Species Invasion in a Network Population Model

Ryan C. Yan

Department of Applied Science,

College of William and Mary,

Williamsburg, VA 23187-8795, USA

Email: rcyan@email.wm.edu

Abstract

The introduction and spread of invasive species is increasingly driven by the expansion of human-made transportation routes. We formulate a network model of biotic invasion incorporating logistic growth and dispersal along a network, and present analyses of the model. We introduce small world networks and use them to investigate the role of network properties and long-distance dispersal on spread dynamics. Lastly we present comparisons between the stochastic and deterministic models to illustrate the effects of stochasticity on invasive species spread dynamics.

Contents

List of Figures

Chapter 1

Introduction

1.1 Background on Invasive Species

The National Invasive Species Council defines invasive species as non-native species whose introduction is likely to cause harm (Committee et al., 2006). As a common and pervasive cause of environmental and economic damage, the problem of invasive species has come to be one of the most pressing issues in ecology today. Invasive agricultural pests alone are estimated to cost the United States $120 billion annually. In addition, 42% of endangered or threatened species in the U.S. are at risk primarily due to invasive species (Pimentel et al., 2005). The severity of the problem is clear, yet attempts at guarding ourselves against invasive species introductions have historically, and still do, fail. Though a portion of these failures can be contributed to sociological issues, such as lack of public awareness and government support to address the issue, we still have only a limited understanding of the spread and proliferation of terrestrial invasive species (Mack et al., 2000). It is the role of modelers to understand these processes well enough to produce accurate insights on them.

The invasion process proceeds in several stages, as explained in Williamson (1989). First is the introduction of propagules or organisms into the new environment. Although

an introduction may occur, a large proportion of of these introductions fail to establish stable colonies. Those that succeed in the establishment phase often exhibit a lag time - a period of low growth followed by rapid proliferation until the population reaches its carrying capacity (Mack et al., 2000). This particular pattern of accelerating and decelerating growth is called logistic growth and will be described further.

Once established in a central location, further range expansion can rapidly occur - the "spread" phase. There are an incredible amount of varied models used to describe the spread or dispersal phase of invasive species spread. Historically, logistic population growth has been paired with diffusive spread to generate predictive models of species spread. Yet many species are known to be spread via long-distance dispersal events via human transport vectors (Carlton, 2003).

Humans and animals have long acted as vectors of dispersal for invasive organisms. Gypsy moths lay their eggsacs on the underside of car bumpers, bivalves can be taken in with ballast water on ships, and exotic pest species can be shipped in infected shipments of agricultural or construction materials (Carlton, 2003). Biotic invasions have occurred long before our time (oftentimes, plant seeds are physically dispersed by hitching onto an animal or person), but since the industrial revolution, the rapid expansion of production alongside the development of new global trade routes has hastened the rate of introductions significantly. (Hulme, 2009).

The perceived importance of long-distance dispersal events in invasive species spread, particularly relating to the human-made transportation network, leads us to choose a network model framework for our research.

1.2 Modeling Background

The invasive species literature is rich with a large volume of research to draw upon to formulate our model. The purpose of this section is twofold in both motivating this study and providing a brief summary of the common modeling techniques in invasive species

literature.

We begin by discussing models of single-patch population growth, discounting dispersal between locations for now. The most common model of population growth in biology is the continuous logistic equation, defined as:

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}) \tag{1.1}$$

where N is the population size, r is the intrinsic growth rate of the population, and K is the carrying capacity.

One criticism of this model is that lacks the inclusion of an Allee effect, characterized as a positive association between population density and individual fitness. As explained in Korolev et al. (2014), a strong Allee effect describes a phenomenon that leads a population to extinction once it dips below a certain threshold. Thus the extension of the logistic growth function with a strong Allee effect describes a population which grows at intermediate population levels but whose growth rate declines at both low and high population. Allee effects can be caused by many different mechanisms that are positively density dependent, such as pack hunting behavior or mate-finding. Concretely, the logistic equation can be modified to include an Allee effect as such:

$$\frac{dN}{dt} = rN(1 - \frac{N}{K}(\frac{N}{A} - 1)) \tag{1.2}$$

where A is a positive constant which defines the threshold below which extinction is ensured. This has significant implications for the growth trajectory of a small populations. The figure below illustrates this.

We can see that in the population undergoing logistic growth without an Allee effect, the growth rate decreases with increasing population size. However, under the Allee effect, the change in density dependence creates a difference in stable states between the logistic growth model with and without an Allee effect. In the context of invasive species, the this may be highly significant in explaining the low success rate of introductions in establishing stable colonies.

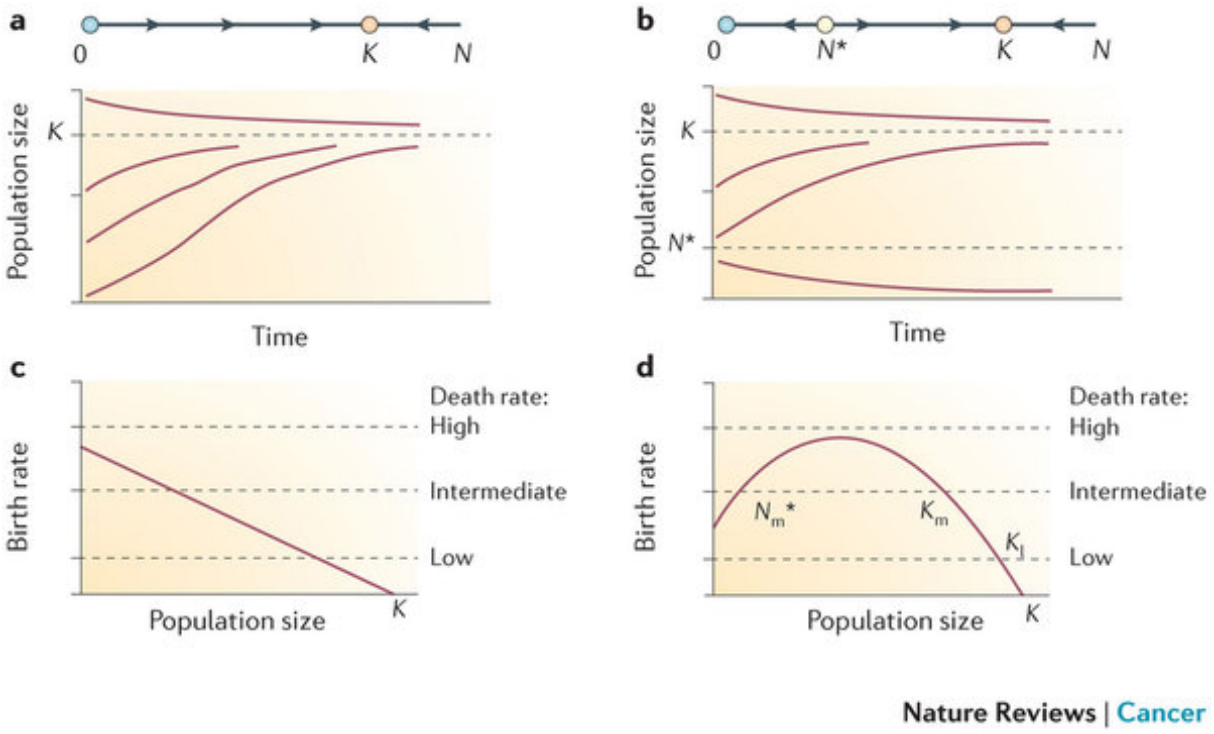

Figure 1.1: Population and per capita growth rate trajectories of populations under logistic growth with and without an Allee effect (Korolev et al., 2014).

The previous models describing single-patch population growth are not spatially explicit. Models of species spread combine population growth with some sort of dispersal mechanism. Historically, a common and relatively simple approach to achieve this is based on reaction-diffusion equations. An early example of this is a form of Fisher's equation, taken from Neubert and Parker (2004):

$$\frac{\partial N}{\partial t} = rN(1 - \frac{N}{K}) + D\frac{\partial^2 N}{\partial x^2} \tag{1.3}$$

Here, we present a simple one-dimensional form, with $N(x,t)$ representing the population density at time t and location x. The parameters r and K denote the intrinsic growth rate and the environmental carrying capacity respectively. Finally, D denotes the diffusion coefficient. This type of model presents an oversimplified view of population spread,

failing to take into account variables such as age-structure or environmental factors such as landscape heterogeneity. The result is that these models describe species invasion as a solid advancing front, where in reality, the spread dynamics are much more complex. For example in recorded terrestrial invasions such as the Coypu rodent in Europe and North America, the spread of the species is not synchronous, but rather there exist isolated colonies ahead of the front as well as locations behind that front that remain uncolonized by the species (Reeves and Usher, 1989).

Several different approaches developed to incorporate various aspects left out by the relatively basic reaction-diffusion equations. Integro-differential equations (IDEs) as described in Neubert and Parker (2004) express the process of spread in two phases. For a model of population spread along a single dimension, we would first calculate the change in local population density according to the equation

$$N(y, t+1) = f[N(y, t)] \tag{1.4}$$

where $N(y, t+1)$ is the population density at location y at time $t+1$, which is arrived at from applying a growth function f on the population at time t. Secondly, the population is redistributed by the density kernel as shown below:

$$N(x, t+1) = \int_{-\infty}^{+\infty} k(x, y) f[N(y, t)] dy \tag{1.5}$$

where $N(x, t+1)$ is the probability of redistributing from point y to point x given the probability density function $k(x, y)$. The main advantage of IDEs are their flexibility in choosing the density kernel, which allows for more complex redistribution of population. However, Neubert and Parker (2004) also describes a method for converting an IDE into a stage-structured matrix model. Stage-structured models are common in population biology and allow us to distinguish between birth, death, and migration rates of different life stages, which are often dissimilar.

In recent years the growth in computational power and abundance and communication of ecological data has spurred the use of more data-driven species distribution models

(SDMs). These models, as explained in Václavík and Meentemeyer (2009), use presence and absence data with environmental data to create a mathematical model of species distribution in environmental space. With the appropriate geographically mapped data, for example, data layers in a Geographical Information System (GIS) data set, one could generate potential landscapes that the species could inhabit.

As reviewed in Elith and Leathwick (2009), there is much debate about model selection and predictive capabilities of SDMs. Historically SDMs have been used to an explanatory effect, answering the question of why the present species distribution is what it is. However, in the case of invasive species we wish to use the model for extrapolation. One problem with this is that a strong assumption underlying SDMs is that the species in question is in equilibrium with its environment. Essentially, this means that we expect the species to be present in all of its suitable habitats. Whether this is a reasonable assumption of an invasive species is debatable, since they face clear dispersal limitations being newly introduced to their environment. Despite these concerns, SDMs are being used more and used for extrapolations, and by linking them with dispersal models to generate predictions of invasive species spread.

Lastly, we reach a relatively new approach and the topic of this thesis: network models of invasive species. Networks are widely studied and applied to a multitude of fields, including invasive species biology. In this field, marine species have received more attention than terrestrial species. A recent, highly cited paper by Kaluza et al. (2010) posed a network framework of marine bioinvasion, with links between ports weighted by observable shipping traffic along these routes. Despite the acknowledged importance of studies such as (Kaluza et al., 2010) and others (Floerl et al., 2009), much of this research is focused on applying a data-driven model to only a few species. Yet not as much work has been done in the general analysis of network models of invasive species, although we are seeing a rise in their usage.

The spread of invasive species across disparate terrestrial landscapes lends itself to

modeling in a network framework. In addition, the new availability of data, both ecological species data and human-centric transportation data will facilitate studies such as these for terrestrial species. Following this premise, the goal of our research is to present a general model of terrestrial invasive species spread and analyze the individual components of that model. Over time, data driven, computationally-intensive models such as these may become more attractive due to increases in data availability and computational power. This will afford us to more powerful predictive tools, but we should aim to understand the general properties of network science and invasion biology that underlie their use.

1.2.1 Thesis Outline

In this thesis we will formulate and analyze a network model of biotic invasion using steady state analysis as well as numerical modeling. We divide the thesis into three main sections discounting the introduction and conclusion. In the first section we will derive the deterministic model and analyze its steady state properties in 1 and 2 dimensions. We also define the n dimensional vectorized model we use for computational simulations. In the section on small world networks, we introduce the characteristics of these networks and their relation to biotic invasion. Lastly, we introduce the stochastic model and explore the role of stochasticity in different parameter regimes in the invasion model.

Chapter 2

Deterministic Modeling

2.1 1-Dimensional Model Formulation

2.1.1 Rationale

We begin by defining a single-patch model, where the dynamics are described simply by logistic growth. Logistic growth is chosen as it is a standard model of a population of individuals growing within an area with restricted resources, described by equation (1.1) (Strogatz, 2014).

In the wider model of invasive species, this represents the establishment phase of invasive species spread, once it has been newly introduced to a location. In this section, we define a discrete time logistic map analog to the continuous logistic equation. The main reason we chose not to use the continuous time logistic equation was because using our map formulation here reduced the computational time significantly over integration of the continuous logistic growth function.

2.1.2 Model Specification

We begin by deriving the discrete map analog to the standard logistic equation presented above in equation 1.1. This will map a population within a single patch forward in yearly time steps.

Through separation of variables and direct integration, we come to the general solution to the continuous time logistic equation:

$$N(t) = \frac{Ce^{rt}}{1 + \frac{Ce^{rt}}{K}} \tag{2.1}$$

where C is the constant of integration. Setting the initial population at time zero equal to some value N_0, we solve for C, and arrive at the equation below:

$$C = \frac{N_0}{1 - \frac{N_0}{K}} \tag{2.2}$$

Replacing C in equation 2.2 yields the particular solution

$$N(t) = \frac{N_0 e^{rt}}{1 + \frac{N_0(e^{rt}-1)}{K}} \tag{2.3}$$

Now noting that given an initial population value N_0 at time $t = 0$, we can calculate the population at $t = 1$. We define the resulting discrete map with a 1-year time step with the equation:

$$N_{t+1} = g(N_t) = \frac{N_t e^{r}}{1 + \frac{N_t(e^{r}-1)}{K}} \tag{2.4}$$

2.1.3 Alternative Logistic Models

A common discretization of the logistic differential equation (1.1) is the standard form of the recurrence relation form shown below:

$$N_{t+1} = rN_t(1 - \frac{N_t}{K}) \tag{2.5}$$

Although relatively simple looking, this formulation of the discrete logistic map reveals undesirable bifurcative properties and chaos as we increase the intrinsic growth rate r

from 0 to 4. We chose to use this formulation because of the lack of chaotic behavior at large values of r. Our formulation exhibits the typical sigmoidal curve characteristic of the logistic equation without atypical behavior.

2.2 1-Dimensional Stability Analysis

2.2.1 Fixed points

How does the population grow as described by the discrete logistic map we formulated in equation 2.4. To answer this question, a common approach is to investigate the steady state stability of the system. We follow the approach described in Strogatz (2014). We first look to discover the fixed points of the system. A fixed point of g is a point N^* that satisfies $N^* = g(N^*)$ where $g(N)$ is given by 2.4. Solving the equation below for N^*,

$$N^* = \frac{N^* e^r}{1 + \frac{N^*(e^r - 1)}{K}} \tag{2.6}$$

we find the fixed points $N^* = 0$ and $N^* = K$ satisfy this equation. The zero fixed point, $N^* = 0$ is the extinct state. The positive fixed point is K, which is the carrying capacity.

To determine how the system will react to perturbations from a fixed point, we investigate their stability.

2.2.2 Stability of Fixed Points

Concretely, we will investigate the response of the system to a small perturbation defined by ϵ_t. Suppose the population at N_t is at the fixed point N^*, then

$$N_t = N^* + \epsilon_t, \text{ where } |\epsilon_t| << 1$$

We can approximate how the population changes at the next time step by finding the Taylor series expansion around the fixed point as shown below:

$$N_{t+1} = g(N^* + \epsilon_t)$$

$$N_{t+1} \approx g(N^*) + g'(N^*)\epsilon_t + O(\epsilon_t^2)$$

$$N_{t+1} \approx N^* + g'(N^*)\epsilon_t$$

Because ϵ_t is small, we ignore the higher order terms because they will be negligible. The size of the perturbation at time $t+1$ is then defined as:

$$\epsilon_{t+1} \approx g'(N^*)\epsilon_t$$

A fixed point will be stable if the perturbations near it get smaller and smaller as time goes on. Perturbations will grow larger and away from an unstable fixed point in its vicinity. If $|g'(N^*)| < 1$, then N^* is a stable fixed point. If $|g'(N^*)| > 1$, then N^* is an unstable fixed point. Note that the assumption of ϵ_t is small means that these conclusions only hold in the local neighborhood of the fixed point.

We calculate the value of $g'(N^*)$ for fixed points $N^* = 0$ and $N^* = K$ in our model, using the equation:

$$g'(N_t) = \frac{e^r}{(1 + \frac{N_t(e^r - 1)}{K})^2} \tag{2.7}$$

For the fixed point $N^* = 0$, we find that $g'(N^*) = e^r$, and for the fixed point $N^* = K$, we find that $g'(N^*) = e^{-r}$. This means that the zero fixed point is stable for $r < 0$ and unstable for $r > 0$. The positive fixed point is stable for $r > 0$ and unstable for $r < 0$.

2.2.3 Interpretation

In a biological sense, the fixed points represent where the population neither grows nor shrinks. When the population is 0, it tends to stay at zero. When the population has reached the environmental carrying capacity, the birth rate matches the death rate and growth halts.

Recall that the parameter r is the intrinsic rate of growth for the population N. Though negative growth rates are possible, the populations we are modeling under typical logistic growth have positive growth rates associated with them. In terms of invasive

species populations, the results of our fixed point analysis reveal that once an invasive is introduced to a new location, it will tend to grow until it reaches its carrying capacity. However, in more intricate models, this will not always be the case. For example, introduction of a strong Allee effect will induce a negative growth rate at low population densities, making the zero fixed point stable.

2.3 Model Formulation in Higher Dimensions

In this section we define a general form of the model in n dimensions. With multiple patches, population dynamics depend on both within-node logistic growth as well as immigration and emigration between nodes.

The migration rates between the nodes are determined by the transition matrix $\mathbf{P}$,

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1j} & \cdots & p_{1n} \\ \vdots & \ddots & & & \vdots \\ p_{i1} & & p_{ij} & & p_{in} \\ \vdots & & & \ddots & \vdots \\ p_{ni} & \cdots & p_{nj} & \cdots & p_{nn} \end{bmatrix} \tag{2.8}$$

where element p_{ij} represents the proportion of population in j transferred from patch j to patch i in a single time step. The columns of $\mathbf{P}$ sum to 1, meaning that the outgoing population from each node is conserved.

We use a directed network graph to illustrate the connections defined by the transition matrix in a 3-patch model in figure 2.1.

The population of each node at each time t is stored in the state vector $\mathbf{s}_t$, an n

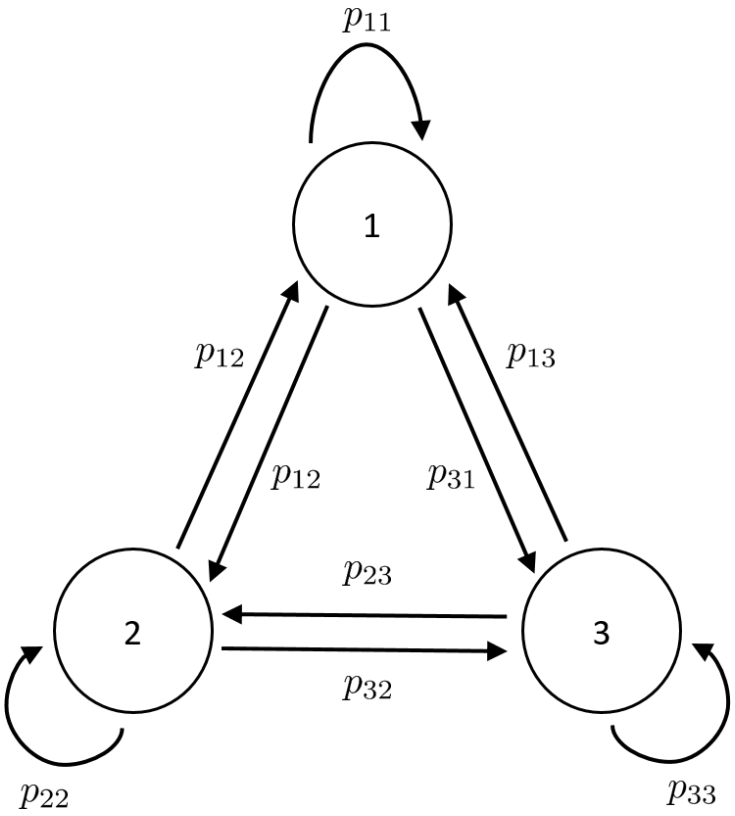

Figure 2.1: This is a connected network model in 3 dimensions. Each patch is connected with both its neighbors and itself, with edge weights denoted by p_{ij} which represent the proportion of population transfered from patch j to i in a single year.

dimensional column vector defined as

$$\mathbf{s}_t = \begin{bmatrix} s_{t,1} \\ \vdots \\ s_{t,i} \\ \vdots \\ s_{t,n} \end{bmatrix} \tag{2.9}$$

where element $s_{t,i}$ represents the number of individuals in node i in year t.

We define a mapping of state vector $\mathbf{s}_t$ from year t to $t+1$ as

$$\mathbf{s}_{t+1} = \mathbf{P}\mathbf{g}(\mathbf{s}_t) \tag{2.10}$$

with the vector function

$$\mathbf{g}(\mathbf{s}_t) = \begin{bmatrix} g(s_{t,1}) \\ \vdots \\ g(s_{t,i}) \\ \vdots \\ g(s_{t,n}) \end{bmatrix} \tag{2.11}$$

Thus by applying the growth function defined in equation 2.4 element-wise to each population in $\mathbf{s}_t$ and multiplying the transition matrix $\mathbf{P}$ by the resulting column vector $\mathbf{g}(\mathbf{s}_t)$, we have our n dimensional determinstic network model of population growth and spread on the network.

2.4 Analysis of 2-Dimensional System

With the new notation in mind, we can describe the most general case of a two-patch system for analysis:

$$s_{t+1,1} = f_1(\mathbf{s}_t) = \frac{p_{11}s_{t,1}e^r}{1+\frac{s_{t,1}(e^r-1)}{K}} + \frac{p_{12}s_{t,2}e^r}{1+\frac{s_{t,2}(e^r-1)}{K}} \tag{2.12}$$

$$s_{t+1,2} = f_2(\mathbf{s}_t) = \frac{p_{22}s_{t,2}e^r}{1+\frac{s_{t,2}(e^r-1)}{K}} + \frac{p_{21}s_{t,1}e^r}{1+\frac{s_{t,1}(e^r-1)}{K}} \tag{2.13}$$

Similarly to the 1-dimensional analysis, we want to find the fixed points of the above system, s_1^* and s_2^*. We define s_i^* as the fixed point value of patch i in the state vector, where $f_i(s_i^*) = s_i^*$. Checking for the fixed points as we did before, we find that the extinct state at $(0,0)$ remains. Intuitively following the results from the 1-dimensional analysis, we predict the existence of a stable positive non-zero state. Under the assumption of symmetry in our transition matrix, meaning in this case that $p_{11}+p_{12} = 1$ and $p_{21}+p_{22} = 1$,

we find the fixed point to be (K, K), as predicted. Without this assumption of symmetry, in the model's most general form, we solve a complicated polynomial expression to arrive at an analytical form for the positive fixed point. We do not attempt to solve it here.

2.4.1 Stability Analysis

Similar to in 1-d analysis, we investigate the stability of the extinct state to see how the system evolves from a small perturbation over time. From Strogatz (2014), we can describe the the evolution of the system characterized by the Jacobian matrix:

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial s_1} & \frac{\partial f_1}{\partial s_2} \\ \frac{\partial f_2}{\partial s_1} & \frac{\partial f_2}{\partial s_2} \end{bmatrix}$$

We assume $K = 1$ for simplicity to calculate the Jacobian. We can make this assumption without loss of generality, assuming that each value s_i, t represents a fractional value of the number of individuals the environment can sustain at carrying capacity.

From Strogatz (2014), we can characterize the stability of the steady states by determining the eigenvalues of the Jacobian evaluated at the fixed points. We evaluate the Jacobian at $(s_1^*, s_2^*) = (0, 0)$ below:

$$J_{0,0} = \begin{bmatrix} p_{11}e^r & p_{12}e^r \\ p_{21}e^r & p_{22}e^r \end{bmatrix}$$

Solving the characteristic equation for the eigenvalues we arrive at

$$\lambda_1, \lambda_2 = e^r(a_{11} + a_{22}) \pm 2e^r\sqrt{a_{11}^2 + a_{22}^2 + a_{11}a_{22} + a_{21}a_{12}}$$

If the magnitude of the eigenvalues are less than 1, then the fixed point is stable. We take the positive root

$$\lambda_1 = e^r(a_{11} + a_{22}) + 2e^r\sqrt{a_{11}^2 + a_{22}^2 + a_{11}a_{22} + a_{21}a_{12}}$$

and set up the inequality describing the relation of the transfer coefficients in the case that λ_1 is an unstable fixed point.

$$\lambda_1 > 1$$

and arrive at the following relation:

$$\ln(a_{11} + a_{22} + 2\sqrt{(a_{11} + a_{22})(a_{11} - a_{22}) + a_{11}a_{22} + a_{21}a_{12}}) < r$$

We are still working to come up with a succinct form or intuitive expression of r, but we can come up with a reasonable way to assure ourselves that 0 is an unstable fixed point. If we reason that $a_{11}, a_{22} \approx 1$ and $a_{21}, a_{12} << 1$, which is true, we can simplify this expression to approximately $\ln(4) > r$. So $r \approx 1.386$ is the bifurcation point. We use small values of $r << 1$ in this study so we can assume that 0 is an unstable fixed point for the 2 dimensional system.

2.5 Numerical Methods

Due to the complexity of analyzing the general network model in higher dimensions, in this section we introduce the deterministic models we use for numerical simulations. This includes the discrete map as well as an integrated ODE model.

2.5.1 Discrete Map

The discrete map defined by equation 2.10 was implemented in R directly and was used for all aggregate results in the small world network section.

2.5.2 ODE Model

We used an integrated ODE model to compare against the stochastic model in the stochastic modeling section. To see whether our discrete map derivation agrees with the original ordinary differential equation (ODE) model, we generate an integrated ODE model using the 'deSolve' package with the logistic equation 1.1 which we modify to incorporate the immigration and emigration terms derived from the transition matrix $\mathbf{P}$ in 2.8. We define

this as:

$$\frac{dN_i}{dt} = rN_i(1 - \frac{N_i}{K}) + \sum_{j=1, j\neq i}^{n} \mathbf{P}_{ij}N_j - \sum_{j=1, j\neq i}^{n} \mathbf{P}_{ji}N_i \tag{2.14}$$

The term $\mathbf{P}_{ij}N_j$ is the number of individuals patch j sends to patch i. The immigration term then, is the sum of $\mathbf{P}_{ij}N_j$ over all patches j that are not the patch in question, i. Emigration is defined similarly, though it is the sum of all outgoing individuals, $\mathbf{P}_{ji}N_i$ over all patches that are not the original patch i.

The values of p_{ij} in the transition matrix rely on the migration rate v, which is a newly introduced parameter. Recall element p_{ij} is the proportion of population each patch transfers out over a yearly time step. The parameter v is the proportion of population from patch j that is sent to each of individual outward connections. Because the population is conserved, the values along the matrix diagonal values are set by the equation:

$$P_{jj} = 1 - d_{\text{out}}v$$

where d_{out} is the out degree of patch j. All off-diagonal non-zero entries are then set to v.

To simulate a biotic invasion over a given time period, we initialize the state vector as $\mathbf{s}_0 = \vec{0}$. Then, we set the value of the patch where the invasion begins in $\mathbf{s}$ equal to carrying capacity K, and all other patches equal to 0. Then we apply the particular model: discrete map, integrated ODE, or Monte Carlo, which will be introduced later, for the given time period. Shown in figure 2.2 is a comparison between the two deterministic methods described above.

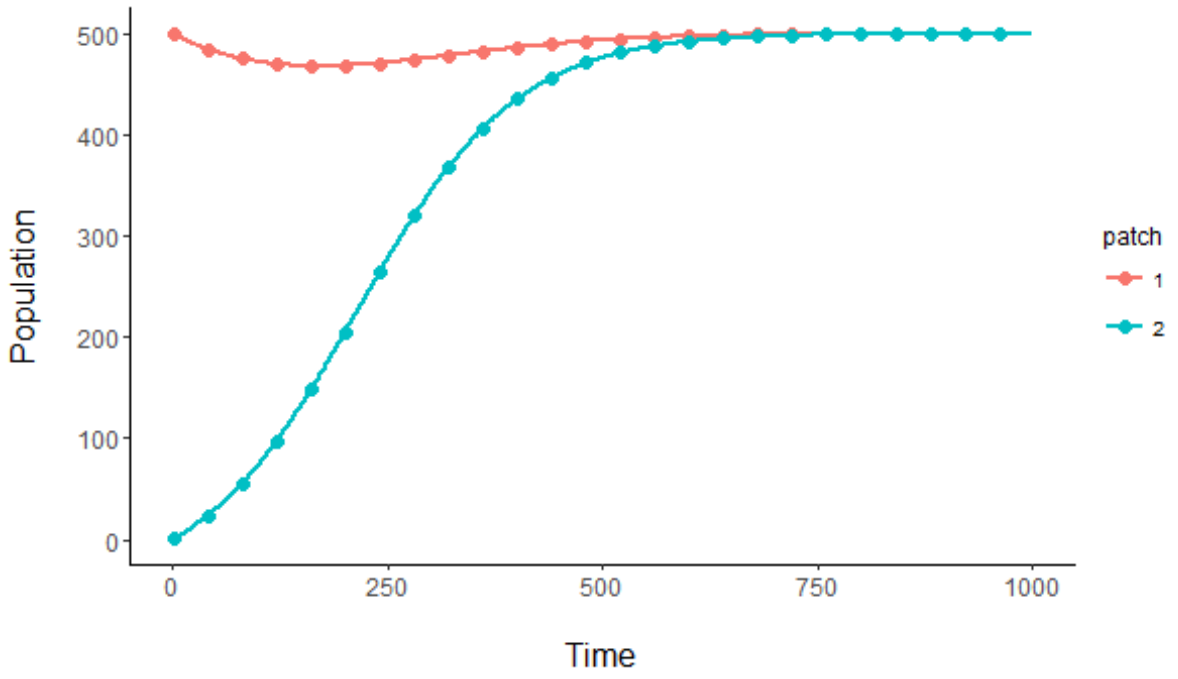

Figure 2.2: Figure showing the agreement between the ODE model and the discrete map. The solid lines are the patches in the ODE model, and are overlaid with points every 40 time steps from the discrete map model. The parameter values given are listed: $n = 2$ patches, migration rate $v = 0.001$, equal birth and death rates $b = d = 0.01$, and carrying capacity $K = 500$.

Chapter 3

Small World Networks

3.1 Background

Networks are often characterized by commonly studied properties of their nodes and edges. Two of these properties are called the clustering coefficient and characteristic path length, and are discussed at length in Watts and Strogatz (1998). Consider a vertex V with k_V neighbors on an undirected graph. Then this vertex can have at most $k_V(k_V - 1)/2$ connections with its neighbors. Let clustering coefficient C_V denote the fraction of these connections that actually exist. Intuitively, the average clustering coefficient of a graph, C, would relate the cliquishness of its nodes - how close neighborhoods of nodes are to being complete graphs, where each pair of distinct nodes is connected by a unique edge. The characteristic path length, L of a graph is the shortest path length between two vertices, averaged over all unique pairs of vertices.

Historically, many real world networks have been classified as mostly regular or random. Regular networks are defined as networks where each node has the same number of connections. Random networks are defined wherein the connections between nodes are determined randomly. Small world networks were introduced by Watts and Strogatz (1998) as a sort of middleground between the two. It is a common property of regular

networks, to be highly clustered. Random graphs, on the other hand would lack clustering but exhibit small characteristic path lengths due to the existence of many randomly introduced connections between distant nodes.

We are interested in small world networks because they could be analogous to the movement network of invasive species. Population networks of invasive species often develop by a combination of movement between neighboring habitats and dispersal along a long distance transport vector, for example, a truck carrying a soil shipment across the United States. Indeed, there have been many research papers discussing small-world properties of real world networks, including transportation networks (Latora and Marchiori, 2001). By analyzing the impact of randomness on a small world network, we hope to learn about the effect of long-distance dispersal in biotic invasions.

3.2 Small World Methods

The original paper generates a small world network by rewiring a regular ring lattice into a random graph. We begin with the regular ring lattice, which we define as a regular network where the nodes are arranged into a ring, and begin randomly choosing edges and randomly reassigning their endpoints, based on a parameter of rewiring probability p. An example of this process is shown in figure 3.1.

The increase in rewiring probability is correlated with a decline in characteristic path length and mean clustering coefficient, seen in figure 3.2.

3.2.1 Extension to Invasion Model

In this section we implement a portion of the procedures including generating the small world network in R using the package 'igraph'.

The algorithm for constructing the small world network using igraph begins by constructing a ring lattice with each of the n nodes connected to its nearest $2m$ neighbors.

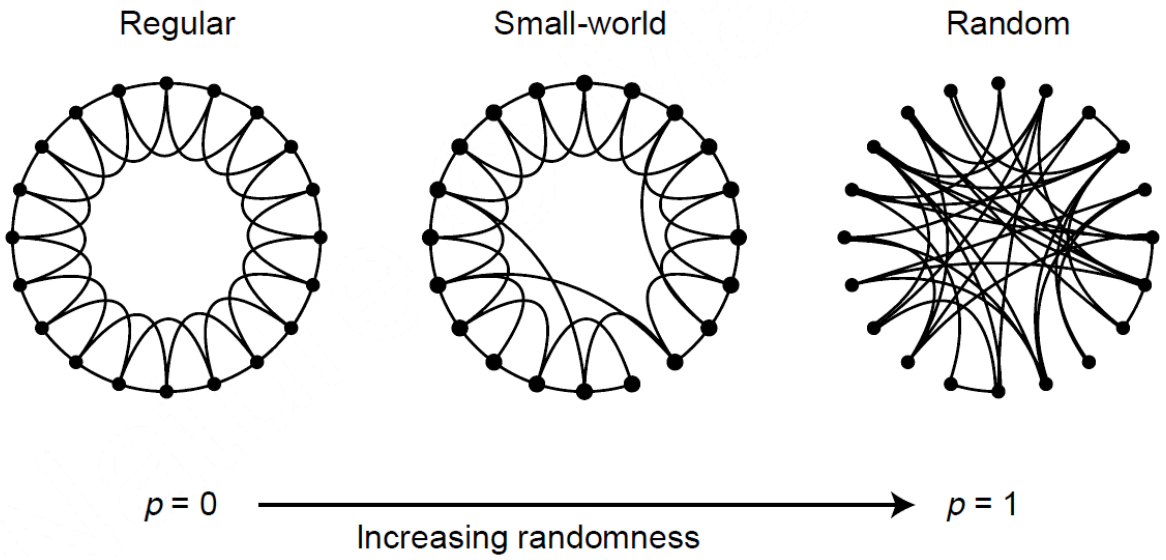

Figure 3.1: Diagram shows the process of rewiring a regular network to introduce small world properties. Starting with a regular network, increasing the probability of rewiring from $p = 0$ to $p = 1$ results in an increasingly random network. (Watts and Strogatz, 1998)

We choose $m = 2$ and $n = 50$. Rewiring is done by replacing each edge e_{ij} with edge e_{xy} with probability p. Nodes x and y are drawn with uniform probability from all possible values. Rewiring events resulting in a self-loop or a duplicated path are redrawn by the igraph algorithm. We imposed an additional check at the end of the rewiring process so that we entirely reconstructed any disconnected networks, meaning networks where there are unreachable nodes. This implementation slightly differs from the original Watts and Strogatz method in that both endpoints of the edge e_{ij} are randomized here as opposed to just one in the original method. However, this difference did not seem significant to the underlying small world network properties. To validate the igraph method we recreated the plot of characteristic path length and clustering coefficient from Watts and Strogatz

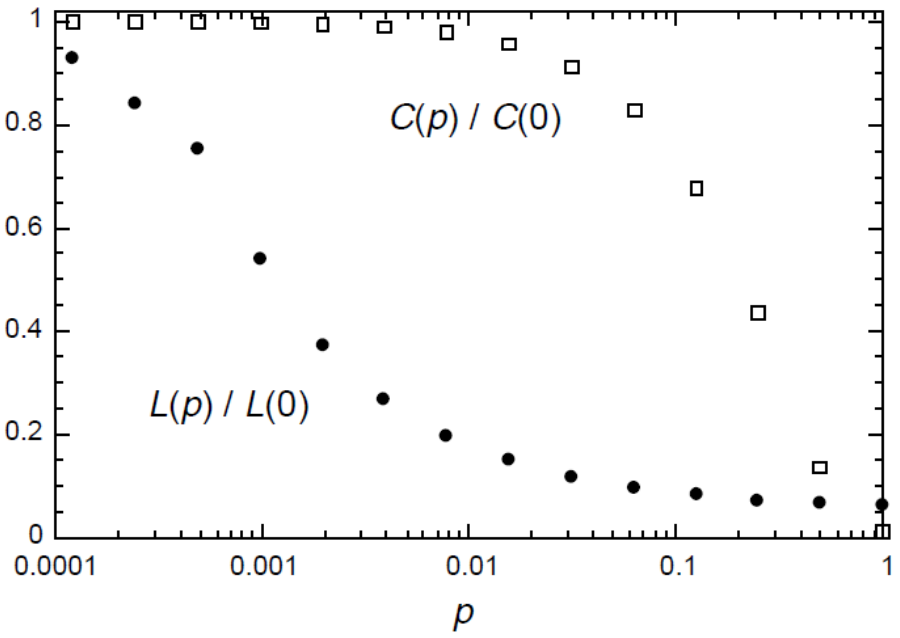

Figure 3.2: Plots showing relationship between characteristic path length and clustering coefficient with rewiring probability. $C(p)$ and $L(p)$ refer to the values of each property for a network generated with given rewiring probability p. Notice that the plotted values are normalized by their values calculated on a regular graph. (Watts and Strogatz, 1998)

(1998) in 3.3. Viewing these graphs side by side, we can tell that the functional forms are nearly identical.

To adapt this small world network to fit our model of species invasion, we transform its adjacency matrix. The adjacency matrix of a network is defined as a matrix $\mathbf{A}$ such that every element $\mathbf{A}_{ij}$ has a binary value 1 or 0. A value of 1 indicates a connection between nodes i and j whereas 0 indicates no connection. Because the small world network is undirected, the adjacency matrix is symmetric, so that $A_{ij} = A_{ji}$.

To map the adjacency matrix to the transition matrix, we perform the following transformations:

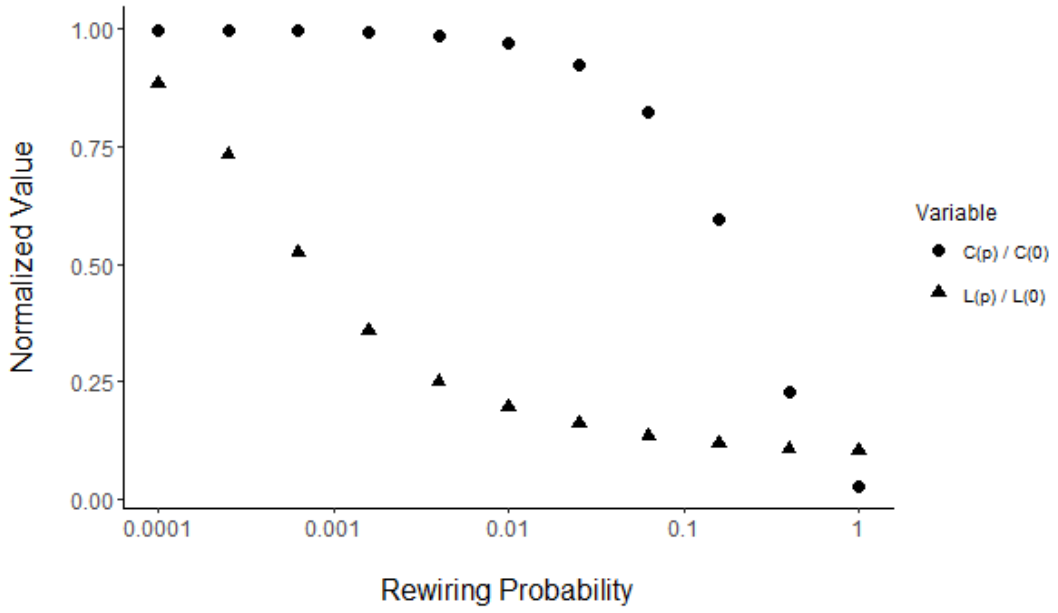

Figure 3.3: Recreation of figure 3.2 in 'igraph' R package. Note there is a small discrepancy between the values due to an algorithmic difference in the rewiring process between the original paper and igraph implementation.

1. $\mathbf{P}_{ij} = v\mathbf{A}_{ij}$ for $i \neq j$

2. $p_{jj} = 1 - \sum_{i=0}^{n} v\mathbf{A}_{ij}$

In this definition, $\mathbf{P}$ refers to the transition matrix defined in 2.8, and v is the migration rate, the proportion of population sent from node p_{ij} to each other node.

Consider the network in figure 3.4 of 50 nodes labeled from 1 to 50 counterclockwise in a ring lattice. We use the transformed adjacency network to map an invasion with yearly time steps over the course of 1000 years. We record a metric $T(p)$, called the time to establishment, which we define on a network rewired with probability p, as the number of time steps it takes for the population to spread from the initial node 1 to the one diametrically opposite, node 26, and surpass a population threshold arbitrarily set to 100 individuals.

Consider one experiment to be defined as generating a small world network with rewiring probability p, transforming the adjacency matrix, and running a simulation as described above.

We performed a series of experiments at different values of p and aggregated the results over 1000 networks for each value of p. The results are presenting in the next section.

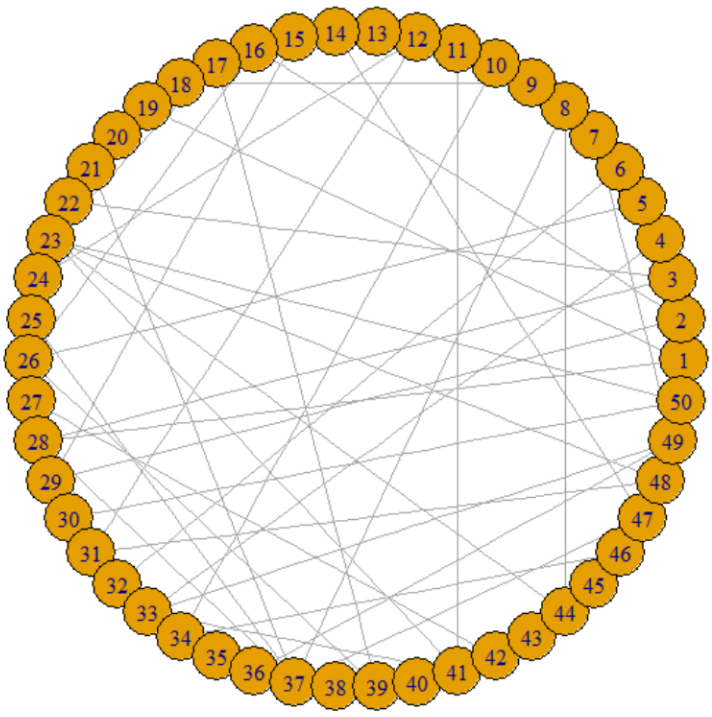

Figure 3.4: Example of the smallworld network generated by igraph with $n = 50$, $m = 2$, and $p = 0.25$.

3.3 Results

We plot the normalized time to establishment, that is, $T(p)/T(0)$ versus the rewiring probability p in figure 3.5. We plot the absolute time values (non-normalized), in figure 3.6.

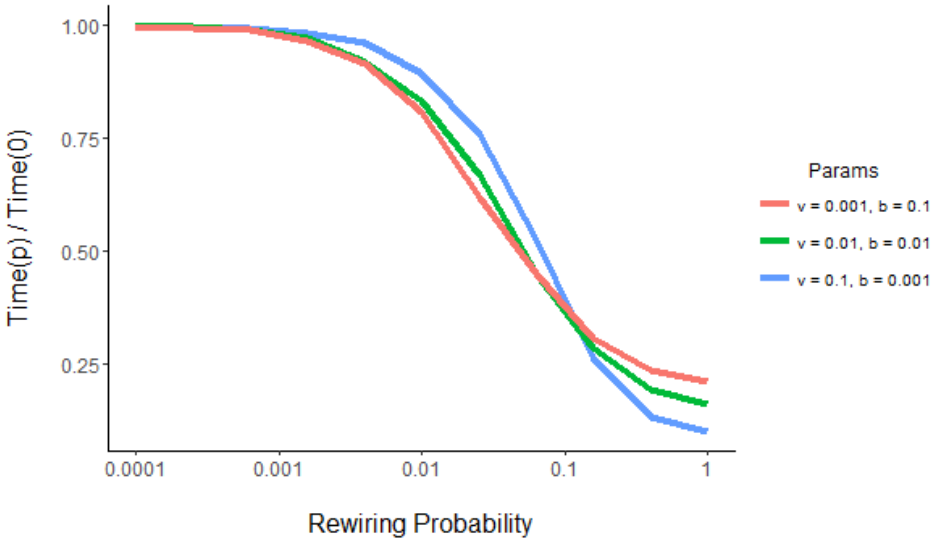

Figure 3.5: Plot of the time to establishment of a species invading a small world network with varying sets of migration rates v and birth and death rates b. Time is normalized by the time to establishment on a regular ring lattice.

Comparing the plots of normalized time to establishment to figure 3.3, we find that the normalized time to establishment, $T(p)/T(0)$, shares the same functional form as $C(p)/C(0)$. This is interesting because we expected these paths to relate to the characteristic path length since it is directly related to the number of long-distance connections between distant nodes on the ring. It is not immediately clear why the time to establishment should appear more analogous to the clustering coefficient than the characteristic path length, but warrants further investigation.

Following the paths in the figure 3.5, we see that $T(p)/T(0)$ is largely dependent on

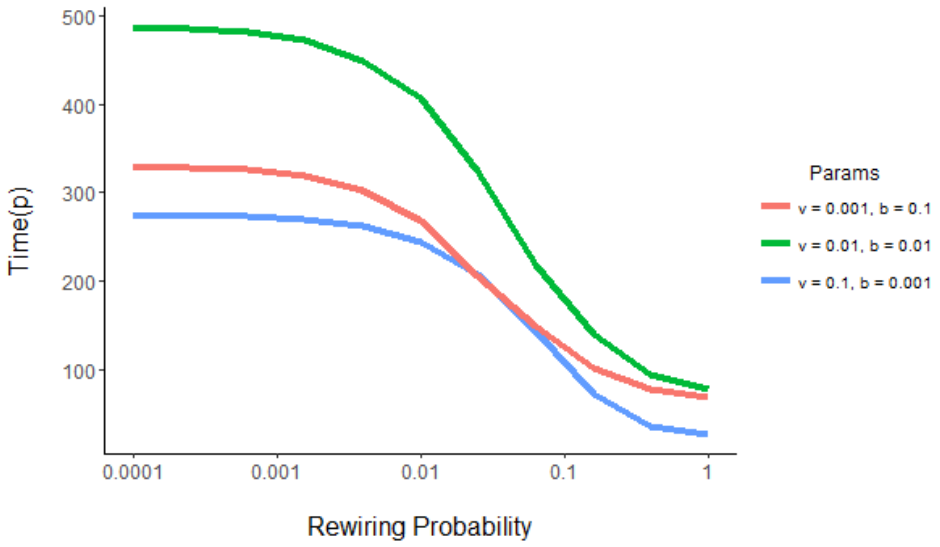

Figure 3.6: Plot of the time in years to establishment of a species invading a small world network with varying sets of migration rates v and birth and death rates b.

migration rate v. Starting from $p = 0$, the paths diverge and settle at $p = 1$ in order of the magnitude of the migration rate. The larger v is the smaller $T(1)/T(0)$. This suggests that as the number of long-distance connections increases, the effect of migration rate on the speed the invasion occurs increases. This is reasonable, as in a highly connected graph with high migration rate, the first few time steps could find the invasive species spread all around the network, whereas this would not occur as quickly with a slower migration rate.

If we compare this to the paths of charting absolute time to establishment in figure 3.6, that is, the number of time steps, we notice two main differences. First, we notice that the values of $T(p)/T(0)$ between the paths diverge as p increases, the absolute values of $T(p)$ converge as p increases. One possible reason for this is that in the initial regular graph, the shortest path to between opposite nodes is longer than in the random graph. Different combinations of parameters v and b could produce slower or faster spreading invasions. As that path length decreases as p increases, any difference in the spread rate

between parameter combinations is diminished since there is less distance to cover.

Additionally, we find that the curves tracing absolute time to establishment vs. p converge in a different order at $p = 1$ than those of normalized time to establishment. Here, the curve with the intermediate value of v and the lowest value have switched places. It seems that in absolute terms, the curve with $v = 0.1$ still results in the fastest establishment. There is possibly some interplay between birth and death rate values and invasion dynamics that we do not currently understand.

Chapter 4

Stochastic Modeling

4.1 Stochastic Model

The stochastic simulations are carried out in a continuous-time Monte Carlo simulation using the Gillespie algorithm (Gillespie, 1977). Generally, we calculate what happens over a series of time steps, where in each time step one individual is either born, dies, or migrates between nodes. We keep track of the number of individuals in each patch and determine what happens at each time step based on the probability of each state transition: birth, death, or migration.

The birth and death rates b_i and d_i respective, for a population N_i in each node, were chosen as:

$$b_i = rN_i \tag{4.1}$$

$$d_i = r\frac{N_i^2}{K} \tag{4.2}$$

This choice is easily related back to the logistic ODE in equation 1.1. We can expand the equation and rearrange the terms as:

$$\frac{dN}{dt} = rN - r\frac{N^2}{K} \tag{4.3}$$

where the birth and death rates are represented by the positive and negative terms respectively.

The total migration rate M is defined as the total number of individuals in the state vector. Notice this migration rate is defined for the implementation of the Monte Carlo simulation, and is different from v, the parameter used to populate the transition matrix $\mathbf{P}$.

$$M = \sum_{i=1}^{n} N_i \tag{4.4}$$

Transitioning between different population states is a Poisson process with a mean rate of λ. λ is equal to the sum of all birth, death, and migration rates,

$$\lambda = \sum_{i}^{n} b_i + \sum_{i}^{n} d_i + \sum_{i}^{n} N_i$$

and the time between events is then exponentially distributed with mean λ^{-1}.

The algorithm to simulate a biotic invasion is implemented as described: First, initialize the state vector s_t as described for the deterministic models. At each time step, record the values in each patch in s_t. To generate the time to the next event τ, we first draw a uniformly distributed random number u, between 0 and 1. This is converted to an exponentially drawn number by equation

$$\tau = \frac{ln(\frac{1}{u})}{\lambda} \tag{4.5}$$

After we have found the time step, we determine whether a birth, death, or migration event has occurred. This is done by assigning a numerical range to each event proportional to the fraction of its rate over the total rate λ, then seeing where u falls after normalizing it to the width of the range. We repeat the process of drawing uniform random numbers and placing them within a range to determine which patch or patches are involved in the event that occurs.

To generate a plot of population vs. time that we can compare against the deterministic model, we run the algorithm described above for 1000 time steps, and aggregate the results over 1000 replicates.

4.1.1 Results

In the introduction we stated that a large proportion of invasive species introductions do not establish stable colonies, and that those that do experience a lag period before growing. However, in the deterministic model analysis section, we showed that the extinct state of our population growth function is unstable. This means deterministically we would expect any introduction to grow to carrying capacity. The difference between the deterministic model and biological reality is that in reality stochastic effects can play a large role in a colony's establishment. Stochastic effects are particularly significant in populations with a small number of individuals and migration rates, both of which are properties of invasive species introductions.

Here we present some results from the comparison of the deterministic ODE model and the stochastic Monte Carlo simulation. The following are multi-panel plots of 1000-year simulations of the deterministic model (solid line) and the stochastic model (dashed line) with constant parameter values $K = 500$, $b = d = 0.01$. These figures illustrate the increasing magnitude of stochastic effects at low population levels and migration rates.

We note that in all of these simulations, the overall pattern is that all populations reach equilibrium. This is reasonable because our birth and death processes mimic the logistic growth function, which has an unstable extinct state. At low population, the migration out of a patch is negligible, so the probability of a series of chance events driving a new population experiencing logistic growth to extinction is incredibly small when there is a source feeding into it. As a result the stochastic model more or less traces the deterministic model with small perturbations.

However, these figures show that the fit between the deterministic and stochastic model breaks down over certain time and parameter regimes. Specifically the regime where stochastic effects are most apparent is in the initial lag period before logistic growth is high, when population is low and population dynamics are primarily controlled by migration. While in the deterministic model, the populations grow at the same rate, this

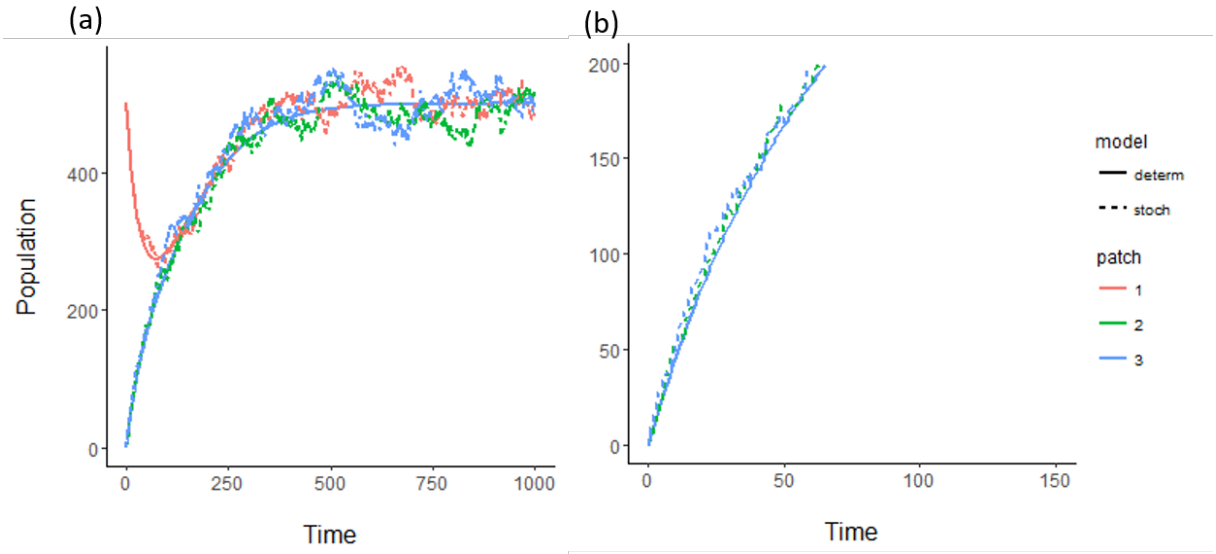

Figure 4.1: Migration rate $v = 0.01$. Plot (a) shows the entire 1000-year trajectory of each population patch while plot (b) shows a shorter time frame where stochastic effects are most apparent.

is not true of the stochastic model. Due to low migration rates and birth rates at low population, if one patch happens to receive an introduction early one, its population will grow much more rapidly than the others, as seen in 4.4.

Previously we discussed Allee effects as more biologically realistic logistic growth models. The subject of Allee effects in stochastic simulations is discussed as a possible topic of future work.

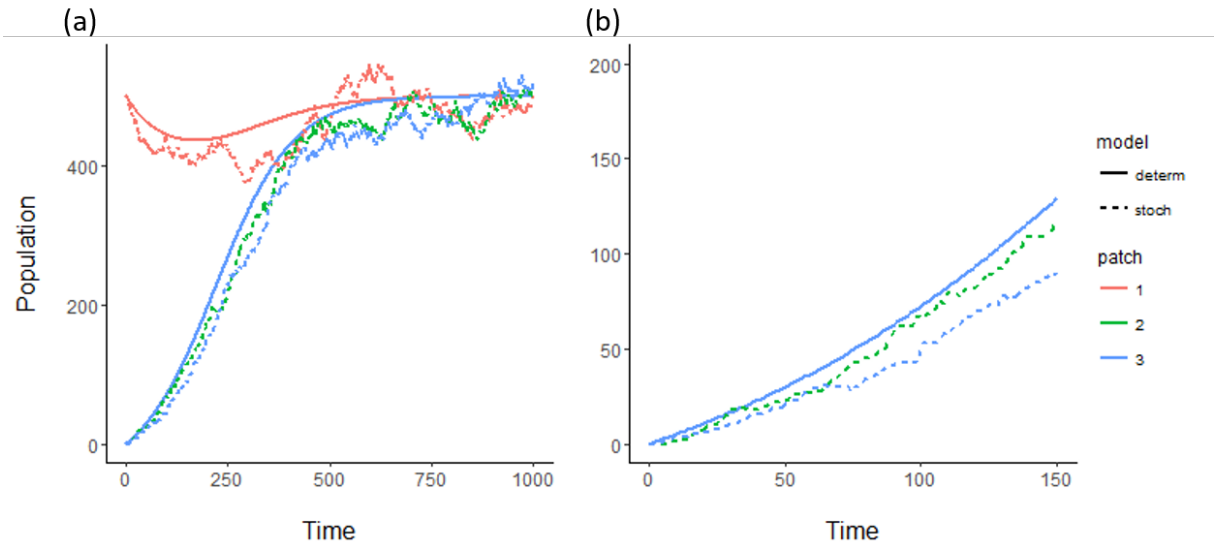

Figure 4.2: Migration rate $v = 0.001$. Otherwise parameters are unchanged from figure 4.1

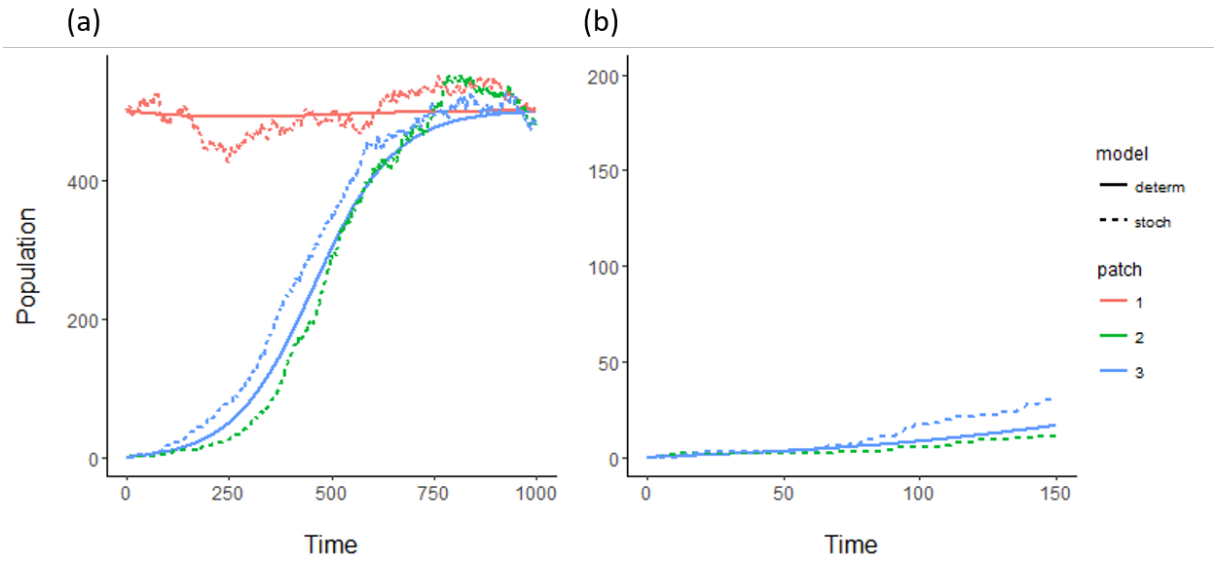

Figure 4.3: Migration rate $v = 0.0001$. Otherwise parameters are unchanged from figure 4.1.

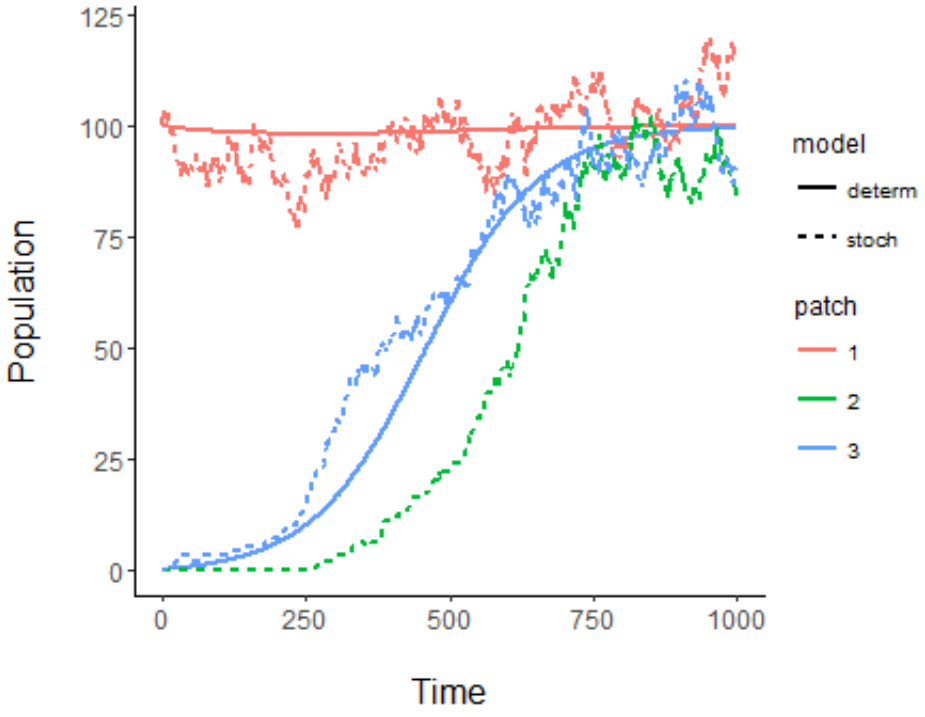

Figure 4.4: Migration rate $v = 0.0001$, $K = 100$. Otherwise parameters are unchanged from figure 4.1. This plot shows that patch dynamics can become highly varied when the population sizes are low.

Chapter 5

Conclusions

5.1 Summary

We developed our n dimensional deterministic network model. We found the fixed points and stability of logistic population growth and presented a portion of the analysis of the 2 dimensional system. We assume that for higher dimensions the analysis can be extended, so that the extinct state is unstable, and there exists a stable positive fixed point in the vicinity of the carrying capacity.

We implemented deterministic and stochastic versions of this model in R and found that the results from the numerical simulations generally followed a growth pattern expected from our analysis of logistic growth steady states. We identified regimes where stochastic effects were most relevant, however, we found that long-term behavior of the deterministic and stochastic models agreed. Allee effects would be a good direction to take future work.

We introduced small world networks and used them to study the effects of random dispersal events on the time for an invasive species to spread across a network. We also related invasive species spread on a network to two network properties - characteristic path length and clustering coefficient. Further work is investigation into these and other

network properties and how they influence spread dynamics is warranted.

5.2 Future Work

5.2.1 Allee Effect

We previously introduced the Allee effect as an extension of the logistic growth equation. The resulting ODE incorporating an Allee effect into logistic growth is defined in equation 1.2. We noted that there is a difference in steady states between this model and the original ODE in 1.1. In the Allee model, there are three fixed points: 0, A (the Allee threshold), and K. Recall that $0 < A < K$. The extinct state, 0 is stable, the Allee threshold is unstable, and the carrying capacity is stable.

The difference in steady states and stability underly a significant difference between our current stochastic model and a biologically realistic one. In real landscapes introduced populations do not necessarily tend towards carrying capacity; instead, many small populations fail to establish. This effect is described by the change in steady state stability of a logistic growth model under an Allee effect. A future direction for research would be the implementation and analysis of a growth function incorporating an Allee effect. Updating the continuous time Monte Carlo simulation using birth and death rates associated with that model would likely yield interesting results when tested under low migration rate and low population.

5.2.2 Networks

In the small world section we noted that the normalized time to establishment metric we tested followed the same functional form as the normalized mean clustering coefficient. It might be fruitful to further investigate how clustering affects spread dynamics and also what other network properties might be important in the same respect.

Besides the effect of long-distance connections, another topic of interest in many networks are hubs. Hubs, defined non-technically, are nodes which experience a large volume of of incoming and outgoing traffic relative to other nodes. Hubs are highly connected and have a large influence on network dynamics. A hub in the context of invasive species might be in a population center. For example, in Kaluza et al. (2010), researchers found that highly trafficked ports were on coastlines that had high numbers of marine invasive species. Hubs as hotspots for the spread of invasive species could be a possible direction for future research. This would be particularly useful for environmental management efforts, which could use this type of knowledge to best allocate limited resources to have the greatest effect.

5.3 Acknowledgements

Thank you to my incredible advisor Dr. Leah Shaw, without whom I could never have taken this project this far. Thank you for not only for demystifying mathematical analysis and always being ready to teach me something new, but for guiding me and pushing me forward through this experience and all of its ups and downs.

Bibliography

James Carlton. *Invasive species: vectors and management strategies*. Island Press, 2003.

Invasive Species Advisory Committee et al. Invasive species definition clarification and guidance white paper. national invasive species council. *US Department of Agriculture, National Agricultural Library. Washington, DC Available at http://www. invasivespeciesinfo. gov/docs/council/isacdef. pdf*, 2006.

Jane Elith and John R Leathwick. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1):677, 2009.

O Floerl, GJ Inglis, K Dey, and A Smith. The importance of transport hubs in stepping-stone invasions. *Journal of Applied Ecology*, 46(1):37–45, 2009.

Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

Philip E Hulme. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1):10–18, 2009.

Pablo Kaluza, Andrea Kölzsch, Michael T Gastner, and Bernd Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48): 1093–1103, 2010.

Kirill S Korolev, Joao B Xavier, and Jeff Gore. Turning ecology and evolution against cancer. *Nature Reviews Cancer*, 14(5):371–380, 2014.

Vito Latora and Massimo Marchiori. Efficient behavior of small-world networks. *Physical review letters*, 87(19):198701, 2001.

Richard N Mack, Daniel Simberloff, W Mark Lonsdale, Harry Evans, Michael Clout, and Fakhri A Bazzaz. Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological applications*, 10(3):689–710, 2000.

Michael G Neubert and Ingrid M Parker. Projecting rates of spread for invasive species. *Risk Analysis*, 24(4):817–831, 2004.

David Pimentel, Rodolfo Zuniga, and Doug Morrison. Update on the environmental and economic costs associated with alien-invasive species in the united states. *Ecological economics*, 52(3):273–288, 2005.

Stuart A Reeves and Michael B Usher. Application of a diffusion model to the spread of an invasive species: The coypu in great britain. *Ecological modelling*, 47(3):217–232, 1989.

Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Westview press, 2014.

Tomáš Václavík and Ross K Meentemeyer. Invasive species distribution modeling (isdm): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling*, 220(23):3248–3258, 2009.

Duncan J Watts and Steven H Strogatz. Collective dynamics of small-worldnetworks. *nature*, 393(6684):440–442, 1998.

Mark Williamson. Mathematical models of invasion. 1989.