

# 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](mailto: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

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background on Invasive Species . . . . .	1
1.2	Modeling Background . . . . .	2
1.2.1	Thesis Outline . . . . .	7
<b>2</b>	<b>Deterministic Modeling</b>	<b>9</b>
2.1	1-Dimensional Model Formulation . . . . .	9
2.1.1	Rationale . . . . .	9
2.1.2	Model Specification . . . . .	10
2.1.3	Alternative Logistic Models . . . . .	10
2.2	1-Dimensional Stability Analysis . . . . .	11
2.2.1	Fixed points . . . . .	11
2.2.2	Stability of Fixed Points . . . . .	11
2.2.3	Interpretation . . . . .	12
2.3	Model Formulation in Higher Dimensions . . . . .	13
2.4	Analysis of 2-Dimensional System . . . . .	15
2.4.1	Stability Analysis . . . . .	16
2.5	Numerical Methods . . . . .	17
2.5.1	Discrete Map . . . . .	18
2.5.2	ODE Model . . . . .	18

2.5.3	Deterministic Model Agreement . . . . .	19
<b>3</b>	<b>Small World Networks</b>	<b>23</b>
3.1	Background . . . . .	23
3.2	Small World Methods . . . . .	24
3.2.1	Extension to Invasion Model . . . . .	24
3.3	Results . . . . .	27
<b>4</b>	<b>Stochastic Modeling</b>	<b>33</b>
4.1	Stochastic Model . . . . .	33
4.1.1	Results . . . . .	35
<b>5</b>	<b>Conclusions</b>	<b>39</b>
5.1	Summary . . . . .	39
5.2	Future Work . . . . .	41
5.2.1	Allee Effect . . . . .	41
5.2.2	Networks . . . . .	41
5.3	Acknowledgements . . . . .	42

# List of Figures

1.1	(a) Flow diagram showing steady states and stability of Equation 1.1. (b) Flow diagram showing steady states and stability of Equation 1.2. . . . .	4
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 transferred from patch $j$ to $i$ in a single year. . . . .	14
2.2	Figure showing a sample network with edges equally weighted. The parameter $v$ is the migration rate between nodes and $k_j$ is the outdegree of the node $j$ . . . . .	19
2.3	Figure showing the agreement between the ODE model and the discrete map at typical parameter values. The solid lines are the abundances 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 $v = 0.001$ , birth rate $r = 0.01$ , and carrying capacity $K = 500$ , over a course of 1000 time steps. . . . .	21
2.4	Figure showing discrepancy between the ODE model and the discrete map at large $v$ values. The parameter values given are listed: $n = 2$ patches, migration rate $v = 0.25$ , birth rate $r = 0.01$ , and carrying capacity $K = 100$ , over a course of 10 time steps. . . . .	22

3.1	A regular ring lattice is rewired with increasing values of $p$ , the rewiring probability. . . . .	25
3.2	Plots showing relationship between normalized 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 $p$ . The plotted values are normalized by their values calculated on a regular graph. Parameter values are $n = 1000$ , $m = 5$ , the results are averaged over 20 realizations. . . . .	26
3.3	Example of the smallworld network generated by igraph with $n = 50$ , $m = 2$ , and $p = 0.25$ . . . . .	28
3.4	Plot of the normalized time in years to establishment of a species invading a small world network with varying values of $v$ . Parameters $n = 50$ , $m = 2$ , $r = 0.01$ , and establishment threshold is set at 100. . . . .	29
3.5	Plot of the normalized time in years to establishment of a species invading a small world network with varying values of $r$ . Parameters $n = 50$ , $m = 2$ , $v = 0.01$ , and establishment threshold is set at 100. . . . .	30
3.6	Plot of the normalized time in years to establishment of a species invading a small world network with varying sets of migration rates $v$ and birth rate $r$ . Parameters $n = 50$ , $m = 2$ , $r = 0.01$ , and establishment threshold is set at 100. . . . .	31
3.7	Plot of the absolute time in years to establishment of a species invading a small world network with varying sets of migration rates $v$ and birth rate $r$ . Parameters $n = 50$ , $m = 2$ , $r = 0.01$ , and establishment threshold is set at 100. . . . .	32
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. Other parameters $K = 500$ , $r = 0.01$ . . . . .	36

4.2	Migration rate $v = 0.001$ , other parameters $K = 500$ , $r = 0.01$ . . . . .	36
4.3	Migration rate $v = 0.0001$ , other parameters $K = 500$ , $r = 0.01$ . . . . .	37
4.4	Migration rate $v = 0.0001$ , other parameters $K = 100$ , $r = 0.01$ . This plot shows that patch dynamics can become highly varied when the population sizes are low. . . . .	38





# 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 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 is an incredible number 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. Throughout this thesis, we will use the terms “node” and “patch”. Generally, a node in our network model is meant to represent a discrete population, or “patch” in the context of biology, and so these terms may be used interchangeably, **with the context implied.**

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:

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

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

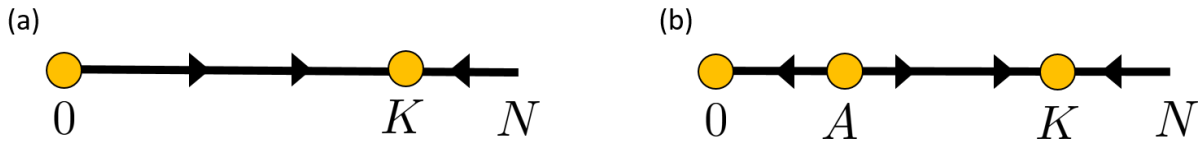
One criticism of this model is that lacks the inclusion of an Allee effect, characterized as a positive association between population density and individual fitness. As explained in Korolev et al. (2014), a strong Allee effect is a phenomenon that leads a population to extinction once it dips below 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, becoming negative at low populations below the threshold. Allee effects can be caused by many different mechanisms that are positively density dependent, such as pack hunting behavior or mate-finding. The logistic equation can be modified to include an Allee effect as follows:

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

where  $A$  is a positive constant that defines the threshold below which extinction is ensured.

Steady states and stability for Equations 1.1 and 1.2 are shown in Figure 1.1. In the population undergoing logistic growth without an Allee effect (Figure 1.1a), the extinct state (zero population) is unstable and grows towards carrying capacity with a small positive perturbation. However, under a strong Allee effect (Figure 1.1b), another steady state at the Allee threshold  $A$  is introduced. The extinct state becomes stable, with

$A$  being unstable and the carrying capacity remaining a stable point. As discussed in Section 1.1, the introduction of a small number of individuals of an invasive species to a new location will often fail to establish a stable colony. The steady state properties of the logistic equation predict that all introductions will establish stable colonies. We know that in nature, this is not true. A likely reason for this is that many species experience population growth with positive density dependence at low populations as modeled by the logistic equation with a strong Allee effect.



**Figure 1.1:** (a) Flow diagram showing steady states and stability of Equation 1.1. (b) Flow diagram showing steady states and stability of Equation 1.2.

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 model spatial dispersal is based on reaction-diffusion equations. An early example of this is a form of Fisher's equation, taken from Neubert and Parker (2004):

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

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. Models with simple diffusion 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 Coyote 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 were developed to incorporate various aspects left out by the relatively basic reaction-diffusion equations. Integro-difference 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

$$N(y, t + 1) = f[N(y, t)] \quad (1.4)$$

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  $k(x, y)$ , the probability of redistributing from point  $y$  to point  $x$ . The resulting distribution of population at time  $t + 1$  is then described by the IDE:


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

The main advantage of IDEs are their flexibility in choosing the density kernel, which allows for more complex redistribution of population. A similar type of model, integrodifferential equations, are used in Sharov and Liebhold (1998) to model the spread of Gypsy moths in the United States. Integrodifferential models share a similar form to integrodifference models, with the difference being that the former model continuous time processes whereas the latter is a discrete time model.

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 explain present species distributions. However, in the case of invasive species we wish to use the model for extrapolation. One problem with SDMs is that an underlying assumption 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. This assumption is likely false for invasive species, since they face clear dispersal limitations being newly introduced to their environment. Despite these concerns, SDMs are increasingly being used more for extrapolations 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. Floerl et al. (2009) investigated transport hubs as centers for species dispersal. Using data from marinas near New Zealand, they found that locations categorized as high traffic and connectivity “hubs” were much more likely to be infected early and be the source of spread to many secondary locations than less visited locations.

There have also been recent papers focused on building predictive models of terrestrial invasive species. For example, in Ferrari et al. (2014), researchers presented a predictive model of hemlock woolly adelgid spread, using a dynamic network model to explore which nodes were most active in dispersal. In Koch et al. (2014), researchers investigated the spread of forest pests in the United States and Canada via firewood movement, commonly held as a vector for forest pest introduction.  Despite the acknowledged importance of

these studies, much of the presently available research has been focused on case studies of particular species and their specific distribution networks. ~~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. The goal of our research is to present a general network model of terrestrial invasive species spread and analyze the individual components of that model. Over time, data driven, computationally-intensive models may become more attractive due to increases in data availability and computational power. This will lead 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 in addition to the introduction and conclusion. In Chapter 2 we will derive the deterministic model and analyze its steady state properties in 1 and 2 dimensions. We also define the  $n$  dimensional model we use for computational simulations. In Chapter 3, we introduce the characteristics of a particular type of network, small world networks, and their relation to biotic invasion. Lastly, in Chapter 4, we introduce a stochastic model and explore the role of stochasticity in different parameter regimes in the network 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:

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

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:

$$C = \frac{N_0}{1 - \frac{N_0}{K}} \quad (2.2)$$

Replacing  $C$  in Equation 2.1 yields the particular solution

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

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:

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



### 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:

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

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 the formulation in Equation 2.4 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

What are the steady state properties of the model discrete logistic map we defined in Equation 2.4? We discussed the steady state of the ordinary differential equation form of the logistic equation, Equation 1.1, in Chapter 1. We expect the stability results here to reflect those of the logistic equation since we derived the map from the logistic equation. We follow the approach described in Strogatz (2014) to analyze the steady state stability of our discrete map system from Equation 2.4. 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 Equation 2.4. Solving the equation below for  $N^*$ ,

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

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  $\epsilon_t$ . Suppose the population at  $N_t$  is at the fixed point  $N^*$ , then

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

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:

$$\begin{aligned} N_{t+1} &= g(N^* + \epsilon_t) \\ N_{t+1} &\approx g(N^*) + g'(N^*)\epsilon_t + O(\epsilon_t^2) \\ N_{t+1} &\approx N^* + g'(N^*)\epsilon_t \end{aligned}$$



Because  $\epsilon_t$  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~~  $\epsilon_t$  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:

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

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  $r < 0$ . These are the results that we expected from knowing the steady state stability of the continuous time logistic equation.



### 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