. The measures of infection risk are total infected nymphs (above) and nymphal infection prevalence (below).

It is essential to consider that the relationship between biodiversity and infection risk depends on the varied parameters and is not a universal attribute of a system. Shifting a different parameter, such as the contact rate of nymphs earlier, may result in a different observation of dilution and amplification, despite the system being otherwise identical. Hence, the analysis of the question of dilution in one case must not be

generalised and will always depend on system dynamics with respect to the parameter varied. Our results are well-aligned with those of Ratti, Winter and Wallace, who used a different Lyme disease model and considered host preference as a choice between different functions describing tick feeding, and found both dilution and amplification, depending on the population of a more competent host [108].

The ODE-based Lyme disease model does have limitations, however. Ecosystems in the real world with *Ixodes scapularis* typically have a higher species richness than three or four [19], and the interspecific relationships formed may develop according patterns beyond what our models can capture. We also do not include disease-induced mortality in the model, despite some studies indicating that it plays a role in reducing Lyme disease incidence in humans [4].

Our Lyme disease model shows the calculation of elasticity in systems with multiple species using numerical techniques. Our analysis shows how the relationship between biodiversity and infection risk is not bound by one causal link. Observing a dilution or amplification effect depends strongly on how biodiversity and infection risk are defined and the specific relationship between both metrics and the corresponding change in the ecosystem [116, 119]. One notable intervention that has featured prolifically in the debate is the culling of deer, and our results show how the efficacy of deer culls as a public health strategy depends on the relative importance of deer to ticks as both a food source and a reservoir of infection. This work is hoped to inform the ongoing debate about biodiversity and Lyme disease and support decisions surrounding public health measures that may take place to mitigate infection risk.

8 Conclusion

In examining the ecology of infectious diseases, we aimed to address the role of scale and spatial dynamics in the observation of dilution and amplification effects. Despite the debate on the subject, the bulk of the literature concerning the dilution effect was focused on ODE-based models that assumed homogeneous mixing [49, 116]. In addressing this, we developed two types of agent-based model, each taking a different approach to modelling an ecosystem and having its own benefits and liabilities.

The lattice-based model we developed is lightweight, and handles its ecological and epidemiological processes without substantial computational burden. It can be adapted to account for a number of different interspecies and intraspecies interactions, though we primarily discussed a predator-prey system in our analysis. Its principal advantage over an ODE-based approach is in its ability to have a spatial component to births and deaths, as well as to infections and recoveries, producing spatially heterogeneous dynamics. We also demonstrated in an agricultural setting that an observed dilution effect can stem from multiple possible causes, including complementarity and decreased interspecies transmission, in line with expectations informed by existing discussion of dilution [50, 106, 148].

In contrast, the spatially explicit model we developed comes with a larger computation burden. It models agent movement using a correlated random walk, while tracking the position of each agent in continuous space. Infection is handled by directly calculating the number of infectious individuals within a certain range, while in our analysis we mostly focus on the time scale where the effects of demography are negligible, as handling demography in the same level of detail as infection serves to undesirably increase computation time. In exchange, the direct handling of space allows for more detail in boundary conditions and dynamics, and in setting probability distribution functions for epidemiological properties such as recovery time.

In general, we observe that the ability to handle demography quickly while scaling $\mathcal{O}(N)$ for the number of agents makes the lattice-based approach more fitting for larger systems on longer time-scales. Where more detail is required on a shorter time-scale, or with fewer agents, our $\mathcal{O}(N^2)$ spatially explicit approach may be preferable. Our work also adapts existing techniques for numerical settings, including the use of ecology quantities such as the Simpson and Shannon indices to quantify biodiversity.

We also use numerical methods to quantify the dilution effect by implementing spline interpolation and calculating elasticity. We also use the dynamics of the earliest phase

of an outbreak in an agent-based model to estimate the basic reproduction number R_0. The validity of assuming that the initial growth of an outbreak is exponential holds only when the parameters are selected such that the infection does not fully infect the population in the local area before spreading beyond that vicinity.

In terms of the generality of the dilution effect, we establish a number of insights about the relationship between disease risk and biodiversity, particularly it relates to the choice of definition for each. For a predator-prey model on a lattice, we establish that even when most metrics of infection risk decrease with biodiversity, the situation can be more complex when considering infection or prevalence in a host of significance. We also find that introducing a more competent host into a system with a constant total population does not necessarily result in amplification when the rate of interspecies transmission is lower than the within species rate. This metric-sensitivity in agent-based systems corroborates similar results in ODE systems [116].

Overall, our results and modelling strongly support the assertion that the dilution effect is an idiosyncratic phenomenon that can not be relied upon in any given scenario without careful analysis of the mechanisms involved, as described by Salkeld et al. [119]. For example, one agent-based predator prey model showed predominantly dilution when the predator death probability was changed, but in an ODE context for toxoplasmosis in two species, the evolution of the loss of predator aversion by infected prey led to an amplification effect.

The most important argument supported by our data is that the dilution effect is highly dependent on the choice of the measure of infection risk and biodiversity and the relevant region of the parameter space [108, 116]. Some in the literature have suggested that the lack of clarity in the metrics used in arguing for a dilution effect is a significant driver of misunderstanding, and our results support this conclusion [106]. Our analysis of ODE systems focused on Lyme disease and toxoplasmosis, infections with important distinctions in how they are transmitted, found mixed results. This leads us to suggest that any discussion of a dilution or amplification effect should be accompanied by a prominent statement on the choice of measure. This is evidenced by our analyses throughout, many of which provided different results for particular selections. As in experimental and observational studies, selection and publication bias must be considered as a factor when arguing for or against generality using modelling [119, 120].

Of course, modelling as a means of investigation as its limitations, and it is important to acknowledge these shortcomings. Real-world ecosystems often feature a higher degree

of species richness than the maximum in one of our models and capturing all of the interspecies relationships in one system of equations can pose a challenge when suitable data is not readily available. Furthermore, it has been claimed recently that vulnerable species to biodiversity loss are more likely to be diluting hosts, making biodiversity loss a more likely source of dilution observations [50, 71]. Since our models simulate dynamics for known community compositions, they are ill-suited to addressing these questions. Our models also do not account for pathogen or ecosystem evolutionary adaptation to changes driving dilution or amplification, as some experiments have studied [131].

Aside from its evidence for idiosyncratic dilution, the importance of our work is in how it can be used to inform the ongoing progression of knowledge about the dilution effect, and to ecoepidemiology in general, alongside experiemental and observational data. Extending existing methods such as the next-generation matrix [36, 115] and quantifying the dilution effect [116] as an elasticity to agent-based models and complex systems that require numerical analysis adds to the tools available and informs future academic discussion.

8.1 Future Work

The work presented here leaves substantial room for further research. In particular, the agent-based models can be updated to account for different dynamics and add additional functionality. For the spatially explicit model, trialling numerous approaches to handle demography and transmission that increase the computational efficiency of the model would make a substantial difference to the practicality of the approach. Explicit implementation of resources in the environment and movement biased to resource-rich locations is a possible extension.

Further study of the lattice-based model outside of a predator-prey context is possible, applying it to logistic competition systems, or on a more complex food web. The lattice-based model is also ideal for addressing the spatiotemporal patterns of infection, and study of the patterns formed by the spread of the infection may yield interesting results.

All of the agent-based models developed for the thesis could also be applied outside of the context of ecoepidemiolgy. Rather than modelling a pathogen circulating among multiple species, it would also be possible to investigate systems where one species is a host of two or more cocirculating pathogens that confer a degree of cross-immunity between them. Cross-immunity is a phenomenon of great interest in epidemiology, and these tools could prove a useful companion to the deterministic ODE models that

already exist.

In an ODE context, our analysis predominantly concerned systems where a stable stationary state exists, but Lotka-Volterra systems are capable of a wide range of dynamics. Considering the dynamics of an infection within an n-cycle and the relationship between population peaks and infection peaks might make for a useful predictive tool for risk of zoonotic disease spread.

Outside of Lyme disease and toxoplasmosis, there are countless other pathogens and systems that can be analysed. With the 21st century featuring two major epidemics and pandemics as a result of coronaviruses, consideration of respiratory infections in ecologies in line with evidence of their reservoirs is imperative. A major emerging infection of the 20th century was the Ebola virus, with its reservoir thought to be in fruit bats, though it is known to infect multiple species. Modelling transmission in those ecosystems could inform control measures in the regions where the virus is endemic. Overall, our work has applications in many fields, and will be useful in informing future research in ecoepideiology.

9 Bibliography

References

[1] L. J. S Allen. *An introduction to stochastic processes with applications to biology.* CRC Press, 12 2010.

[2] L. J.S. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2:128, 5 2017.

[3] T. F. A. Alves, G. A. Alves, A. Macedo-Filho, and R. S. Ferreira. Epidemic outbreaks on random delaunay triangulations. *Physica A: Statistical Mechanics and its Applications*, 541, 1 2019.

[4] N. S. Andersen, S. Skarphédinsson, F. C. Knudtzen, C. R. Olesen, T. G. Jensen, and P. M. Jensen. Reduction in human Lyme neuroborreliosis associated with a major epidemic among roe deer. *Ticks and Tick-borne Diseases*, 9:379–381, 2 2018.

[5] E. Arashiro and T. Tome. Threshold of coexistence and critical behavior of a predator-prey cellular automaton. *Journal of Physics A: Mathematical and Theoretical*, 40:887–900, 7 2006.

[6] S. S. Artemiev and T. A. Averina. *Numerical analysis of systems of ordinary and stochastic differential equations.* De Gruyter, 1 1997.

[7] K. C. Atwood, K. C. Vercauteren, T. J. Deliberto, H. J. Smith, and J. S. Stevenson. Coyotes as sentinels for monitoring bovine tuberculosis prevalence in white-tailed deer. *The Journal of Wildlife Management*, 71:1545–1554, 7 2007.

[8] E. Austin and C. Anderson. Population dynamics of white-footed mice (peromyscus leucopus) from 2012-2018 in a fragmented landscape. *Proceedings of the National Conference on Undergraduate Research*, 2019.

[9] J. Bao, X. Mao, G. Yin, and C. Yuan. Competitive Lotka–Volterra population dynamics with jumps. *Nonlinear Analysis: Theory, Methods and Applications*, 74:6601–6616, 12 2011.

[10] A. Barabási. *Network Science.* Cambridge University Press, 2016.

[11] S. W. Barthold, M. S. de Souza, J. L. Janotka, A. L. Smith, and D. H. Persing. Chronic Lyme borreliosis in the laboratory mouse. *The American Journal of Pathology*, 143:971, 9 1993.

[12] G. O. Batzli. Population dynamics of the white-footed mouse in floodplain and upland forests. *American Midland Naturalist*, 1977.

[13] M. Berdoy, J. P. Webster, and D. W. Mcdonald. Fatal attraction in rats infected with toxoplasma gondii. *Proceedings of the Royal Society B: Biological Sciences*, 267:1591–1594, 8 2000.

[14] E. L. Berlow, A. M. Neutel, J. E. Cohen, P. C. De Ruiter, B. Ebenman, M. Emmerson, J. W. Fox, V. A. A. Jansen, J. I. Jones, G. D. Kokkoris, D. O. Logofet, A. J. Mckane, J. M. Montoya, and O. Petchey. Interaction strengths in food webs: issues and opportunities. *Journal of Animal Ecology*, 73:585–598, 5 2004.

[15] A. A. Berryman. The origins and evolution of predator-prey theory. *Ecology*, 73:1530–1535, 1992.

[16] I. M. Bomze. Lotka-Volterra equation and replicator dynamics: A two-dimensional classification. *Biological Cybernetics*, 48:201–211, 1983.

[17] E. M. Bosler, B. G. Ormiston, J. L. Coleman, J. P. Hanrahan, and J. L. Benach. Prevalence of the Lyme disease spirochete in populations of white-tailed deer and white-footed mice. *The Yale Journal of Biology and Medicine*, 57:651–659, 1984.

[18] M. J. E. Broekman, H. C. Muller-Landau, M. D. Visser, E. Jongejans, S. J. Wright, and H. de Kroon. Signs of stabilisation and stable coexistence. *Ecology Letters*, 22:1957–1975, 11 2019.

[19] J. L. Brunner, L. Cheney, F. Keesing, M. Killilea, K. Logiudice, A. Previtali, and R. S. Ostfeld. Molting success of Ixodes scapularis varies among individual blood meal hosts and species. *Journal of Medical Entomology*, 48:860–866, 7 2011.

[20] Y. Cai, Y. Kang, M. Banerjee, and W. Wang. A stochastic SIRS epidemic model with infectious force under intervention strategies. *Journal of Differential Equations*, 259:7463–7502, 12 2015.

[21] K. C. De Carvalho and T. Tomé. Anisotropic probabilistic cellular automaton for a predator-prey system. *Brazilian Journal of Physics*, 37:466–471, 2007.

[22] Center for Disease Control. About toxoplasmosis. https://www.cdc.gov/toxoplasmosis/about/index.html. Accessed 04-2024.

[23] Center for Disease Control. Malaria - faqs, 2019. https://www.cdc.gov/malaria/about/faqs.html. Accessed 07-19.

[24] Center for Disease Control. Toxoplasmosis - biology, 2019. https://www.cdc.gov/parasites/toxoplasmosis/biology.html. Accessed 07-19.

[25] Center for Disease Control. Toxoplasmosis - epidemiology and risk factors, 2019. https://www.cdc.gov/parasites/toxoplasmosis/epi.html. Accessed 07-19.

[26] E. J. Chapman and C. J. Byron. The flexible application of carrying capacity in ecology. *Global Ecology and Conservation*, 13:e00365, 1 2018.

[27] R. Cherniha and V. Davydovych. Construction and application of exact solutions of the diffusive Lotka–Volterra system: A review and new results. *Communications in Nonlinear Science and Numerical Simulation*, 113:106579, 10 2022.

[28] D. J. Civitello, J. Cohen, H. Fatima, N. T. Halstead, J. Liriano, T. A. McMahon, C. N. Ortega, E. L. Sauer, T. Sehgal, S. Young, and J. R. Rohr. Biodiversity inhibits parasites: Broad evidence for the dilution effect. *Proceedings of the National Academy of Sciences of the United States of America*, 112:8667–8671, 7 2015.

[29] D. J. Civitello, J. Cohen, H. Fatima, N. T. Halstead, T. A. McMahon, C. N. Ortega, E. L. Sauer, S. Young, and J. R. Rohr. Reply to salkeld et al.: Diversity-disease patterns are robust to study design, selection criteria, and publication bias. *Proceedings of the National Academy of Sciences of the United States of America*, 112:E6262, 11 2015.

[30] C. Cohen, M. Shemesh, M. Garrido, I. Messika, M. Einav, I. Khokhlova, S. Tasker, and H. Hawlena. Haemoplasmas in wild rodents: Routes of transmission and infection dynamics. *Molecular Ecology*, 27:3714–3726, 9 2018.

[31] J. E. Cohen, S. L. Pimm, P. Yodzis, and J. Saldana. Body sizes of animal predators and animal prey in food webs. *The Journal of Animal Ecology*, 62:67, 1 1993.

[32] D. Cooke and R. W. Hiorns. *The Mathematical Theory of the Dynamics of Biological Populations*. Academic Press, 1981.

[33] T. J. Daniels, R. C. Falco, and D. Fish. Estimating population size and drag sampling efficiency for the blacklegged tick (acari: Ixodidae). *Journal of Medical Entomology*, 37:357–363, 2000.

[34] O. Diekmann, J. A.P. Heesterbeek, and T. Britton. *Mathematical tools for understanding infectious diseases dynamics*. Princeton University Press, 2013.

[35] O. Diekmann, J. A.P. Heesterbeek, and J. A.J. Metz. On the definition and the computation of the basic reproduction ratio r0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28:365–382, 6 1990.

[36] O. Diekmann, J. A.P. Heesterbeek, and M. G. Roberts. The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society Interface*, 7:873–885, 6 2010.

[37] K. Dietz. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2:23–41, 3 1993.

[38] M. A. Diuk-Wasser, A. G. Gatewood, M. R. Cortinas, S. Yaremych-Hamer, J. Tsao, U. Kitron, G. Hickling, J. S. Brownstein, E. Walker, J. Piesman, and D. Fish. Spatiotemporal patterns of host-seeking ixodes scapularis nymphs (acari: Ixodidae) in the united states. *Journal of Medical Entomology*, 43:166–176, 3 2006.

[39] M. J. Dornbusch, R. F. Limb, and C. K. Gasch. Facilitation of an exotic grass through nitrogen enrichment by an exotic legume. *Rangeland Ecology and Management*, 71:691–694, 11 2018.

[40] H. Van Den Esker, R. Van Der Hofstad, G. Hooghiemstra, and D. Znamenski. Distances in random graphs with infinite mean degrees. *Extremes*, 8:111–141, 9 2005.

[41] N. T. Fadai, S. T. Johnston, and M. J. Simpson. Unpacking the allee effect: determining individual-level mechanisms that drive global population dynamics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476, 9 2020.

[42] W. F. Fagan and J. M. Calabrese. The correlated random walk and the rise of movement ecology. *The Bulletin of the Ecological Society of America*, 95:204–206, 7 2014.

[43] P. E. M. Fine. Herd immunity: History, theory, practice. *Epidemiologic Reveiws*, 15:265–302, 1993.

[44] J. D. Fridley. The influence of species diversity on ecosystem productivity: How, where, and why? *Oikos*, 93:514–526, 6 2001.

[45] J. Gatkowska, M. Wieczorek, B. Dziadek, K. Dzitko, and H. Dlugonska. Behavioral changes in mice caused by toxoplasma gondii invasion of brain. *Parasitology Research*, 111:53–58, 7 2012.

[46] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81:2340–2361, 1977.

[47] A. D. Gordon, T. A. Henzinger, A. V. Nori, and S. K. Rajamani. Probabilistic programming. In *Future of Software Engineering*, pages 167–181. Association for Computing Machinery, 5 2014.

[48] T. N. Grainger, J. M. Levine, and B. Gilbert. The invasion criterion: A common currency for ecological research. *Trends in Ecology and Evolution*, 34:925–935, 10 2019.

[49] E. A. Gómez-Hernández, F. N. Moreno-Gómez, M. Bravo-Gaete, and F. Córdova-Lepe. Concurrent dilution and amplification effects in an intraguild predation eco-epidemiological model. *Scientific Reports*, 13:1–9, 4 2023.

[50] Fletcher W. Halliday, Jason R. Rohr, and Anna-Liisa Laine. Biodiversity loss underlies the dilution effect of biodiversity. *Ecology Letters*, 23:1611–1622, 11 2020.

[51] S. Halsey. Defuse the dilution effect debate. *Nature Ecology and Evolution*, 3:145–146, 12 2018.

[52] D. Hare, M. Daniels, and B. Blossey. Public perceptions of deer management in scotland. *Frontiers in Conservation Science*, 2:781546, 12 2021.

[53] R. F. Hechinger and K. D. Lafferty. Host diversity begets parasite diversity: bird final hosts and trematodes in snail intermediate hosts. *Proceedings of the Royal Society B: Biological Sciences*, 272:1059–1066, 5 2005.

[54] M. W. Hirsch. Systems of differential equations that are competitive or cooperative ii: Convergence almost everywhere. *SIAM Journal on Mathematical Analysis*, 16:423–439, 5 1985.

[55] M. W. Hirsch. Systems of differential equations which are competitive or cooperative: Iii. competing species. *Nonlinearity*, 1:51–71, 2 1988.

[56] M. W. Hirsch. Systems of differential equations that are competitive or cooperative. iv: Structural stability in three-dimensional systems. *SIAM Journal on Mathematical Analysis*, 21:1225–1234, 9 1990.

[57] G. Hu and K. Wang. Stability in distribution of competitive Lotka–Volterra system with markovian switching. *Applied Mathematical Modelling*, 35:3189–3200, 7 2011.

[58] R. Hu and W. A. Rowley. Relationship between weights of the engorged nymphal stage and resultant sexes in ixodes scapularis and dermacentor variabilis (acari: Ixodidae) ticks. *Journal of Medical Entomology*, 37:198–200, 2000.

[59] C. I. Huang, S. C. Kay, S. Davis, D. M. Tufts, K. Gaffett, B. Tefft, and M. A. Diuk-Wasser. High burdens of Ixodes scapularis larval ticks on white-tailed deer may limit Lyme disease risk in a low biodiversity setting. *Ticks and Tick-borne Diseases*, 10:258–268, 2 2019.

[60] A. Hurwitz. *Selected Papers on Mathematical Trends in Control Theory*. Dover, 1964.

[61] The MathWorks Inc. Matlab, 2019.

[62] W. M. Ingram, L. M. Goodrich, E. A. Robey, and M. B. Eisen. Mice infected with low-virulence strains of toxoplasma gondii lose their innate aversion to cat urine, even after extensive parasite clearance. *PLoS One*, 8:e75246, 9 2013.

[63] C. N. Johnson. Relationships between body size and population density of animals: The problem of the scaling of study area in relation to body size. *Oikos*, 85:565, 6 1999.

[64] P. T. J. Johnson, D. L. Preston, J. T. Hoverman, and K. L. D. Richgels. Biodiversity decreases disease through predictable changes in host community competence. *Nature*, 494:230–233, 2 2013.

[65] R. A. Johnston, L. W. Cowgill, and R. L. Lyman. Estimation of body mass in white-tailed deer (odocoileus virginianus) using cross-sectional geometry of the metapodial. *Journal of Archaeological Science: Reports*, 37:102889, 6 2021.

[66] D. S. Jones, M. Plank, and B. Sleeman. *Differential equations and mathematical biology*. CRC Press, 2 1983.

[67] R. A. Jordan, T. L. Schulze, and M. B. Jahn. Effects of reduced deer density on the abundance of Ixodes scapularis (Acari: Ixodidae) and Lyme disease incidence in a northern New Jersey endemic area. *Journal of Medical Entomology*, 44:752–757, 9 2007.

[68] P. M. Kareiva and N. Shigesada. Analyzing insect movement as a correlated random walk. *Oecologia*, 56:234–238, 2 1983.

[69] H. Kedem, C. Cohen, I. Messika, M. Einav, S. Pilosof, and H. Hawlena. Multiple effects of host-species diversity on coexisting host-specific and host-opportunistic microbes. *Ecology*, 95:1173–1183, 5 2014.

[70] F. Keesing, R. D. Holt, and R. S. Ostfeld. Effects of species diversity on disease risk. *Ecology Letters*, 9:485–498, 3 2006.

[71] F. Keesing and R. S. Ostfeld. Dilution effects in disease ecology. *Ecology Letters*, 24:2490–2505, 11 2021.

[72] E. Kelting, B. Bannish, and S. Laverty. Toxoplasma gondii: A mathematical model of its transfer between cats and the environment. *SIAM Undergraduate Research Online*, 11, 2018.

[73] E. Kenah and J. M. Robins. Second look at the spread of epidemics on networks. *Physical Review E*, 76, 9 2007.

[74] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 115:700–721, 8 1927.

[75] Hussein Khalil, Frauke Ecke, Magnus Evander, Magnus Magnusson, and Birger Hörnfeldt. Declining ecosystem health and the dilution effect. *Scientific Reports*, 6:31314, 8 2016.

[76] A. M. Kilpatrick, A. D.M. Dobson, T. Levi, D. J. Salkeld, A. Swei, H. S. Ginsberg, A. Kjemtrup, K. A. Padgett, P. M. Jensen, D. Fish, N. H. Ogden, and M. A. Diuk-Wasser. Lyme disease ecology in a changing world: consensus, uncertainty and critical gaps for improving control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372, 6 2017.

[77] H. J. Kilpatrick, A. M. Labonte, and K. C. Stafford. The relationship between deer density, tick abundance, and human cases of lyme disease in a residential community. *Journal of Medical Entomology*, 51:777–784, 7 2014.

[78] G. M. Knight, N. J. Dharan, G. J. Fox, N. Stennis, A. Zwerling, R. Khurana, and D. W. Dowdy. Bridging the gap between evidence and policy for infectious diseases: How models can aid public health decision-making. *International Journal of Infectious Diseases*, 42:17–23, 2016.

[79] M. Kohls and T. Hernandez. Expected coverage of random walk mobility algorithm. *Unpublished manuscript*, 11 2016.

[80] K. J. Kugeler, R. A. Jordan, T. L. Schulze, K. S. Griffith, and P. S. Mead. Will culling white-tailed deer prevent Lyme disease? *Zoonoses and public health*, 63:337–345, 8 2016.

[81] K. D. Lafferty and C. L. Wood. It's a myth that protection against disease is a strong and general service of biodiversity conservation: Response to Ostfeld and Keesing. *Trends in Ecology and Evolution*, 28:503–504, 9 2013.

[82] T. Levi, A. M. Kilpatrick, M. Mangel, and C. C. Wilmers. Deer, predators, and the emergence of Lyme disease. *Proceedings of the National Academy of Sciences of the United States of America*, 109:10942–10947, 7 2012.

[83] W.D. Li and X.H. Guo. Using cellular automata to study the effect of competition for epidemic diseases. *Procedia Environmental Sciences*, 13:1010–1018, 1 2012.

[84] L. R. Lindsay, I. K. Barker, G. A. Surgeoner, S. A. McEwen, T. J. Gillespie, and J. T. Robinson. Survival and development of Ixodes scapularis (Acari: Ixodidae) under various climatic conditions in Ontario, Canada. *Journal of Medical Entomology*, 32:143–152, 1995.

[85] K. LoGiudice, R. S. Ostfeld, K. A. Schmidt, and F. Keesing. The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. *Proceedings of the National Academy of Sciences of the United States of America*, 100:567–571, 1 2003.

[86] A. J. Lotka. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry*, 14:271–274, 3 1910.

[87] A. J. Lotka. Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, 6:410–415, 7 1920.

[88] N. K. Madhav, J. S. Brownstein, J. I. Tsao, and D. Fish. A dispersal model for the range expansion of blacklegged tick (Acari: Ixodidae). *Journal of Medical Entomology*, 41:842–852, 2004.

[89] Y. Maki and H. Hirose. Infectious disease spread analysis using stochastic differential equations for SIR model. In *International Conference on Intelligent Systems, Modelling and Simulation*, pages 152–156, 2013.

[90] M. T. Marrelli, R. S Malafronte, M. A. Sallum, and D. Natal. Kerteszia subgenus of Anopheles associated with the Brazilian Atlantic rainforest: current knowledge and future challenges. *Malaria Journal*, 2007.

[91] R. M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467, 1976.

[92] N. Metropolis and S. Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44:335, 9 1949.

[93] D. Mollison. Dependence of epidemic and population velocities on basic parameters. *Mathematical Biosciences*, 107:255–287, 12 1991.

[94] K. D. Moody, S. W. Barthold, G. A. Terwilliger, D. S. Beck, G. M. Hansen, and R. O. Jacoby. Experimental chronic Lyme borreliosis in Lewis rats. *The American Journal of Tropical Medicine and Hygiene*, 42:165–174, 2 1990.

[95] C. J. L. Murray, R. Lozano, M. Naghavi, A. D. Flaxman, C. Michaud, M. Ezzati, K. Shibuya, J. A. Salomon, and A. D. Lopez et al. Disability-adjusted life years (dalys) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the global burden of disease study 2010. *The Lancet*, 380:2197–2223, 12 2012.

[96] M. E.J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66, 7 2002.

[97] A. Nguyen, J. Mahaffy, and N. K. Vaidya. Modeling transmission dynamics of Lyme disease: Multiple vectors, seasonality, and vector mobility. *Infectious Disease Modelling*, 4:28, 1 2019.

[98] L. Nguyen. Simulation 102: Agent-based disease modeling. *Medium*, 2023.

[99] J. V. Noble. Geographic and temporal development of plagues. *Nature 1974 250:5469*, 250:726–729, 1974.

[100] N. H. Ogden, M. Bigras-Poulin, K. Hanincová, A. Maarouf, C. J. O'Callaghan, and K. Kurtenbach. Projected effects of climate change on tick phenology and fitness of pathogens transmitted by the North American tick Ixodes scapularis. *Journal of theoretical biology*, 254:621–632, 10 2008.

[101] N. H. Ogden and J. I. Tsao. Biodiversity and Lyme disease: Dilution or amplification? *Epidemics*, 1:196–206, 9 2009.

[102] R. S. Ostfeld. Biodiversity loss and the rise of zoonotic pathogens. *Clinical Microbiology and Infection*, 15 Suppl 1:40–43, 2009.

[103] R. S. Ostfeld and F. Keesing. Biodiversity and disease risk: The case of Lyme disease. *Conservation Biology*, 14:722–728, 6 2000.

[104] R. S. Ostfeld and F. Keesing. Straw men don't get Lyme disease: Response to Wood and Lafferty. *Trends in Ecology and Evolution*, 28:502–503, 9 2013.

[105] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86:3200–3203, 10 2000.

[106] S. E. Randolph and A. D. M. Dobson. Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology*, 139:847–863, 6 2012.

[107] J. M. Raollaigh. Deer culling to go ahead due to growing population. *RTE*, 12 2023.

[108] V. Ratti, J. M. Winter, and D. I. Wallace. Dilution and amplification effects in Lyme disease: Modeling the effects of reservoir-incompetent hosts on Borrelia burgdorferi sensu stricto transmission. *Ticks and Tick-borne Diseases*, 12:101724, 7 2021.

[109] E. Renshaw and R. Henderson. The correlated random walk. *Journal of Applied Probability*, 18:403–414, 6 1981.

[110] P. Revesz. *Random walk in random and non-random environments*. World Scientific, 1990.

[111] D. Richter, D. B. Schlee, and F. R. Matuschka. Reservoir competence of various rodents for the Lyme disease spirochete Borrelia spielmanii. *Applied and Environmental Microbiology*, 77:3565–3570, 6 2011.

[112] M. G. Roberts. The pluses and minuses of $\mathcal{R}_0$. *Journal of the Royal Society Interface*, 4:949, 10 2007.

[113] M. G. Roberts and J. A.P. Heesterbeek. A new method for estimating the effort required to control an infectious disease. *Proceedings of the Royal Society B: Biological Sciences*, 270:1359, 7 2003.

[114] M. G. Roberts and J. A.P. Heesterbeek. Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. *Journal of Mathematical Biology*, 55:803–816, 11 2007.

[115] M. G. Roberts and J. A.P. Heesterbeek. Characterizing the next-generation matrix and basic reproduction number in ecological epidemiology. *Journal of Mathematical Biology*, 66:1045–1064, 2013.

[116] M. G. Roberts and J. A.P. Heesterbeek. Quantifying the dilution effect for models in ecological epidemiology. *Journal of the Royal Society Interface*, 15, 2018.

[117] J. R. Rohr, D. J. Civitello, F. W. Halliday, P. J. Hudson, K. D. Lafferty, C. L. Wood, and E. A. Mordecai. Towards common ground in the biodiversity-disease debate. *Nature Ecology and Evolution*, 4:24–33, 1 2020.

[118] E. J. Routh. *A treatise on the stability of a given state of motion, particularly steady motion*. Macmillan and Company, 1877.

[119] D. J. Salkeld, K. A. Padgett, and J. H. Jones. A meta-analysis suggesting that the relationship between biodiversity and risk of zoonotic pathogen transmission is idiosyncratic. *Ecology Letters*, 16:679–686, 5 2013.

[120] D. J. Salkeld, K. A. Padgett, J. H. Jones, and M. F. Antolin. Public health perspective on patterns of biodiversity and zoonotic disease. *Proceedings of the National Academy of Sciences of the United States of America*, 112:E6261, 11 2015.

[121] G. B.M. Santos, T. F.A. Alves, G. A. Alves, A. Macedo-Filho, and R. S. Ferreira. Epidemic outbreaks on two-dimensional quasiperiodic lattices. *Physics Letters A*, 384:126063, 1 2020.

[122] J. E. Satulovsky and T. Tomé. Stochastic lattice gas model for a predator-prey system. *Physical Review E*, 49:5073–5079, 1994.

[123] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[124] S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52:591–611, 12 1965.

[125] E. H. Simpson. Measurement of diversity. *Nature 1949 163:4148*, 163:688–688, 1949.

[126] J. G. Skellam. Random dispersal in theoretical populations. *Biometrika*, 38:196, 6 1951.

[127] J. G. Skellam. The formulation and interpretation of mathematical models of diffusional process in population biology. *The Mathematical Theory of the Dynamic of Biological Populations*, 1973.

[128] S. Smale. On the differential equations of species in competition. *Journal of Mathematical Biology*, 3:5–7, 1976.

[129] A. Srivastava. Random-walk, agent-level pandemic simulation (raw-alps) for analyzing effects of different lockdown measures. *Frontiers in Applied Mathematics and Statistics*, 7:638996, 4 2021.

[130] A. C. Steere, E. Taylor, M. L. Wilson, J. F. Levine, and A. Spielman. Longitudinal assessment of the clinical and epidemiological features of lyme disease in a defined population. *The Journal of Infectious Diseases*, 154:295–300, 1986.

[131] A. T. Strauss, S. E. Hobbie, P. B. Reich, E. W. Seabloom, and E. T. Borer. The effect of diversity on disease reverses from dilution to amplification in a 22-year biodiversity×N×CO_2 experiment. *Scientific Reports*, 14:1–13, 5 2024.

[132] L. H. Taylor, S. M. Latham, and M. E.J. Woolhouse. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356:983–989, 7 2001.

[133] S. R. Telford. Deer reduction is a cornerstone of integrated deer tick management. *Journal of Integrated Pest Management*, 8:25, 1 2017.

[134] T. Tomé and R. M. Ziff. Critical behavior of the susceptible-infected-recovered model on a square lattice. *Physical Review E*, 82, 11 2010.

[135] F. Vadillo. Comparing stochastic Lotka–Volterra predator-prey models. *Applied Mathematics and Computation*, 360:181–189, 11 2019.

[136] D. Valle and J. Clark. Conservation efforts may increase malaria burden in the brazilian amazon. *PLoS One*, 8:e57519, 3 2013.

[137] J. A. Vano, J. C. Wildenberg, M. B. Anderson, J. K. Noel, and J. C. Sprott. Chaos in low-dimensional Lotka-Volterra models of competition. *Nonlinearity*, 19:2391–2404, 10 2006.

[138] C. Varga, P. McDonald, W. M. Brown, P. Shelton, A. L. Roca, J. E. Novakofski, and N. E. Mateus-Pinilla. Evaluating the ability of a locally focused culling program in removing chronic wasting disease infected free-ranging white-tailed deer in illinois, usa, 2003–2020. *Transboundary and Emerging Diseases*, 69:2867–2878, 9 2022.

[139] P. Verhulst. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance Mathématique et Physique*, 10:113–121, 1838.

[140] P. Verhulst. Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles*, 18:8, 1845.

[141] P. Verhulst. Deuxième mémoire sur la loi d'accroissement de la population. *Mémoires de l'Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique*, 20:1–32, 1847.

[142] L. J. Verme. Reproductive patterns of white-tailed deer related to nutritional plane. *The Journal of Wildlife Management*, 1969.

[143] V. Volterra. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Memoria della Reale Accademia Nazionale dei Lincei*, 2:31–113, 1926.

[144] D. Wallace, V. Ratti, A. Kodali, J. M. Winter, M. P. Ayres, J. W. Chipman, C. F. Aoki, E. C. Osterberg, C. Silvanic, T. F. Partridge, and M. J. Webb. Effect of rising temperature on Lyme disease: Ixodes scapularis population dynamics and Borrelia burgdorferi transmission and prevalence. *Canadian Journal of Infectious Diseases and Medical Microbiology*, 2019, 2019.

[145] E. C. Y. Wang, T. H. Grasela, and C. A. Walawander. Applying epidemiology-based outcomes research to clinical decision- making: A hypothetical model of biotechnology therapy in gram-negative sepsis. *PharmacoEconomics*, 15:385–393, 4 1999.

[146] J. P. Webster. The effect of toxoplasma gondii on animal behavior: Playing cat and mouse. *Schizophrenia Bulletin*, 33:752–756, 5 2007.

[147] R. H. Whittaker. Evolution and measurement of species diversity. *Taxon*, 21:213–251, 5 1972.

[148] S. Wise, M. Garrido, S. Halle, R. Flatau, C. Cohen, Á. Navarro-Castilla, I. Barja, and H. Hawlena. The dilution effect behind the scenes: testing the underlying assumptions of its mechanisms through quantifying the long-term dynamics and effects of a pathogen in multiple host species. *Proceedings of the Royal Society B: Biological Sciences*, 2021.

[149] C. L. Wood and K. D. Lafferty. Biodiversity and disease: A synthesis of ecological perspectives on Lyme disease transmission. *Trends in Ecology and Evolution*, 28:239–247, 4 2013.

[150] C. L. Wood, K. D. Lafferty, G. DeLeo, H. S. Young, P. J. Hudson, and A. M. Kuris. Does biodiversity protect humans against infectious disease? *Ecology*, 95:817–832, 2014.

[151] C. L. Wood, A. Mcinturff, H. S. Young, D. Kim, and K. D. Lafferty. Human infectious disease burdens decrease with urbanization but not with biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372, 2017.

[152] World Health Organisation. Who — influenza (seasonal), 2023. http://www.who.int/mediacentre/factsheets/fs211/en/. Accessed 04-2024.

[153] World Health Organisation. Ebola virus disease, 2024. https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease, Accessed 04-2024.

[154] Y. Y. Zhu, H. R. Chen, J. H. X. Fan, Y. Y. Wang, Y. Li, J. B. Chen, J. H. X. Fan, S. S. Yang, L. P. Hu, H. Leung, T. W. Mew, P. S. Teng, Z. H. Wang, and C. C. Mundt. Genetic diversity and disease control in rice. *Nature*, 406:718–722, 2000.

10 Appendices

10.1 Lattice-Model Functionality

Here we detail functionality and mechanisms developed for the lattice-based model that were not relevant to the results obtained in Chapters 3 and 4. Given the aim of developing tools that can explore scale as it relates to the dilution effect, we outline those functionalities here for potential future use, and to catalogue the full range of dynamics of the models developed.

10.1.1 Movement on a Lattice

Retaining the SIR-based infection dynamics underpinning the plant-based model, we first consider a system containing only one species. In our system, each site can only be occupied by a single animal, and so it is necessary to relax the assumption of total site occupation. Consequently, not every animal would have a nearest neighbour for disease spread, so we must rethink disease transmission when implementing movement between sites. We reintroduce the matrix A describing the occupation status of each site i, j,

$$A_{i,j} = \begin{cases} 0 & \text{Position (i, j) is unoccupied} \\ 1 & \text{Position (i, j) is occupied} \end{cases} \tag{168}$$

We assume that each animal moves one site in a random direction (either up, down, left or right) provided the site in the chosen direction exists and is unoccupied. Representing this mathematically as a change to the matrix A is highly unintuitive, so for clarify we prefer to talk about how the position of each animal changes. After producing a random order of occupied sites, the matrix A is altered sequentially to account for the movement for each site. We denote the position of the animal k in an occupied site to be given by $a_k(t) = (x, y)$. Then, the position is updated according to

$$a_k(t+1) = \begin{cases} (x, y-1) & \text{if } y \neq 0 \wedge \text{(x, y-1) is empty } \wedge p < 0.25 \\ (x+1, y) & \text{if } x \neq N \wedge \text{(x+1, y) is empty } \wedge 0.25 < p < 0.5 \\ (x, y+1) & \text{if } y \neq N \wedge \text{(x, y+1) is empty } \wedge 0.5 < p < 0.75 \\ (x-1, y) & \text{if } x \neq 0 \wedge \text{(x-1, y) is empty } \wedge 0.75 < p < 1 \\ (x, y) & \text{otherwise} \end{cases}, \tag{169}$$

where $0 < p < 1$ is a uniformly distributed random number. After each location is updated, so too are the corresponding entries of the matrix A, until each occupied site is updated, by which point $A(t)$ has become $A(t+1)$.

We also amend the disease transmission mechanism to be more appropriate in this context. We retain the matrix $S_{i,j}$ describing the infection state of the animal in position (i, j), and the matrix $Q_{i,j}$ describing the number of timesteps for which the animal in (i, j) has remained in either the infected or recovered state. These entries change as individuals move through the various infection states, and shift positions along with the corresponding animal, as outlined in Equation 169. However, instead of using a matrix $P_{i,j}$ to describe the probability of infection for a susceptible animal in position i, j, it is more practical to tie the infection mechanism to "interactions" that occur between animals in neighbouring sites. In essence, if a susceptible animal attempts to move into an occupied site, the two animals interact; interactions between a susceptible and infected animal result in the former becoming infected with probability β. This mode of transmission is density-dependent; i.e. more interactions occur with more animals spread over the same area. Mathematically, if a susceptible animal in position (x, y) interacts with an infected neighbour in position (x', y'), then

$$S_{x,y} = \begin{cases} 1 & \text{if } q < \beta \\ 0 & \text{if } q > \beta \end{cases}, \tag{170}$$

where $0 < q < 1$ is another uniformly distributed random number. In addition to direct transmission, perinatal transmission is also able to be captured by this model, by way of the reproduction process we also implement.

Since the primary focus of the model is epidemiological, we seek to include demography in our model on the scale of the whole system, rather than as a result of interactions between individuals. Define the prevalence $X(t)$ to be

$$X(t) = \frac{1}{N^2} \sum_{i,j}^{N} A_{i,j}(t). \tag{171}$$

In words, this is the proportion of sites that are occupied. After all previous processes in a time step are complete (specifically, movement and disease transmission), we then evaluate the shift in the population, ΔX which occurs as a result of births and deaths.

$$\Delta X(t) = f(X(t)) - g(X(t)), \tag{172}$$

where $f(x)$ and $g(x)$ are functions selected to represent births and deaths, respectively. In cases with no disease, there is no reason to consider anything more than a net population change, including disease dynamics necessitates that births and deaths be considered directly in almost all cases. Multiplying these functions by N (and rounding) gives the corresponding number of births and deaths between time steps. Births are implemented by randomly changing the appropriate number of zero entries of A; similarly deaths are implemented by randomly changing the appropriate number of non-zero entries. The functions $f(x)$ and $g(x)$ must satisfy a number of conditions. Firstly, barring the introduction of new individuals from an outside source, $f(0) = g(0) = 0$, as a population of zero is unable to repopulate or indeed die down further. Secondly, $f(x)$ and $g(x)$ must satisfy $0 \leq X(t) + \Delta X(t) \leq 1$, so that the births and deaths never cause the population to be negative or exceed the capacity of the matrix. A special case of this is that, $f(1) - g(1) \leq 0$, as a totally occupied matrix leaves no room for further births. Any zeroes of $\Delta X(t)$ would correspond to an ecological steady state, the stability of which can be determined as with any discrete map. The simplest choice of functions for $f(x)$ and $g(x)$ mimics the simplest ecological model for exponential growth

$$f(x) = \mu x \tag{173}$$

$$g(x) = \alpha x. \tag{174}$$

The possibility of perinatal (or vertical) transmission between the last generation and any new births can also be accounted for using $\Delta X(t)$. This is the case for a number of infectious diseases, including toxoplasmosis and HIV in humans, and in Rift Valley fever in mosquitos, among other animalian examples. Define the proportion of new births that are infected to be I_{new}. Since this model does not allow for sexes in the population, we assume an equal proportion of each sex. In the simple case where there is a 100% rate of transmission from the mother to the child, then $I_{new} = I(t)f(X(t))N$, where $I(t)$ is the proportion of animals which are infected, though this can easily be adapted for other possibilities, or neglected in the absence of vertical transmission. The infected status is then given at random to an appropriate number of the new births. We can incorporate a variance in births or deaths based on species-wide infection rates to incorporate the effects of disease-related mortality as well, by making appropriate changes to the functions $f(x)$ and $g(x)$.

10.1.2 Explicit Resource Competition

We can also include the effects of resource competition; akin to a logistic model. In the context of animals such a resource could be food, space or water. In a continuous-time context, a basic logistic model with infectious disease dynamics is expressed by the following system of equations:

$$\frac{dS}{dt} = -\frac{\beta IS}{N} + \mu N - \alpha S - \nu NS \tag{175}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \alpha I - \nu NI \tag{176}$$

$$\frac{dR}{dt} = \gamma I - \alpha R - \nu NR. \tag{177}$$

The birth and death rates have been multiplied by $N = S + I + R$, which is representative of the exponential growth in the absence of competition that is typical of the logistic differential equation. The death rate has been supplemented by an additional term proportional to N as well, representing competition for some resource; the higher the population is, the greater this term becomes, constraining the growth of the population beyond a certain carrying capacity, denoted by k. By adding each of the three equations together and recalling the definition of N, we can recover the logistic differential equation, described by

$$\frac{dN}{dt} = (\mu - \alpha)N\left(1 - \frac{\nu N}{\mu - \alpha}\right), \tag{178}$$

where we have the growth rate $r = \mu - \alpha$, and the carrying capacity $K = (\mu - \alpha)/\nu$. The infectious disease mechanics remain the same as in previous models, though we once again assume that all newborns are born in the susceptible state, where this is not necessarily the case.

In the context of our agent-based model, we can implement something analogous to logistic growth using a new matrix, which we will denote as $F(t)$ (representing food, though other resources can be represented this way). The entries of $F(t)$, $F_{i,j}$, correspond to the amount of the resource in question available at the given site i, j. At the conclusion of each time step, every animal in an occupied site checks to see if the resource is available. If the resource is available, $F_{i,j}(t+1) = F_{i,j}(t) - 1$. Mathematically, we express the updating of this matrix as follows

$$F_{i,j}(t+1) = \begin{cases} F_{i,j}(t) - 1 & \text{if } A_{i.j} = 1 \wedge Fi, j \geq 1 \\ F_{i,j}(t) & \text{otherwise} \end{cases} \tag{179}$$

Otherwise, the animal in question dies, providing a means of death that features spatial dyanmics, rather than one that is consigned to a function that applies randomly over the whole array. Said resource can regenerate at some rate $\psi(t)$. This rate can be zero (modelling water in a floodplain), constant (grass growing at a constant rate), or seasonal (leaves on a deciduous tree species), depending on the scenario being described by the model. With this approach, setting an initial condition for the resource distribution is required; keeping this constant can result in seasonal forcing, where all cells produce a new unit of food at the same time, potentially causing periodic rises and falls in population. Adding a uniformly distributed random number between 0 and 1 to the initial value for each square can avoid this dynamic where it is not desired. Further to this, seeding the initial values in other ways can also introduce more varied spatial dynamics, such as resource replenishment as a travelling wave.

While this explicit mechanism is useful, logistic growth can also be included on a population-level by simply amending the functions dictating births and deaths $f(x)$ and $g(x)$. This has the principal benefit of minimising computation time, at the cost of spatial dynamics. Where the death function has a logistic term, it is amended to be given by

$$g(x) = \alpha x - \nu x^2. \tag{180}$$

In having both spatially explicit and population-level incorporations of logistic effects, the agent-based model is more adaptable to the requirements of any particular scenario. These demographic and movement mechanics also create new opportunities for disease transmission other than direct contact. An example of this is the possibility for environmental contamination and transmission, which will be touched on in the following section.

10.1.3 Environmental Transmission

A further possible inclusion is environmental transmission. Many infections are not transmitted directly from host to host, and are instead spread via the environment. One way this can happen is through a vector. Malaria, for example, is caused by *Plasmodium* protozoa spread by mosquitos, while Lyme disease is caused by *Borrelia* bacteria and spread by ticks [23]. Other infections are spread via soil, water or other such environmental features. Waterborne infections include the protozoal giardiasis and the bacterial cholera and dysentry, while the *Toxoplasma gondii* protozoan can be soilborne by way of the faeces of an infected cat.

Incorporating these into our single-species system of differential equations can be

done in various ways depending on the role of the environment in the infection cycle. We present a simple example in the following system of differential equations, adapted from Kelting: [72]

$$\frac{dS}{dt} = -\omega ES \tag{181}$$

$$\frac{dI}{dt} = \omega ES - \gamma I \tag{182}$$

$$\frac{dR}{dt} = \gamma I \tag{183}$$

$$\frac{dE}{dt} = \lambda I - \eta E, \tag{184}$$

where $E(t)$ is the amount of the pathogen which persists in the environment at time t. The pathogen is spread to the environment at a rate of λ per infected individual, while η is the rate at which pathogen in the environment is lost. Taking the role of β in the standard SIR model, ωES is the rate at which the pathogen spreads from the environment to susceptible individuals. The basic reproduction number and final size equation can be found to be

$$R_0 = \sqrt{\frac{N\lambda\omega}{\gamma\eta}} \tag{185}$$

$$R = N(1 - e^{-R_0^2 R/N}). \tag{186}$$

Even this simple formulation is applicable to multiple potential contexts. We could consider a system where E represented a vector that persists at a constant population with turnover rate η, transferring the pathogen from infected individuals to susceptible individuals without transmission between the vector. We could equivalently consider a system in which a pathogen is shed by infected individuals into the environment, where it cannot survive indefinitely and eventually decays, while susceptible individuals come into contact with it and become infected. It would be assumed in this context that susceptibles consuming the pathogen would not significantly reduce the amount of it in the environment; this would be because many millions of the pathogen would be shed into the environment by each infected individual, while it typically only takes a few of the pathogen to enter a host for infection to take hold [72]. We also assume for both of these contexts that recovered individuals ease to be infectious, though it is important to note here in particular that this is not always the case.

We can also add the effects of demography to this model, producing the following:

$$\frac{dS}{dt} = -\omega ES + \mu - \alpha S \tag{187}$$

$$\frac{dI}{dt} = \omega ES - \gamma I - \alpha I \tag{188}$$

$$\frac{dR}{dt} = \gamma I - \alpha R \tag{189}$$

$$\frac{dE}{dt} = \lambda I - \eta E. \tag{190}$$

Returning to our agent-based model, we can implement environmental transmission analogously in various ways depending on the mechanism we seek to describe. We can add a new matrix to the system $E(t)$, with entries $E_{i,j}$ that denote the amount of a pathogen present in the environment at the site i,j. Should a susceptible individual move into a site for which $E_{ij} > 0$, it becomes infected with some probability $0 \leq \omega(E) \leq 1$, which is a function of the amount of the pathogen at the site, and should satisfy $\omega'(E) \geq 0 \; \forall E$, so that an increase in pathogen prevalence always results in an increase in risk of infection. The matrix updates by considering an exponential decay in the amount of the pathogen at a site from time step to time step, while assuming that an infected individual always sheds a constant amount of the pathogen into its immediate vicinity at the site it currently occupies. We represent this mathematically as

$$E_{i,j}(t+1) = \frac{E_{i,j}}{\eta} + \lambda\delta(S_{i,j} - 1), \tag{191}$$

,

where λ is the constant rate of shedding of an infected individual, and $\eta > 1$ is a constant rate of decay per time step. Hence the amount of pathogen remaining that was shed by an infected individual at a time step $t = 0$ can be evaluated as $\lambda\eta^{-t}$, where t is the number of time steps which have passed since $t = 0$.

Considering vectors that would move and interact independently is substantially more complicated. Even the simplest vector would need to be implemented as a separate species, which would not work in the single species framework discussed in this section. Therefore, we leave discussion of potential vectors for the section that follows, which extends the agent-based model to multiple species.

10.1.4 Interspecies Dynamics

We now consider the extension of this model to the case where n species exist in the array. As with the plant-based precursor to this model, we must account for situations

where parameters such as the transmission rate and birth rate are different between species. The transmission rate β again becomes an $n \times n$ matrix with elements $\beta_{i,j}$, which represent the probability of a susceptible individual of species i becoming infected after interacting with an individual of species j. The rates of recovery and loss of immunity (where relevant) for the species are represented by vectors with n elements, γ_i and ξ_i, respectively. The function(s) dictating population turnover are similarly indexed to describe birth and death rates for a specific species i.

With the coexistence of multiple host species, it is important to consider the possible interactions in the ecosystem we are simulating. It is possible to include various forms of symbiotic relationships, including parasitism, mutualism and commensalism, by adjusting the ecological and/or epidemiological parameters of the model to depend on the prevalence of another species, but we will not consider these in detail. We instead turn our attention towards the incorporation of predator-prey dynamics, as well as inter-species resource competition, which we will detail here.

There are various possible ways species can interact, one predator species can consume another as prey, two or more species could compete for resources like space or a food source, or some form of symbiotic relationship could exist. Before considering how this is implemented in the case of our agent-based model, we consider the continuous-time equivalent. For n species, a generalised system of $3n$ equations which can capture a wide range of dynamics is described by the following system of differential equations:

$$\frac{dS_k}{dt} = \mu_k N_k - \alpha_k S_k - S_k \sum_{i=1}^{n} \frac{\beta_{k,i} I_i}{N_i} - S_k \sum_{i=1}^{n} \phi_{k,i} N_i \tag{192}$$

$$\frac{dI_k}{dt} = -\alpha_k I_k + S_k \sum_{i=1}^{n} \frac{\beta_{k,i} I_i}{N_i} - \gamma_k I_k - I_k \sum_{i=1}^{n} \phi_{k,i} N_i \tag{193}$$

$$\frac{dR_k}{dt} = -\alpha_k R_k + \gamma_k I_k - R_k \sum_{i=1}^{n} \phi_{k,i} N_i, \tag{194}$$

where $N_k = S_k + I_k + R_k$ is the number of each species present. Like in the single species case, we assume that for each species, all newborns are put into the susceptible state, and deaths occur equally from each state, as a result either of natural deaths or of interactions between species. The parameter $\phi_{i,j}$ describes the interactions between each species. A positive value corresponds to a decrease in growth as a result of these interactions, while conversely, a negative value corresponds to an increase in growth as a result of interactions. In addition, the ϕ_{ii} terms in these equations correspond to the intraspecies resource competition resulting in logistic growth discussed in the previous section. The possible dynamics described by ϕ_{ii} are appear in Table 2.

In the context of our agent-based model, we describe what interactions occur between species using an analogous matrix ϕ, whose entries $\phi_{i,j}$ describe the interactions with other species and the environment. The diagonal entries describe whether the species consumes resources in the environment, whereas the other entries indicate whether a species consumes others. Logically, its meaning is represented as

$$\phi_{i,j} = \begin{cases} 1 \wedge i \neq j & \text{species } i \text{ predates species } j \\ 1 \wedge i = j & \text{species } i \text{ consumes an environmental resource} \\ 0 & \text{species does not consume the relevant resource/species} \end{cases} \tag{195}$$

The mechanism for resource competition between species already exists in the single-species model, but we need to introduce a new mechanism to describe predation. To do this, we reuse the mechanism for "interactions" that we used for transmission in a single species context, where an interaction consists one individual attempts to move into a site occupied by another. For each combination of predator i and prey j, a probability of success $\rho_{i,j}$ is preselected. Should either of the two species predate the other, we generate a uniformly distributed random number $0 < p < 1$, and if $p < \rho_{i,j}$, the predator's attack is successful, and the prey is removed from the array, much like a chess piece would be removed from a chessboard.

Predation-prey interactions have implications for disease transmission. Should an pathogen be able to infect a susceptible predator, consuming infected prey is likely to result in infection. In this case, we can implement this either by assuming guaranteed transmission, transmission with probability $\beta_{i,j}$, or any other probability deemed appropriate. In the case the predator fails to consume the prey, the interaction proceeds as normal, meaning that transmission occurs with the corresponding probability β if one individual is infected and the other susceptible.

10.2 Spatially Explicit Model Functionality

As with the lattice-based model, not all of the mechanisms developed for the spatially explicit model were used in the derivation of the results in those sections. Other functionalities of the spatially explicit model are recorded here.

10.2.1 Logistic Growth

For a discrete analogue of our logistic differential equations, rather than filling in the environment with finite levels of a resource to be depleted (which is impractical in the

spatially-explicit case) we can employ the discrete logistic map. This map, popularised by May [91], relates the population of individuals from one time-step to the next by the following equation:

$$x_{t+1} = rx_t(1 - x_t), \tag{196}$$

where $x_t \in [0, 1]$ is the ratio of the population to the environment's carrying capacity, or maximum possible population, and t is the current time-step.[91] The parameter $r \in [0, 4]$ encodes two dynamics at once. For small population sizes relative to the carrying capacity, r is close to the growth rate of the species as a result of reproduction as well as by natural deaths. For larger population sizes, r also informs the rate at which additional population is lost through logistic competition for resources. The range of values for r is the range for which x_t will remain between 0 and 1 given an approximate initial condition.

This map is well-studied, and has rich dynamics. For $0 < r < 1$, the population dies out, while for $1 < r < 3$, the population will approach a fixed point at $\frac{r-1}{r}$. For $3 < r < 4$, the population will display different behaviour, including cascading cycles and eventually chaos as r increases. For the purposes of our single-species model, we are most interested in the cases where $1 < r < 2$. In this specific range of values, the population will approach the fixed point relatively quickly, without displaying the damped oscillations towards the fixed point that typify behaviour from $2 < r < 3$.

Implementing this map in the context of our model, since $0 \leq x \leq 1$, we are required to provide the carrying capacity K as a parameter. Since the map only describes a growth rate for low populations, we must also decide on a death rate in the absence of competition. While this isn't relevant to the dynamics of the map, it is important in our case because births and deaths have epidemiological ramifications that cannot be neglected in constructing our model. Births and deaths represent a way for both immune individuals to be removed from the system, and for the system to be repopulated with susceptible individuals. We retain the use of μ and α to represent the birth and death rates respectively, and, by definition $r = \mu - \alpha$. Using the definition of the carrying capacity, we have that $N = Kx_t$, and the number of births and deaths in each time-step are therefore given by

$$\Delta N_{births} = \mu K x_n \tag{197}$$

$$\Delta N_{deaths} = K(\alpha x_n + r x_n^2). \tag{198}$$

It is important to note that the map treats population as continuous, while in our model, it is discrete, and so these population changes are rounded down to the nearest individual.

Both the population at the fixed point and the rate of convergence increase with r in our desired region. Since we would expect infection to occur on a faster timescale than births and deaths in most cases, we can choose to only update our population at a certain frequency. We denote this time interval τ. Updating our demography only after τ time-steps allows us to control the relative timescales of infection and demography without affecting the rate at which the population changes or the population density at the fixed point. Doing this would have the undesirable effect of causing sudden population jumps and declines. These sharp changes are circumvented by randomly distributing ΔN_{births} and ΔN_{deaths} across the time-steps $[1, 2, ..., \tau]$ after each update of the map. A degree of stochastic variation occurs in the population is observed in the model as a result, but the smoother changes in population allow for a more realistic rate of change of the contact rate between individual agents, allowing for more realistic infection modelling.

The process for births assumes that all newborns are born into the susceptible state, though the possibility of perinatal transmission or neonatal immunity can be readily accounted for as well. Newborns are individually assigned the position of a randomly selected "parent" as their starting location, though from the next step onward they will move independently of their parent, with a random starting direction for their correlated random walk. The information corresponding to the position, direction, infectious status and timer status of newborns is appended to the bottom row of each relevant matrix. Deaths occur immediately following births in each timestep, and the dead individuals are randomly selected from the population excluding the newborns of the current step. Analogous to births, for any dead individuals, their rows of the relevant matrices are deleted.

10.2.2 Spatial Interspecies Dynamics

As with the single-species spatially-explicit case, we choose to forego making the demography of the system spatially explicit through introducing systems for predation and resource competition. Predator behaviour in this case would require adjusting the correlated random walk to obtain realistic behaviour in the case where a predator must hunt its prey, and we have more of an interest in simulating the transmission of the infectious disease.

Instead we choose to update our logistic map, to one that uses multiple species. In general, in a system with n species and time step t, this would take the form:

$$x_{t+1}^k = r_k x_t^k - \sum_{i=1}^{n} \phi_{ki} x_t^i x_t^k, \tag{199}$$

where x^k is the proportion of the k-th species relative to its carrying capacity and ϕ_{ij} again refers to the loss in fecundity of species i as a result of species j. Like with the grid class of models or the differential equations, this parameter can be positive or negative. Owing to the complexity of the behaviour from the logistic map with just a single species, we must be extremely careful with our choice of parameters in order to obtain the desired behaviour.

Without an explicit system for predators and prey, we must consider various possible interactions between infectious status and demography. We incorporate interspecies transmission through predation. We do this by infecting the closest susceptible to predator to a dead infected prey with some presupposed probability. In determining births and deaths, we also consider susceptible, infected and recovered individuals separately, and assign each a different probability of being selected as prey, or as being removed through competition or natural deaths. Factoring these in directly through the logistic map poses somewhat of an obstacle.

We can have predators and prey, and introduce different behaviours based on this role and infection status. Chasing is handled by choosing the angle to go towards something, without path tracking; running away is done the same way. We can also amend the probability of infection when interactions occur depending on species and infection status, which is much simpler. Overall, these possibilities are scope for future work.

10.3 Agent-Based Model Code

Here, we leave the MATLAB code used our various models for perusal [61]. The two models presented are the predator-prey agent-based model, and the spatially explicit model.

10.3.1 Lattice-Based Model

Below is the code for the predator-prey lattice-based model used in Chapter 4.3.

```
function [grid,infection_grid, population_states] =...
    predatorpreyautomatonsir(N, time)


    grid = repmat('Z', N, N); % 'Z' represents empty cells

    % Array to track population counts
    population_states = zeros(2, 3, time);

    % Set initial positions for predators and prey
    predator_indices = randperm(N^2, round(0.1 * N^2)); % 10% predators
    prey_indices = randperm(N^2, round(0.3 * N^2));     % 30% prey
    init_predator = numel(predator_indices);
    init_prey = numel(prey_indices);

% Select 30% of the initial predator and prey population for infection
    inf_pred_indices = predator_indices(randperm(init_predator,...
        round(0.3 * init_predator)));
    inf_prey_indices = prey_indices(randperm(init_prey, ...
        round(0.3 * init_prey)));

    % Assign 'Y' to predator positions
    grid(predator_indices) = 'Y';

    % Assign 'X' to prey positions
    grid(prey_indices) = 'X';

    % Set infection matrix, where S is susceptible, I infected, R
    % recovered and E for an empty cell with no host.
    infection_grid = repmat('E', N, N);
    infection_grid(predator_indices)= 'S';
    infection_grid(prey_indices)= 'S';
    infection_grid(inf_pred_indices)= 'I';
    infection_grid(inf_prey_indices)= 'I';

    % Simulation parameters
    num_iterations = time;
    x = 0.06; % Probability empty becomes prey
    y = 0.08; % Probability prey becomes predator
    z = 0.04;  % Probability predator dies
    p_infect_predator = 1; % Prob. predator infected from prey
    p_infect_prey = 0.5; % Prob. prey infected by infected neighbours
    p_recovery = 0.1; % Probability of recovery
    p_loss = 0.05; % Probability of loss of immunity

    % Create a figure for visualization
```

```
47      figure;
48      for t = 1:num_iterations
49          % Update the grid
50          [grid, infection_grid] = ...
51              updateGrid(grid, infection_grid, x, y, z, ...
52              p_infect_predator, p_infect_prey, p_recovery, p_loss);
53          % Convert characters to numeric values for visualisation
54          numeric_grid = zeros(N, N);
55          numeric_grid(grid == 'X') = 1; % Prey
56          numeric_grid(grid == 'Y') = 4; % Predator
57
58          % Darken the colours for infected cells
59          numeric_grid(infection_grid == 'I') = ...
60              numeric_grid(infection_grid == 'I') + 1;
61          numeric_grid(infection_grid == 'R') = ...
62              numeric_grid(infection_grid == 'R') + 2;
63
64          % Visualise the grid
65          imagesc(numeric_grid);
66          colormap([1 1 1; 0 1 0; 0 0.4 0; 0 0.7 0; 1 0 0; 0.4 0 0; ...
                0.7 0 0]);
67          % White, dark green, dark red for Z, X, Y
68          caxis([0 6]);
69          drawnow;
70
71          pause(0.001);
72
73          %Get data
74          prey_susceptible = sum((grid == 'X') & ...
75              (infection_grid == 'S'),'all');
76          prey_infected = sum((grid == 'X') & ...
77              (infection_grid == 'I'), 'all');
78          prey_recovered = sum((grid == 'X') & ...
79              (infection_grid == 'R'), 'all');
80          predator_susceptible = sum((grid == 'Y') & ...
81              (infection_grid == 'S'), 'all');
82          predator_infected = sum((grid == 'Y') & ...
83              (infection_grid == 'I'), 'all');
84          predator_recovered = sum((grid == 'Y') & ...
85              (infection_grid == 'R'), 'all');
86
87          % Record data
88          population_states(1, 1, t) = prey_susceptible;
89          population_states(1, 2, t) = prey_infected;
90          population_states(1, 3, t) = prey_recovered;
91          population_states(2, 1, t) = predator_susceptible;
92          population_states(2, 2, t) = predator_infected;
```

```
93          population_states(2, 3, t) = predator_recovered;
94
95      end
96
97  end
98
99  function [new_grid, new_infection_grid] = ...
100     updateGrid(grid, infection_grid, x, y, z, ...
101     p_infect_predator, p_infect_prey, p_recovery, p_loss)
102
103     N = size(grid, 1);
104     new_grid = grid; % Copy grid to update state
105     new_infection_grid = infection_grid; % Copy infection grid to ...
            update
106
107     for i = 1:N
108         for j = 1:N
109
110             % Calculate neighbour indices with periodic boundary ...
                    conditions
111             north = mod(i-2, N) + 1;
112             south = mod(i, N) + 1;
113             west = mod(j-2, N) + 1;
114             east = mod(j, N) + 1;
115
116             neighbours = [grid(north,j), grid(south,j),...
117                 grid(east, i), grid(west,i)];
118             inf_neighbours = ...
119                 [infection_grid(north,j), infection_grid(south,j), ...
120                 infection_grid(east, i), infection_grid(west,i)];
121             % Count the number of each type in the neighbourhood
122             numPredators = sum(neighbours(:) == 'Y');
123             numPrey = sum(neighbours(:) == 'X');
124             numInfectedPrey = sum(inf_neighbours(:)== 'I' & ...
125                 neighbours(:)=='X');
126
127             if grid(i,j) == 'Y'
128
129                 % Predators die with probability z
130                 if rand() < z
131                     new_grid(i,j) = 'Z';
132                     new_infection_grid(i,j) = 'E'; % Infection ...
                            status reset
133
134                 elseif infection_grid(i,j) == 'I' && rand() < ...
                        p_recovery
135                     new_infection_grid(i,j) = 'R'; % Predator recovers
```

```
136
137                 elseif infection_grid(i,j) == 'R' && rand() < p_loss
138                     new_infection_grid(i,j) = 'S'; % Predator ...
                            immunity loss
139                 end
140
141
142             elseif grid(i,j) == 'X'
143
144                 % Prey becomes predator with probability y*numPredators
145
146                 if rand() < y * numPredators/4
147                     new_grid(i,j) = 'Y';
148                     new_infection_grid(i,j) = 'S'; % Prey becomes ...
                            predator
149                     if rand() < p_infect_predator * numInfectedPrey/4
150                        new_infection_grid(i,j) = 'I'; % Predator ...
                               infected
151                     end
152
153                 elseif infection_grid(i,j) == 'S' &&...
154                         rand() < p_infect_prey * numInfectedPrey/4
155                     new_infection_grid(i,j) = 'I'; % Prey infected
156
157                 elseif infection_grid(i,j) == 'I' && rand() < ...
                        p_recovery
158                     new_infection_grid(i,j) = 'R'; % Prey recovers
159
160                 elseif infection_grid(i,j) == 'R' && rand() < p_loss
161                     new_infection_grid(i,j) = 'S'; % Prey loses ...
                            immunity
162                 end
163
164             elseif grid(i,j) == 'Z'
165                 % Empty becomes prey with probability x*numPrey
166                 if rand() < x * numPrey/4
167                     new_grid(i,j) = 'X';
168                     new_infection_grid(i,j) = 'S';
169                     % Empty becomes susceptible prey
170                 end
171             end
172         end
173     end
174 end
```

10.3.2 Spatially Explicit Model

Below is the code for the spatially explicit model used in Chapter 5.5. The commented out code can be utilised to calculate the basic reproduction number statistically, using a script with one initial infected individual, and counting its secondary infection.

```
function [data]=spatialexplicitmultispecies(time, invader)

% Set initial positions and movement directions

initi=30;

% Set parameters

agents = 1200;
correl = [0.1 0.1 0.1];
step = [0.05 0.05 0.05];
length = 10;
radius = 0.04;
beta = [0.3 0.05 0.05; 0.05 0.3 0.05; 0.1 0.1 0.6];
max = [200 10001];

% Plotting styles
plotstyle = ["ob","vb","^b"; "or","vr","^r" ;"og","vg","^g"];

% Specify species counts

speccount = nnz(correl);

species = [ones(round(agents - invader)/2, 1); ...
    2 * ones(round(agents - invader)/2, 1), ...
    ;3 * ones(invader, 1)];

pos=length*rand(agents,2);
h=ones(1,agents);
v=2*pi*rand(agents,1);
data=zeros(agents,3,time);

% Set infection initial conditions

sir=[3*ones(1, initi),zeros(1, agents-initi)];
sir=sir(randperm(agents))';
timer=zeros(agents,1);

%Graphics prep

```

```
figure
hold on
axis([0 length 0 length])

for i=1:time

    if nnz(sir==3)+nnz(sir==1)==0
        break
    end

    moves=rand([agents,2]);
    order=randperm(agents);

    for j=1:agents

        act=order(j);
        spec = species(act);



      if moves(act,1)<correl(spec)
        theta=moves(act,2)*2*pi;
        pos(act,:)=pos(act,:)+step(spec)*[cos(theta),sin(theta)];
        v(act)=theta;
      else
          theta=v(act);
          pos(act,:)=pos(act,:)+step(spec)*[cos(theta),sin(theta)];
      end

    end


    I=mod(sir,2);
    dist=((pos(:,1)*h-h'*pos(:,1)').^2 ...
        +(pos(:,2)*h-h'*pos(:,2)').^2).^(0.5);
    contag=dist.*I;
    %ph=nnz(sir==3);

    %if ph ≠ 0
     %   orig=double(eq(sir,3));
      %  disp(size(orig))
       % rcount=dist.*orig;
        %disp(size(rcount))
    %end


    for k=1:agents
```

```
88
89          if sir(k)==0
90
91              power = zeros(1,3);
92              neighbours = contag(:,k);
93              kspec = species(k);
94
95
96              for l = 1:speccount
97
98                  specneigh = neighbours.*double(species==l);
99                  power(l) = nnz(specneigh < radius)-nnz(specneigh==0);
100             end
101
102             q=rand();
103
104             if nnz(power) > 0 && q > prod((1-beta(kspec,:)).^power)
105                 sir(k)=1;
106             end
107
108         elseif sir(k)==1 || sir(k)==3
109
110             timer(k)=timer(k)+1;
111
112             if timer(k)>max(1)
113                 sir(k)=2;
114                 timer(k)=0;
115             end
116
117         elseif sir(k)==2
118
119             timer(k)=timer(k)+1;
120
121             if timer(k)>max(2)
122                 sir(k)=0;
123                 timer(k)=0;
124             end
125
126         end
127
128
129
130     end
131
132     % Check for out-of-bounds individuals and adjust
133
134       posnew=length/pi*acos(cos(pi*pos/length));
```

```
135         oob=pos-posnew;
136         oob(abs(oob)<1e-12)=0;
137         rebound=mod(sign(oob),2);
138         v=mod(v-2*rebound(:,2).*v,2*pi);
139         v=mod(v+rebound(:,1).*(pi-2.*v),2*pi);
140         pos=posnew;
141 
142         data(:,1,i)=pos(:,1);
143         data(:,2,i)=pos(:,2);
144         data(:,3,i)=sir;
145 
146         cla;
147         for j = 1:speccount
148             spos=pos.*double(eq(sir,0)).*double(eq(species,j));
149             spos=spos(any(spos,2),:);
150             ipos=pos.*double(eq(mod(sir,2),1)).*double(eq(species,j));
151             ipos=ipos(any(ipos,2),:);
152             rpos=pos.*double(eq(sir,2)).*double(eq(species,j));
153             rpos=rpos(any(rpos,2),:);
154             plot(spos(:,1),spos(:,2), ...
155                 plotstyle(1,j),'MarkerSize',5,'MarkerFaceColor','b')
156             plot(ipos(:,1),ipos(:,2), ...
157                 plotstyle(2,j),'MarkerSize',5,'MarkerFaceColor','r')
158             plot(rpos(:,1),rpos(:,2), ...
159                 plotstyle(3,j),'MarkerSize',5,'MarkerFaceColor','g')
160         end
161         drawnow;
162         pause(.000001)
163 
164 end
165 
166 
167 end
```