University of Alberta

Library Release Form

Name of Author: Christopher Lee Jerde

Title of Thesis: The arrival and establishment of non-indigenous species: mechanisms, uncertainty, and prediction

Degree: Doctor of Philosophy

Year this Degree Granted: 2008

Christopher Lee Jerde
CW 304, Biological Sciences Centre
Center for Mathematical Biology
University of Alberta
Edmonton, Alberta
Canada, T6G 2E9

Date: ____________________

There are so few left, so few that believe the earth is enough.

– Harry Middleton, *The Earth is Enough*.

University of Alberta

THE ARRIVAL AND ESTABLISHMENT OF NON-INDIGENOUS SPECIES: MECHANISMS, UNCERTAINTY, AND PREDICTION

by

Christopher Lee Jerde

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

in

Ecology

Department of Biological Sciences

Edmonton, Alberta
Spring 2008

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **The arrival and establishment of non-indigenous species: mechanisms, uncertainty, and prediction** submitted by Christopher Lee Jerde in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in *Ecology*.

____

Mark A. Lewis

____

Carol C. Horvitz

____

Subhash R. Lele

____

Mark R. T. Dale

____

Rolf D. Vinebrooke

Date: ______________

To Mom for showing me how to teach,
to Dad for teaching me how to understand,
to Royal for taking me fishing,
and
to my Grandfather, Gale Witham, for seed–corn hats, pocketknives, and dirt under my fingernails.

Abstract

Non-indigenous species (NIS) have reduced native species abundances, increased native species extinction probabilities, and changed ecosystem functioning. In addition to the impact on the environment, economic losses are accrued to the public and industry. The inability to make successful predictions of which species will invade, where, and when, hinders invasive species management. My thesis objective was to develop a risk assessment framework for predicting the arrival and establishment of NIS. I used hierarchical probability models that capture the NIS arrival process, and I evaluated invasion risk using stochastic processes to produce invasion waiting times. This process was tested using experimentally manipulated propagule pressure of scentless chamomile (*Matricaria perforata*), which subsequently validated an invasion waiting time risk assessment approach. The approach was extended using relative measures of ballast water discharge of Chinese mitten crabs (*Eriocheir sinensis*) into North American ports, and the estimated transport of zebra mussels (*Dreissena polymorpha*) by recreational boaters in the United States. With these examples, I showed that relative measures require making mathematical and biological assumptions, which when violated, result in poor predictions. I considered the influence of the Allee effect on the invasion waiting time by formulating a hierarchical probability model of NIS establishment for semelparous, sexual species. The Allee effect is detected in the net geometric per capita growth rate, and I evaluated persistence with a stochastic process. With large fecundity, the influence of the Allee effect is negligible and invasion risk can be evaluated using the invasion waiting time based on the probability of one surviving, fertilized female persisting post arrival. This approach was applied to Chinese mitten crab and apple snail (*Pomacea canaliculata*).

Acknowledgements

I am grateful to Mark Lewis for providing support, encouragement, enthusiasm, and critical feedback throughout the dissertation process. Maybe the greatest contribution to my development as a scientist has been the personal interactions arising from my involvement with other scientists at the Centre for Mathematical Biology (CMB). Mark's direction of the CMB provided this environment. Members of the CMB that helped shape my thinking and development include: G. de Vries, T. Hillen, A. Potapov, F. Lutscher. E. Noonburg, L. Pachepsky, C. Cobbold, A. Pielaat, T. Robbins, M. Wonham, C. Bampfylde, T. de Camino Beck, J. Lee, W. Nelson, F. Hilker, A. Dawson, R. Eftimie, M. Krkosek, H. McKenzie, P. Molnar ,A. Hurford, J. Muirhead, H. Rajakaruna and J. Marleau. Additionally, C. Hutchinson, R. Bechtel, K. Schreiner, C. Zinchuk, C. Williams and M. Duller provided crisis management for administrative and computing concerns.

T. de Camino Beck, along with helping in the field, was a confidant throughout. M. Wonham inspired me often. C. Bampfylde tirelessly provided guidance, feedback, and support. In the absence of these three individuals, I am sure my work (and life) would be left wanting.

S. Lele, a committee member and friend, taught me to simplify problems, think clearly, struggle with scientific evidence, and go to the board. M. Taper, took a chance on me as a scientist, and he started me on this path when I was looking for a direction. In this group also belongs W. Nelson, whom over coffee, made me see bigger pictures of science – and the fine details.

I discovered the joy in working on science, not only with a good scientist, but also with a good friend. Thank you DR. Visscher.

I have experienced beneficial collaborations and discussions with many people

and groups during my tenure. In closing I would like to acknowledge the contributions to my work from H. MacIsaac Lab, D. Lodge Lab, J. Bossenbroek, J. Drake, L-M. Herborg, E. Merrill Lab, The ISIS work group, the Mixed Models Graduate Student work group, the support staff of the University of Alberta Department of Biological Sciences, and my thesis committee: Dr. M. Lewis, Dr. S. Lele, Dr. M.R.T. Dale, Dr. R. Vinebrooke, and Dr. C. Horvitz. Thank you all.

Funds and resources for this dissertation were provided by The University of Alberta, The Department of Biological Sciences, The Department of Mathematical and Statistical Sciences, The Alberta Research Council, The Pacific Institute of Mathematical Science, The Natural Sciences and Engineering Research Council of Canada's Collaborative Research Opportunities programme, The Canadian Aquatic Invasive Species Network, The Ecological Society of America, and The U.S. National Science Foundation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Why do we care about invasive species?

The impact of invasive species in ecological processes has been the subject of ongoing discussion for some time (Elton, 1958; Davis et al., 2001). However, ecological curiosity is transitioning to environmental concern because biological invasions are second only to habitat degradation as a threat to biodiversity (Glowka et al., 1994; Williamson, 1996). Many justifications have been put forth for studying invasion biology. There are the notable ecological impacts, such as changes in ecosystem functioning (Chapin et al., 2000) and the resulting loss of earth's biodiversity (Clavero and Garcia-Berthou, 2005). Others have demonstrated how ecological theory is advanced by uncovering the role Allee effects play in the population dynamics of invasive species (Taylor and Hastings, 2005), or how species richness changes when new species are added to a system (Tilman, 2004). In these instances, invasive species are an unfortunate and unintended experiment with a benefit for testing theoretical insights (Shea and Chesson, 2002). More pragmatically, invasive species are of economic concern (Pimentel et al., 2000, 2005). These justifications are not independent. For example, evaluating the economic benefit gained by preventing or eradicating a species largely depends on the population dynamics and the risk of invasion (Leung et al., 2002; Simberloff, 2003). Despite these justifications, the general properties and conditions that lead to successful invasion remain largely elusive (Carlton, 1996), but with some tactical successes (see, for example Williamson, 1996; Bossenbroek et al., 2007).

1.2 What is an invasive species?

Exotic, alien, invasive, pest, weed, nuisance – there are many names for non-indigenous species (NIS; Colautti and MacIsaac, 2004), and the terminology used in reference to NIS is inconsistent (Richardson et al., 2000). One definition of NIS is the "condition of a species being moved beyond its natural range or natural zone of potential dispersal including all domesticated and feral species and hybrids" (Non-indigenous Aquatic Nuisance Prevention and Control Act of 1990, Public Law 101-646, 16 USC 4701-4741). However this legal definition suffers from linguistic uncertainty (Regan et al., 2002) because the natural range or zone implies some fixed state of the world's flora and fauna to which non-indigenous can be referenced to, but fixed states are rarely defined and inconsistent definitions of non-indigenous species persists (Shrader-Frechette, 2001).

Alternatively, non-indigenous species are sometimes defined solely by the characteristic of anthropogenic transport, without any reference to a natural range or zone (Colautti and MacIsaac, 2004). However, there exists a problem for classifying invasions after a NIS is introduced to a location, establishes, and is able to propagate. Are the offspring of the NIS also NIS if they disperse by means unrelated to human movement and transportation? I would argue that offspring of NIS are also NIS – irrespective of the method dispersal. One way to overcome this nuance is to allow the classification of a NIS to change through the invasion process (Colautti and MacIsaac, 2004). This would imply that a species' dispersal and establishment history defines its status as a NIS, and thus I must return to the initial human transport aspect as the defining feature of NIS. But where do I place the proverbial line in the sand for human transport of species so they may be classified as NIS? Therefore, I return the problem of having a reference state that classifies where species are indigenous to.

Recorded natural history and the history of human transportation will necessarily form the basis for discriminating indigenous and non-indigenous species (at least for the Americas and the species I will consider in my thesis). Without written accounts, we have no reference to what species persisted prior to the regular arrival and settlement of the Americas by European explorers (*circa* 1500 A.D.). Therefore a non-indigenous species is operationally defined for my thesis as a human transported species that arrived after 1500 A.D., and includes the offspring of these human transported species that continue to persist. The distinction between NIS and invasive species is primarily a social construct where invaders are a subset of NIS that have some negative ecological impact (Parker et al., 1999; Colautti and MacIsaac, 2004).

I use four species to discuss the arrival and establishment of NIS: scentless chamomile (*Matricaria perforata*), Chinese mitten crabs (*Eriocheir sinensis*), apple snails (*Pomacea canaliculata*), and zebra mussels (*Dreissena polymorpha*). All of these species have first record sightings in North America after 1500 A.D., and there is evidence that these species arrived by human mitigated pathways of introduction. The details of these arrivals are presented with the examples.

1.3 What is the invasion process?

The invasion process is often separated into arrival, establishment, and spread stages (Vermeij, 1996; Sakai et al., 2001). The arrival stage is characterized by the uptake, transport, and delivery of NIS into new locations (Carlton, 1996; Floerl and Inglis, 2005). Also referred to as initial dispersal (Puth and Post, 2005), research on this stage has emphasized predicting potential NIS (Kolar and Lodge, 2001) through identifying characteristics of successful invaders (Goodwin et al., 1999) and by assessing the rate at which NIS are introduced into new locations (propagule pressure: Schneider et al., 1998; Lockwood et al., 2005).

In contrast, the establishment stage of the invasion process focuses on the population dynamics of NIS at low population abundances (Sakai et al., 2001; Drake et al., 2005; Drake and Jerde, In Press). This includes consideration of the role of demographic stochasticity (Lande, 1993; Drake and Lodge, 2006), biotic resistance (Elton, 1958), fluctuating environmental conditions (Davis et al., 2000; Tilman, 2004), and Allee effects (Allee, 1938; Dennis, 1989; Leung et al., 2004) on the probability of establishment and the probability of going extinct. The transition between establishment and spread can occur almost immediately after the species arrives (e.g. Schneider et al., 1998), but the transition can also experience a lag where species are present in the system at low abundances only later to rapidly increase in abundance and begin to spread within the invaded landscape (e.g. Wangen and Webster, 2006).

Theoretical considerations and applied investigations of the spread of invasive species have received proportionally more attention than both the arrival and establishment stages (Puth and Post, 2005). The spread stage is characterized by population dynamics at larger abundances that fluctuate due to environmental stochasticity, and the spread of populations into continuous landscapes and discrete patches (Shigesada and Kawasaki, 1997). Similar to establishment phase, the spread of invasive species may also consider environmental variability (Moyle and Light, 1996) and dynamics such as Allee effects (Lewis and Kareiva, 1993).

The “tens rule” for biological invasions (Williamson, 1996) estimates that from a pool of potential invaders only ten percent of arriving species will establish, of the establishing species only ten percent will spread, and of spreading species only ten percent become pests by having an ecological (Parker et al., 1999) and/or economic (Pimentel et al., 2005) impact. This generalized pattern of the invasion process was proposed as a probabilistic ‘rule of thumb’ to predicting invasions in response to the criticism of Gilpin (1990) who said, “we are never going to have a scheme to predict the success of invading species....” Since this exchange, much of the attention in invasion biology has been on developing theoretical frameworks and methods for predicting invasions (Kolar and Lodge, 2001), in part because arrival represents the most likely stage where active management can reduce the risk of invasions (Sakai et al., 2001; Wonham et al., 2005).

1.4 Thesis overview

Chapter 2: A working framework for arrival and establishment In this chapter, I formulate a mechanistically based, hierarchical model of NIS arrival and demonstrate simplifications leading to a marginal distribution of the number of surviving, introduced individuals from parameters of survival probability and propagule pressure. The marginal distribution is extended as a stochastic process from which establishment emerges with a waiting time distribution. This provides a probability of NIS establishment within a specified period and may be useful for identifying patterns of successful invaders (Gilpin, 1990). However, estimates of both the propagule pressure and individual survival probability are rarely available for NIS making estimates of the probability of establishment difficult. Alternatively, researchers are able to estimate proportional values of propagule pressure through models of NIS transport, such as gravity models, or survival probability through habitat matching indexes measuring the similarity between potentially occupied and native NIS ranges. Therefore, I formulate the relative waiting time between two locations and the probability of one location being invaded before the other.

Chapter 3: Experimentation and patterns of invasion Preventing invasions requires predicting which locations are most likely to become invaded, and when. In Chapter 2, I proposed the invasion waiting time method for evaluating the risk of invasion using parameters of the propagule pressure and survival. Because the establishment process uses count data, it is subject to greater than expected variation in the number of establishing individuals. Failure to account for this overdispersion leads to overestimates in the probability of at least one individual establishing and increases the right skew of the resulting waiting time distribution. I experimentally manipulated the arrival of an established invader, *M. perforata* in two locations and produced spatially replicated observations of establishment. After accounting for significant levels of overdispersion, the locations have probabilities of at least one individual establishing equal to 0.55 and 0.37. Using the spatial replication, I validated the waiting time distributions for these two locations as being geometrically distributed and further support invasion waiting times as a useful construct for estimating invasion risk, predicting invasions, and uncovering the patterns of successful invasions.

Chapter 4: Making connections through uncertainty When there is variability in predictions or when mechanisms of how systems or processes work are unknown, there is uncertainty. From medicine to engineering, and biology to economics, uncertainty is an important consideration when testing hypotheses and predicting outcomes. In this chapter, I explore the role of uncertainty on explaining and predicting invasions from a biological perspective. I begin by reviewing a classification of uncertainty that conforms to how many biologists and statisticians perceive the role of uncertainty in their scientific explorations. I then narrow my focus by evaluating the uncertainty in explanations and predictions from a gravity model of zebra mussel invasion into the Western United States. This includes providing bootstrapped confidence intervals on parameter estimates from survey data and evaluating the predictive performance of the gravity model on a subset of economically and ecologically valuable lakes using probability theory and receiver operator characteristic curves. Many of these evaluations of uncertainty are uncommon in current gravity model applications to invasive species and are generalizable to other modeling approaches and the larger concern of predicting successful invasions.

Chapter 5: Establishment, Allee effects, and demographic stochasticity Predicting which species will invade, which locations will become invaded, and when invasion will occur is a lofty goal for invasion biology. But management, seeking to intervene in the invasion process, and bio-economic assessments, used to justify management actions, need probabilities of these events occurring in order to be effective. Here, I have formalized the establishment process for a sexual, semelparous organism using hierarchical probability modeling from life history parameters of survival, probability of being female, the probability of a female being fertilized by a male, and the expected fecundity. From this probability model, I have shown how to calculate the expected net per capita growth rate, the expected net population change, and probability of extinction, as a function of the initial population size. An Allee effect is observed if either the expected net population growth rate or net population change decrease as the initial population size decreases. A weak Allee effect is observed when expected growth is decreasing but remains positive, and a strong Allee effect is observed when the expected growth rate decreases and becomes negative. The model can be further extended as a stochastic process to evaluate the probability of extinction in subsequent generations. Two semelparous, sexual organisms appear on the ICUN's list of 100 world's worst invaders, the Chinese mitten crab (*E. sinensis*) and the apple snail (*P. canaliculata*). Using point estimates from the literature, I evaluate the strength of the Allee effect and conclude that apple snails experience a weak Allee effect and the Chinese mitten crab experiences a strong Allee effect. The stochastic process reveals that invasion risk can be estimated by the probability of one fertilized female surviving, because the expected fecundity for one surviving female overwhelms the system such that population persistence is almost certain. As a consequence of this dynamic, Chinese mitten crabs or apple snails, under repeated introductions, fit into an invasion waiting time framework.

Chapter 6: Synthesis In my closing chapter, I take a broader, critical, and sometimes speculative, approach to the ideas and work presented in the previous four chapters. I demonstrate how my dissertation merges or complements existing approaches of NIS risk assessment, distinguish where my dissertation advances NIS risk assessment, and identify where future invasive species research needs to proceed. I structure this discussion by dissecting the arrival and establishment stages of the invasion process (Sakai et al., 2001), critiquing the application of risk assessment for NIS (Suter, 1993), and connecting my work to applications in risk management and conservation biology (D'Antonio et al., 2001; Lodge et al., 2006).

Bibliography

Allee, W. 1938. The social life of animals. W.W. Norton and Co., New York, NY, USA.

Bossenbroek, J., J. Ladd, B. Peter, and D. Lodge. 2007. Forecasting the expansion of Zebra Mussels in the United States. Conservation Biology **21**:800–810.

Carlton, J. 1996. Pattern, process, and prediction in marine invasion ecology. Biological Conservation **78**:97–106.

Chapin, F., E. Zavaleta, V. Eviner, N. Rosamond, P. Vitousek, H. Reynolds, D. Hooper, S. Lavorel, O. Sala, S. Hobbie, M. Mack, and S. Diaz. 2000. Consequences of changing biodiversity. Nature **405**:234–242.

Clavero, M., and E. Garcia-Berthou. 2005. Invasive species are a leading cause of animal extinctions. Trends in Ecology and Evolution **20**:110.

Colautti, R., and H. MacIsaac. 2004. A neutral terminology to define invasive species. Diversity and Distributions **10**:135–141.

D'Antonio, C., L. Meyerson, and J. Denslow, 2001. Conservation Biology: research priorities for the next decade, Chapter exotic species and conservation: Research needs, pages 59–80 . Island Press, Washington, D.C., USA.

Davis, M., J. Grime, and K. Thompson. 2000. Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology **88**:528–534.

Davis, M., K. Thompson, and J. Grime. 2001. Charles S. Elton and the dissociation of invasion ecology from the rest of ecology. Diversity and Distributions **7**:97–102.

Dennis, B. 1989. Allee effects: population growth, critical density, and the chance of extinction. Natural Resource Modeling **3**:481–538.

Drake, J., P. Baggenstos, and D. Lodge. 2005. Propagule pressure and persistence in experimental populations. Biology Letters **1**:480–483.

Drake, J., and C. Jerde, In Press. Bioeconomics of Invasive Species: Integrating Ecology, Economics, Policy, and Management, Chapter stochastic models of propagule pressure and establishment . Oxford University Press.

Drake, J., and D. Lodge. 2006. Allee effects, propagule pressure and the probability of establishment: risk analysis for biological invastions. Biological Invasions **8**:365–375.

Elton, C. 1958. The ecology of invasion by animals and plants. Methuen, London, UK.

Floerl, O., and G. Inglis. 2005. Starting the invasion pathway: the interaction between source populations and human transport vectors. Biological Invasions **7**:589–606.

Gilpin, M. 1990. Review: ecological prediction. Science **248**:88–89.

Glowka, L., F. Burhenne-Guilmin, and H. Synge, 1994. A guide to the conservation of biological diversity. Technical report, ICUN, Gland.

Goodwin, B., A. McAllister, and L. Fahrig. 1999. Predicting invasiveness of plant species based on biological information. Conservation Biology **13**:422–426.

Kolar, C., and D. Lodge. 2001. Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution **16**:199–204.

Lande, R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. American Naturalist **142**:911–927.

Leung, B., J. Drake, and D. Lodge. 2004. Predicting invasions: propagule pressure and the gravity of Allee effects. Ecology **85**:1651–1660.

Leung, B., D. Lodge, D. Finnoff, J. Shogren, M. Lewis, and G. Lamberti. 2002. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proceedings of the Royal Society of London, B **269**:2407–2413.

Lewis, M., and P. Kareiva. 1993. Allee dynamics and the spread of invading organisms. Theoretical Population Biology **43**:141–158.

Lockwood, J., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. Trends in Ecology and Evolution **20**:223–228.

Lodge, D., S. Williams, H. MacIsaac, K. Hayes, B. Leung, S. Reichard, R. Mack, P. Moyle, M. Smith, D. Andow, J. Carlton, and M. McMichael. 2006. Biological invasions: recommendations for U.S. policy and management. Ecological Applications **16**:20352054.

Moyle, P., and T. Light. 1996. Fish invasions in California: Do abiotic factors determine success? Ecology **77**:1666–1670.

Parker, I., D. Simberloff, W. Lonsdale, K. Goodell, M. Wonham, P. Kareiva, M. Williamson, B. Von Holle, P. Moyle, J. Byers, and L. Goldwasser. 1999. Impact: Toward a framework for understanding the ecological effects of invaders. Biological Invasions **1**:3–19.

Pimentel, D., L. Lach, R. Zunign, and D. Morrison. 2000. Environmental and economic costs of non-indiginous species in the United States. BioScience **50**:53–67.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien–invasive species in the United States. Ecological Economics **52**:273–288.

Puth, L., and D. Post. 2005. Studying invasion: have we missed the boat? Ecology Letters **8**:715–721.

Regan, H., M. Colyvan, and M. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications **12**:618–628.

Richardson, D., M. Pysek, M. Rejmanek, M. Barbour, F. Panetta, and C. West. 2000. Naturalization and invasion of alien plants: concepts and definitions. Diversity and Distributions **6**:93–107.

Sakai, A., F. Allendorf, J. Holt, D. Lodge, J. Molofsky, K. With, S. Baughman, R. Cabin, J. Cohen, N. Ellstrand, D. McCauley, P. O'Neil, I. Parker, J. Thompson, and S. Weller. 2001. The population biology of invasive species. Annual Reviews of Ecology and Systematics **32**:305–32.

Schneider, D., C. Ellis, and K. Cummings. 1998. A transportation model assessment of the risk to native mussel communities from Zebra Mussel spread. Conservation Biology **12**:788–800.

Shea, K., and P. Chesson. 2002. Community ecology theory as a framework for biological invasions. Trends in Ecology and Evolution **17**:170–176.

Shigesada, N., and K. Kawasaki. 1997. Biological Invasions: Theory and Practice. Oxford University Press, New York, NY.

Shrader-Frechette, K. 2001. Non–indigenous species and ecological explanation. Biology and Philosophy **16**:507–519.

Simberloff, D. 2003. Eradication – preventing invasions at the outset. Weed Science **51**:247–253.

Suter, G. 1993. Ecological Risk Assessment. Lewis Publishers, Chelsea, Michigan.

Taylor, C., and A. Hastings. 2005. Allee effects in biological invasions. Ecology Letters **8**:895–908.

Tilman, D. 2004. Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assesmbly. Proceedings of the National Academy of Science, U.S.A. **101**:10854–10861.

Vermeij, G. 1996. An agenda for invasion biology. Biological Conservation **78**:3–9.

Wangen, S., and C. Webster. 2006. Potential for multiple lag phases during biotic invasions: reconstructing an invasion of the exotic tree *Acer plantanoides*. Journal of Applied Ecology **43**:258–268.

Williamson, M. 1996. Biological Invasions. Chapman and Hall, London, United Kingdom.

Wonham, M., M. Lewis, and H. MacIsaac. 2005. Minimizing invasion risk by reducing propagule pressure: a model for ballast water exchange. Frontiers in Ecology and the Environment **3**:473–478.

Chapter 2

Waiting for invasions: a framework of the arrival of non–indigenous species

[1]

2.1 Introduction

The invasion of non-indigenous species (NIS) into new locations can be divided into three stages: arrival, establishment, and spread (Vermeij, 1996; Sakai et al., 2001). The establishment and spread stages garner considerable theoretical attention, but the arrival stage receives proportionally little consideration (Puth and Post, 2005), even though it represents a critical phase for preventing the establishment and potential impacts of new invaders (Parker et al., 1999). Arrival is the process by which individuals are transported from a source, through a dispersal pathway, to a destination where the NIS are introduced and attempt to establish (Carlton, 1996).

Many studies have attempted to identify the characteristics of successful invaders to predict future invaders (Chapter 6; Richardson and Rejmanek, 2004), while other studies suggest propagule pressure is the primary predictor of invasion success (Lockwood et al., 2005), but these studies are often limited to case–by–case assessments of invasion risk (Gilpin, 1990; Williamson, 1996). Such assessments

---

[1] The chapter was published as: Jerde, C.L., and M.A. Lewis. 2007. Waiting for invasions: A framework of the arrival of non–indigenous species. *The American Naturalist* 170:1-9.

are effectively tactical approaches for a particular system or a specific species (Godfray and Rees, 2002). By way of contrast, strategic models can provide a more general approach indicative of the invasion process rather than a specific case. The absence of a general framework for NIS arrival limits one to a tactical approach to predicting invasion that is potentially insufficient for developing a general foundation for predicting invasions. It has been suggested that by following a tactical approach to NIS arrival "we are never going to have a scheme to predict the success of invading species" (Gilpin, 1990), and this is likely one reason why an arrival framework for predicting invasions remains elusive (Carlton, 1996). However, Gilpin advocates a strategic approach to develop a framework that includes stochastic elements for predicting the arrival of NIS. This is the approach we adopt in this paper.

Throughout the arrival process, stochasticity influences the eventual establishment success or failure of the NIS. Stochastic events include abiotic changes in the source region modifying the number of individuals available for transport, establishment of new dispersal pathways, and variable environmental conditions of the destination influencing survival (Carlton, 1996). Also, some biological invasions occur with only a few introduced individuals. Therefore, demographic stochasticity, including Allee effects, is likely to contribute to the establishment or failure of new NIS populations (Shaffer, 1981; Lande, 1993; Leung et al., 2004). Taken in total, stochasticity leads to uncertainty in spatial and temporal predictions of invasion (Kolar and Lodge, 2001; Drake and Bossenbroek, 2004) and is likely one reason why tactical approaches have been emphasized previously (Gilpin, 1990; Carlton, 1996).

Here we develop a strategic, probabilistic framework for the arrival process of NIS from source to destination. Being a contingent process (Puth and Post, 2005), arrival emerges as a hierarchical model. We show the model can be simplified to a marginal distribution describing the number of surviving individuals introduced into the destination as a function of the propagule pressure and the individual probability of survival.

The hierarchical structure is suitable to estimate the probability of

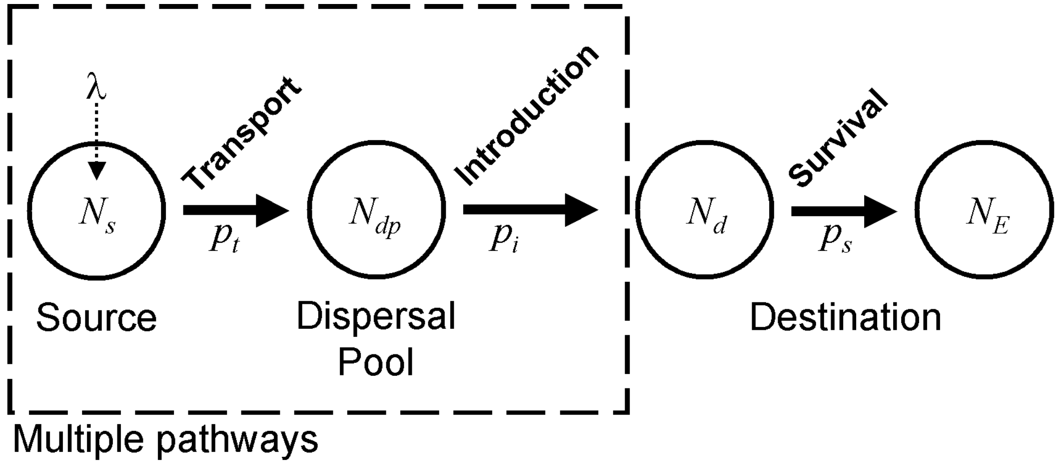


Figure 2.1: Flow diagram of the arrival process from source to destination. Variables and parameters are defined in Table 1.

establishment for one time step (discrete time). The time step formulation is flexible and may consider seasonal or yearly patterns of arrival. The increasing number of established invaders is likely linked to repeated introductions (Cohen and Carlton, 1998), so we examine the marginal distribution of establishment as a stochastic process (time–to–event model) and produce invasion waiting time distributions. The results of the hierarchical modeling and the time–to–event modeling produce the probability of establishment at the next time step and the probability of establishment within a defined period of time, respectively. While providing a theoretical framework with which to study the arrival process, the framework is also directly related to ecological risk assessment for NIS (Suter, 1993). I demonstrate how to estimate the waiting time risk (probability) and relative risk (ratio of probabilities) of two locations being invaded.

2.2 The arrival process

Conceptualizations of the arrival stage of the invasion process take many forms (Wonham et al., 2000; Sakai et al., 2001; Lockwood et al., 2005). Here, I adopt Carlton's (1996) description of source, dispersal, and destination pools (Figure 1). All parameters and variables are defined in Table 1. Throughout, I refer to individuals, populations, and pools for a single species unless otherwise noted.

Parameter or Variable	Description
N_E	Number of established individuals
$\sum N_d$	Propagule pressure into a destination
N_d	Number of introduced individuals into the destination from a single pathway
N_{dp}	Number of individuals in the dispersal pool
N_s	Number of individuals in the source population
λ	Central tendency of the source population
k	Index of introduction pathways
n	Number of introduction pathways
p_s	Probability an individual survives introduction into a destination
p_i	Probability an individual is introduced from the dispersal pool
p_t	Probability an individual is transported
ϕ	Expectation of propagule pressure from k pathways
γ	Probability at least one individual establishes
α	Shape parameter for considering Allee effects
T	Waiting time random variable
$\overline{N}_E$	Expected surviving propagule pressure
c	Constant of proportionality
R	Ratio of expected waiting times for two locations

Table 2.1: Definitions for the hierarchical model and arrival waiting time.

First, individuals at the source site are subject to environmental conditions such that the source population, N_s, is a random variable. The individuals are then transported from the source, N_s, to the dispersal pool, N_{dp}. Transported NIS constitute a dispersal pool of individuals. From the dispersal pool, individuals are then introduced into the destination. The process from source to destination represents one unique pathway of introduction. However, as the box of Figure 1 demonstrates, there are likely many pathways, such as several ships (Drake and Lodge, 2004). The input of individuals to a destination from all pathways is the propagule pressure ($\sum N_d$) of a species to a destination (Lockwood et al., 2005). Lastly, I consider survival of the introduced individuals (p_s). This does not capture the growth and spread but simply the physiological tolerance of the organism to the new environment (i.e. Spidle et al., 1995), also called inoculant survival (Smith et al., 1999). Additionally, p_s includes properties of habitat invasibility. Biotic (such as competition) and abiotic interactions at any new location have the potential to increase or decrease the probability of surviving and establishing (Jules et al., 2002; Von Holle, 2005).

The entire arrival process, from all pathways, occurs over time. For example, the arrival of aquatic NIS may be the number of introductions occurring over one year, or the arrival of seeds into a location during a particular season. Generally, I refer to the period of time considered in the arrival model as a time step. Hereafter, establishment refers to the event of an organism surviving in the new system beyond the arrival time step. More complexity, such as source population dynamics, population dynamics during transport, or different types of release mechanisms may be significant for a given species or pathway, but my simplified process is a reasonable, strategic skeleton of arrival.

2.3 A hierarchical model of NIS arrival

The usefulness of hierarchical modeling comes by reducing a larger process to a collection of simple processes. In its simplest form, a hierarchical model is a random variable with a distribution that is conditional on a random variable (see, for example, Casella and Berger, 2002; Clark, 2003). I assume individuals act independently from each other, and by following individuals, the model captures demographic stochasticity (Shaffer, 1981; Lande, 1993). In a probabilistic framework, an individual experiences a Bernoulli trial of being transported from one pool to another (such as being transported from the source to the dispersal pool). Independent, identically distributed (i.i.d.) Bernoulli trials for a group of individuals yields a binomial or Poisson distribution for the transition (Casella and Berger, 2002). Linking the transitions between source and destination, with the use of the binomial and Poisson distributions, leads to a hierarchical model of the contingent arrival process:

$$N_E \mid N_d \sim \text{Binomial}\,(N_d, p_s) \tag{2.1}$$

$$N_d \mid N_{dp} \sim \text{Binomial}\,(N_{dp}, p_i) \tag{2.2}$$

$$N_{dp} \mid N_s \sim \text{Binomial}\,(N_s, p_t) \tag{2.3}$$

$$N_s \sim \text{Poisson}\,(\lambda) \tag{2.4}$$

The Poisson distribution describes the number of individuals entering the arrival pathway while the binomial distributions are used as transition processes. Future investigations may consider alternative distributions both for the number of individuals entering the system and the transition processes. The binomial and Poisson distributions follow the probability mass functions found in most statistical texts (e.g. Casella and Berger, 2002) and are also found in Appendix A.

From the arrival process I can write a random variable for the population size at the source location (N_s) with a parameter λ describing the central tendency (equation 2.4). I use a 'constant only' model (λ) of the source population dynamics for simplicity, but future considerations may investigate the influence of alternative source population dynamic models such as autoregressive processes. A hierarchy emerges when the random variable for the number of individuals entering the dispersal pool (N_{dp}, equation 2.3) is formulated as a binomial distribution with the parameter p_t capturing the probability of an individual being transported and is conditional on the random variable N_s. Similarly, the random variable for the number of individuals being introduced into the destination (N_d, equation 2.2) can be written as a binomial process with the probability of being introduced (p_i) to a particular source and conditional on the contribution of individuals from the dispersal pool (Equation 2.2). Lastly, the probability of survival (p_s) parameterizes the distribution of the number of introduced individuals into the destination that survive to establish. Again, I model this as a random variable, N_E, conditional on the number of individuals released into the destination (Equation 2.1).

Equations 2.1 to 2.4 are a hierarchical system describing the transfer of individuals from a single source to a single destination, but for purposes of inference, I am interested in the marginal (unconditional) distribution of N_E and not the mixture distribution of $N_E \mid N_d$ (Casella and Berger, 2002). In order to find the marginal distribution of N_E, it is necessary to simplify the process.

Hierarchial processes can sometimes be simplified using conditional probability,

$$\text{Prob}(X = x) = \sum_y \text{Prob}(X = x \mid Y = y)\,\text{Prob}(Y = y). \tag{2.5}$$

From my formulation, a hierarchy is the number of transported individuals (equation 2.3) defined by the probability of being transported, p_t, and the number of individuals in the source population where N_s is a random variable from a Poisson distribution. Using conditional probability, the hierarchical statements can be reduced to a marginal distribution of N_{dp},

$$N_{dp} \sim \text{Poisson}(\lambda p_t). \tag{2.6}$$

The details of this simplification can be found in Appendix A.

Similarly, I can use the marginal distribution of N_{dp} and the conditional statement of the number of introduced individuals (equation 2.2) to formulate the marginal distribution of the number of introduced individuals,

$$N_d \sim \text{Poisson}(\lambda p_t p_i), \tag{2.7}$$

and follows the same procedure as demonstrated in section A of the online appendix. The marginal distribution of N_d is the number of individuals introduced to the destination from a single source.

To consider n independent pathways, I assume the input of multiple Poisson random variables. A useful mathematical result is that the sum of independent Poisson distributions is also Poisson distributed,

$$\sum N_d \;\sim\; \sum_{k=1}^{n} \text{Poisson}(\lambda_k \; p_{t,k} \; p_{i,k}) = \text{Poisson}(\phi), \tag{2.8}$$

where $\sum N_d$ is the distribution of the propagule pressure into the destination from n pathways, and $\phi = \sum_{k=1}^{n} \lambda_k \; p_{t,k} \; p_{i,k}$ is the expectation of the propagule pressure. The last hierarchical simplification is to collapse the propagule pressure (equation 3.1) into the conditional survival process (equation 2.1) to find the marginal distribution of the number of surviving individuals, N_E. Because the propagule pressure is Poisson distributed, I have the same simplification as observed with both the transport and introduction hierarchical statements. This leads to,

$$N_E \;\sim\; \text{Poisson}(p_s \; \phi). \tag{2.9}$$

The distribution of N_E captures the process variability of transporting individuals from multiple sources to a single destination and the demographic stochasticity of the individual probability of survival.

Estimating the probability of establishment of NIS may require consideration of the underlying population dynamics at the destination (Drake et al., 2005). Leung et al. (2004) have formulated the establishment process for two models, one that assumes independence in the arriving propagules and one that contains a shape parameter to test for the presences of Allee effects. The independence model of NIS establishment used in Leung et al. (2004) can be formulated from the arrival process modeled here by first considering the probability of not observing an establishment event. Using the marginal distribution for N_E, this is

$$\text{Prob}(N_E = 0) = e^{-p_s \, \phi}. \tag{2.10}$$

Its complement, the probability of observing at least one established individual,

$$\gamma = \text{Prob}(N_E \geq 1) = 1 - e^{-p_s\,\phi}, \tag{2.11}$$

is similar to the independence model of establishment proposed by Leung et al. (2004). The difference between the two formulations is here I have justified a Poisson distributed random variable, N_E, and Leung et al. (2004) used a binomial random variable where $\gamma = \text{Prob}(N_E \geq 1) = 1 - (1 - p_s)^{\phi}$. The Poisson is an approximation of the binomial when ϕ is large and p_s is small (Casella and Berger, 2002). The effect of small p_s on γ can be found by expanding e^{-p_s} in equation 2.11 as $1 - p_s + \frac{p_s^2}{2}\ldots$. Keeping the first two terms in this expansion yields γ for the binomial distribution. Thus, the arrival framework resulting from hierarchical modeling is a link to establishment phase of biological invasions and justifies the independence model found in Leung et al. (2004).

Leung et al. (2004) assess the presence of Allee effects by testing for the significance of a shape parameter, α (c in their formulation), on the independence model. Equation 2.11 of NIS arrival may be modified to include a shape coefficient test developed by Leung et al. (2004), $\gamma = 1 - e^{-(p_s\,\phi)^{\alpha}}$, but the Allee model formulation for NIS establishment is not further justified here, and I continue my investigation of NIS arrival following the independence model (Equation 2.11).

Suter (1993) refers to equations 2.10 and 2.11 as an endpoint. The endpoint terminology comes from ecological risk assessment and is used to define a "formal expression of the environmental values to be protected." In this case, the value to be protected is the absence of the NIS from a particular location (equation 2.10), and the risk (probability) associated with failure of this objective is the complement (equation 2.11). The endpoint is flexible and could be adjusted to monitor for the successful establishment of 10, 100, or more individuals. I use the endpoint defined by equations 2.10 and 2.11 hereafter.

2.4 Time-to-event stochastic processes

Given the probability of the event that at least one individual establishes is γ, it is possible to extend the dynamics of this system through time as a series of Bernoulli trials (Taylor and Karlin, 1998) to consider the contribution of repeated introductions. This assumes that, over discrete time steps, the arrival process can be modeled as an independent Bernoulli random variable. A property of i.i.d. Bernoulli trials is the geometric distribution of the waiting time, T, to the first event (establishment),

$$T \sim \text{Geometric } (\gamma)$$

with probability mass function,

$$\text{Prob}(T = t) \ = \gamma(1 - \gamma)^{t-1}, \tag{2.12}$$

where γ is defined by equation 2.11. Consider three locations with different propagule pressures (ϕ = 100, 50, and 10 individuals) and the same probability of survival, $p_s = 0.001$. Equation 2.11 yields γ = 0.10, 0.05, and 0.01. The waiting time distributions for three realization following equation 12 are illustrated in Figure 2. The geometric distribution has a mean, $1/\gamma$, and variance, $(1 - \gamma)/\gamma^2$ (Casella and Berger, 2002). This results in the property that, as the probability of at least one establishing individual becomes small, the mean and variance become large. Thus, by attempting to reduce NIS establishment events, we are increasing the uncertainty in when the NIS will establish. However, there is a trade–off in that we can now accurately state that an invasion is unlikely to occur. Yet, this never precludes the possibility of an invasion occurring in the next time step if $\gamma > 0$. Given the waiting time distribution, I can analytically determine the probability of an establishment by time t according to

$$\text{Prob(establishment by } t) = \text{Prob}(T \leq t) = 1 - (1 - \gamma)^t. \tag{2.13}$$

Using Figure 2 as an example and setting $t = 10$ time steps, the risk of establishment would be the sum from 1 to 10 of the geometric distributions for

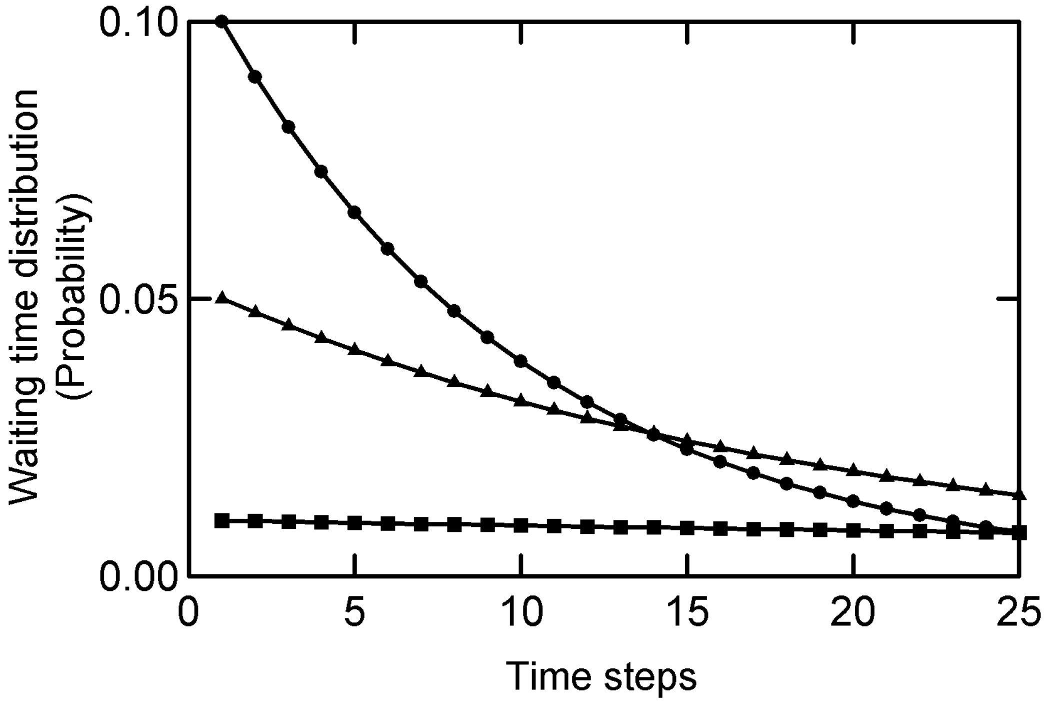


Figure 2.2: The geometric probability mass function for the time to establishment where $\gamma = 0.1$ (circles), 0.05 (triangles), and 0.01 (squares). Using the geometric waiting time distribution, as $\gamma \rightarrow 0$, the distribution becomes flat. Thus by reducing the propagule pressure, ϕ, or the survival probability, p_s, we increase the uncertainty in when an invasion will occur. However, there is a reduced probability (equation 2.13) that invasion will occur in the near future (i.e if $t = 3$).

$\gamma = 0.10$, 0.05, and 0.01. This results in a probability of an establishment event (Equation 2.13) within the next 10 time steps equal to 0.65, 0.40, and 0.10, respectively.

The probability of at least one individual establishing in a single time step, γ, and the probability of invasion by time t are two different, albeit closely related, endpoints. When $t = 1$ the two probabilities are equal, (i.e. Equation 2.13 yields γ). Thus, γ provides one time step predictions of invasion, while the probability of invasion by time t allows for prediction over a longer time frame.

2.5 Patterns of invasion

There are many reasons why estimating the individual survival probability, p_s, and the propagule pressure, ϕ, may be difficult. It may be economically infeasible to monitor continuously for the introduction of NIS, whose detection is further complicated when only a few individuals are introduced. This is the dilemma community ecologists face when accounting for rare species (Longino and Colwell, 1997). Estimating individual survival probabilities face another set of difficulties in that uninvaded locations likely require not only the standard experimental efforts and logistics (e.g. Von Holle and Simberloff, 2005), but also substantial security measures as to prevent the experiment from being the source of an NIS establishment event. However, possibly the most limiting factor to all of these estimation considerations is time. Experiments and observations take time to conduct and collect while inaction in preventing establishment may allow for the invasion to occur (Simberloff, 2003).

The approach many ecologists have taken is to use relative measures of individual survival or propagule pressure to assess invasion risk in a timely manner. For example, ‘gravity models’ (see Chapter 4 for more details) that describe movement rates with a formulation analogous to Newton’s laws of attraction are useful in estimating relative abundance of aquatic NIS transported through recreational boater traffic (Schneider et al., 1998; Bossenbroek et al., 2001; Muirhead and MacIsaac, 2005). Relative measures of survival come from habitat matching studies between source and destinations (e.g. Drake and Bossenbroek, 2004; Herborg et al., 2007). Therefore, applying relative measures of propagule pressure and survival probability would be useful for uncovering the patterns of successful invasions from the arrival framework. The hierarchical model for the arrival process provides a mechanistic basis for comparing invasion waiting times for locations with different propagule pressures and probabilities of survival. Here I show how relative measures can be used to calculate relative waiting times and predict the order invasion into two locations.

I refer to the expected surviving propagule pressure as $\overline{N}_E = p_s \phi$ (see equation 2.9) and can be interpreted as the expected number of establishing individuals from the independence model of Leung et al. (2004). Suppose two locations have different pressures, where I denote $\overline{N}_{E_A}$ for the expected surviving propagule pressure in location A and $\overline{N}_{E_B} = c\ \overline{N}_{E_A}$ for the expected surviving propagule pressure in location B. Intuitively we would expect that $c < 1$ would lead to location A, on average, being invaded before location B, due to reduced expected surviving propagule pressure, and $c > 1$ would lead to location B, on average, being invaded before location A, due to increased expected surviving propagule pressure in B. These are the arguments used to justify ranked risk assessments of locations found in the gravity model literature (Schneider et al., 1998; Bossenbroek et al., 2001; Leung et al., 2004). However, although ranks may be useful in determining which locations are more likely to be invaded, they do not reveal the probability of invasion (probabilistic risk), or the relative odds that one location will be invaded over another (relative risk).

The time-to-event extension of the hierarchical model allows us to forecast probabilities and relative risk associated with the outcomes regarding invasion times for specified locations. I define $T_A \sim$ Geometric(γ_A) (see equation 2.12) to be the random variable describing the waiting time that location A is first invaded, where $\gamma_A = 1 - e^{(-\overline{N}_{E_A})}$ (see equations 2.10 and 2.11). Likewise $T_B \sim$ Geometric(γ_B) is defined as the random variable describing the waiting time that location B is first invaded, where $\gamma_B = 1 - e^{(-c\,\overline{N}_{E_A})}$. The ratio, R, of expected time for invasion of location B to expected time for invasion of location A is

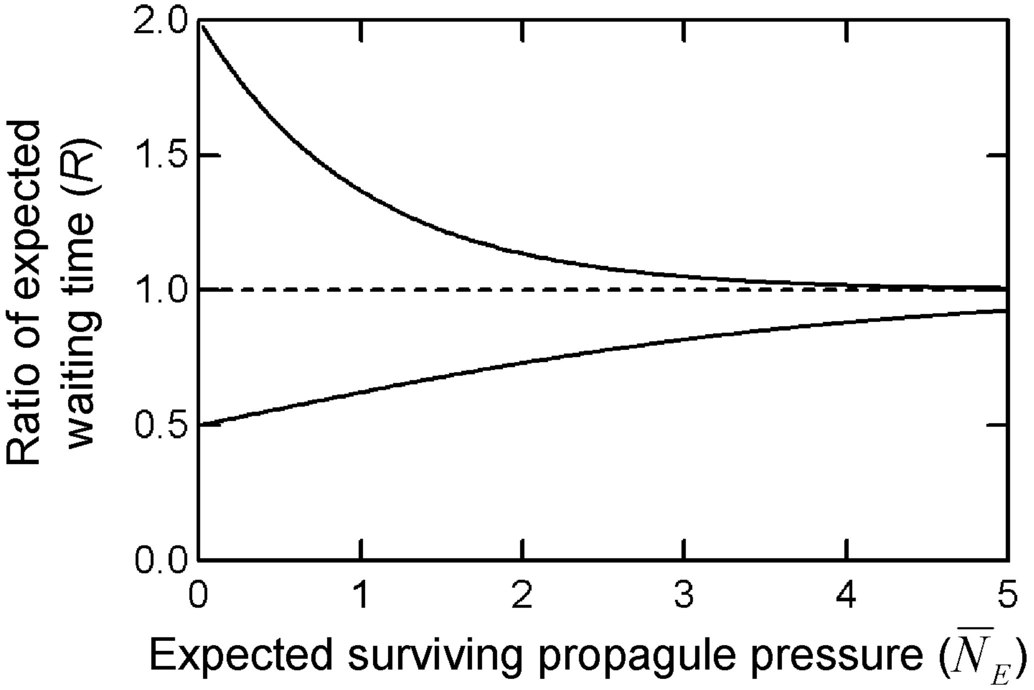


Figure 2.3: The ratio, R, of expected time for invasion of location B to expected time for invasion of location A as a function of $\overline{N}_{E_A}$. The region above the dashed line is for $c < 1$ and the region below for $c > 1$. As $\overline{N}_{E_A}$ becomes large, the ratio becomes 1. As $\overline{N}_{E_A}$ goes to zero, the ratio converges to c^{-1}. Two realizations, $c = 2$ and $c = 0.5$, are shown.

$$R = \frac{\mathrm{E}[T_B]}{\mathrm{E}[T_A]} = \frac{\frac{1}{\gamma_B}}{\frac{1}{\gamma_A}} = \frac{\gamma_A}{\gamma_B} = \frac{1 - e^{(-\overline{N}_{E_A})}}{1 - e^{(-c\,\overline{N}_{E_A})}}. \tag{2.14}$$

If $\gamma_A = 0.4$ and $\gamma_B = 0.1$, then we should expect to wait, on average, $R = 4$ times longer for location B to be invaded than location A. This ratio, R, is a decreasing function of the absolute risk of invasion, as given by the expected surviving propagule pressures $\overline{N}_{E_A}$ and $\overline{N}_{E_B} = c\,\overline{N}_{E_A}$ where $c < 1$, and an increasing function if $c > 1$ (Fig. 3). In the case where γ becomes small, $\overline{N}_{E_A} \to 0$ and $c > 1$, the ratio of expected times in equation 2.14 converges to $c^{-1} = \overline{N}_{E_A}/\overline{N}_{E_B}$, a measure of relative risk. Hence relative risk provides a useful measure for the ratio of expected invasion waiting times when the expected surviving propagule pressure is small.

Empirical evidence suggests propagule pressure is often a predictor of invasions (Lockwood et al., 2005; Von Holle and Simberloff, 2005). If we compare two locations with the same, but unknown p_s, then the expected propagule pressure, ϕ, can give estimates of the relative waiting times such that we should on average expect to wait $c^{-1} = \phi_A/\phi_B$ times longer for location B to be invaded than location A.

Bossenbroek et al. (2001), using a gravity model of boater traffic, provides estimates of Zebra mussel (*Dreissena polymorpha*) dispersal, Q_u, of the relative number of infested boats visiting uninvaded lake u per year. Assuming the zebra mussel propagule pressure is proportional to infected boater traffic and the expected surviving propagule pressure, $\overline{N}_E$, for both locations is small, the ratio of boater traffic is an estimate of the relative expected waiting time of invasion. For example, if lake A has Q_A=100 and lake B has Q_B=5, then we should expect to wait on average 20 times longer for lake B to be invaded than lake A. Thus, estimates of relative propagule pressure, such as Q_u, provide an estimate of the relative waiting time that is more informative than ranks currently provided.

Additionally it is possible to estimate the probabilities of the ordered invasion of two locations. There are three possible outcomes: location A invaded before location B, location A and location B invaded at the same time, and location B invaded before location A. The derivation of these probabilities are presented in section B of the online appendix and the results are,

$$\begin{aligned}\text{Prob}(B \text{ before } \boldsymbol{A}) &= \frac{\gamma_B - \gamma_A\gamma_B}{\gamma_A + \gamma_B - \gamma_A\gamma_B} \\ \text{Prob}(B \text{ and } \boldsymbol{A}) &= \frac{\gamma_A\gamma_B}{\gamma_A + \gamma_B - \gamma_A\gamma_B} \\ \text{Prob}(A \text{ before } \boldsymbol{B}) &= \frac{\gamma_A - \gamma_A\gamma_B}{\gamma_A + \gamma_B - \gamma_A\gamma_B}.\end{aligned} \tag{2.15}$$

Again, γ_A and γ_B are the probabilities of at least one individual establishing in location A and location B, respectively, and the γ estimates are functions of the propagule pressure and individual probability of survival (equation 2.11). When the product of the probability of at least one establishing individual for both locations is small ($\gamma_A\gamma_B \to 0$), the probability of location B being invaded before location A is $1/(1+R)$, the probability of location A being invaded before location B is $R/(1+$

R) (formulations shown in section B of the online appendix). The probability of both locations being invaded at the same time step is approximately zero.

2.6 Examples

The arrival framework can be connected to experimental and observational studies. Here I provide two analyses. First, using the invasive species scentless chamomile, I manipulate the propagule pressure for two locations and demonstrate how parameter estimation of the individual probability of survival can be used to construct γ and waiting time endpoints. I then evaluate the use of relative formulations of waiting time and ordered invasions and compare them to the formulations using γ. In the second example, I use observational data of water-ballast discharge into five U.S. ports to estimate the relative waiting time risk of Chinese mitten crab invasion. Additionally, I calculate the order of invasion for location pairs.

Scentless chamomile (*Matricaria perforata*)

I performed an invasion experiment attempting to have approximately the same individual probabilities of survival, p_s, for two locations (A and B) with one location receiving twice the propagule pressure, ϕ. The experiment was conducted at the Alberta Research Council's experimental field station in Vegreville, Alberta, Canada. Two adjacent plots were subdivided into 50 cells. Location A received 10 seeds per cell while location B received 20 seeds per cell, and the density was 1 seed per $10cm^2$ for both locations. I measured the number of plants in each cell that matured to either rosette or flowering form. The data are presented in Table 2.6.

Using the Poisson distribution with known propagule pressure (see equation 3.1), I estimated the individual probability of survival using maximum likelihood estimation. Bootstrapped confidence intervals (Table 2.6) allowed us to conclude the individual probabilities of survival are not statistically different (Efron and Tibshirani, 1993). Because both locations have a known propagule pressure and an estimated probability of survival, the risk of establishment, γ, can be estimated (Equation 2.11;γ_A and γ_B in Table 2.6). The probability of location A or location B

Location	Propagule pressure	Data	$\hat{p_s}$ [90% C.I.]	γ
Location A	10 seeds per cell	(0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 2, 1, 2, 1, 0, 1, 0, 0, 2, 1, 0, 1, 1, 0, 0, 0, 0, 1, 2, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 3, 1, 1, 0)	0.064 [0.046, 0.083]	0.473
Location B	20 seeds per cell	(2, 2, 0, 2, 0, 0, 0, 1, 1, 0, 1, 3, 0, 0, 0, 1, 1, 1, 4, 3, 3, 0, 1, 1, 0, 2, 1, 0, 0, 0, 0, 1, 2, 0, 0, 2, 1, 0, 2, 0, 3, 0, 1, 1, 1, 2, 5, 4, 1, 1)	0.057 [0.045, 0.070]	0.680

Table 2.2: Scentless chamomile establishment data for two locations with different levels propagule pressure. Individual probability of survival is estimate using maximum likelihood estimation with bootstrapped confidence intervals. γ is calculated using equation 2.11.

being invaded by $t = 10$ (from Equation 2.13) is approximately 1. The relative risk (R from Equation 2.14) is 1.437. Using Equation 2.15 the probability of location B being invaded before A is 0.431, the probability of both locations being invaded at the same time is 0.387, and the probability of location A being invaded before B is 0.182.

All of the above quantities are calculated from the probabilistic formulations presented within the text. However, estimates of the individual probability of survival are not always available. The question is, 'how would my estimates of relative risk and the probabilities of ordered invasion change if I worked this same example but under the assumption that the individual probability of survival for both locations was the same but unknown?' First, the relative risk becomes a ratio of propagule pressures into both locations, $R = \frac{\phi_B}{\phi_A} = \frac{20}{10} = 2$. Then, using R, I estimate the probability of B invading before A, (R/(1+R)=2/3), and similarly the probability of A invading before B, (1/(1+R)=1/3). The probability of both establishing at the same time is assumed zero due to the assumption that γ_A and γ_B are small.

Scentless chamomile is a well established NIS throughout the region surrounding the experimental plots (contamination is discussed further in Chapter 3), and therefore it is not surprising that the individual probabilities of survival are large (see Table 2.6). Even with propagule pressures of 10 and 20 individuals, the resulting γ's for both locations are large. Therefore the relative formulation (using propagule pressure only) provides a poor estimate of the relative waiting time and order of invasion. This experiment identifies the limitation in using relative measures in identifying the patterns of invasion. When γ is large, all locations are likely invaded at the next time step and the temporal patterns (waiting time and order of invasion) are deterministic. However, some species have been repeatedly introduced for many time steps and yet there are no or few records of invasion. We may therefore reasonably assume that γ is small. Such is the case with the Chinese mitten crab.

Chinese mitten crab (*Eriocheir sinensis*)

Location	Ballast water	Waiting time ($\times$ longer than Norfolk)	Prob. of invasion before Norfolk
Norfolk,VA	36.5	1	–
Baltimore, MD	7.6	4.8	0.17
Portland, OR	6.5	5.6	0.15
New York, NY	2.1	17.4	0.05
Seattle, WA	1.3	27.9	0.03

Table 2.3: Ballast water discharged $tonnes \times 10^7\ year^{-1}$ into five United States shipping ports. The relative waiting time, following equation 2.14 suggests Norfolk is the most susceptible to invasion. The ports of Baltimore, Portland, New York, and Seattle are unlikely to be invaded before Norfolk.

Chinese mitten crab has successfully invaded much of Europe and considered a likely invader of North America. Herborg et al. (2007) compared covariates associated with invaded location in Europe to identify ports in the United States with similar covariates. Five ports (Norfolk,VA; Baltimore, MD; Portland, OR; New York, NY; and Seattle, WA) were identified with medium to high habitat similarity, but currently none of these ports are invaded (as of spring 2006, but see Ruiz et al. (2006) for additional discussion). Individual probabilities of survival and propagule pressure are unknown for all ports. However over a 3.5 year period, it is known that 36.5, 7.6, 6.5, 2.1, and 1.3 tonnes ($\times 10^7$) of ballast water was delivered to these ports, respectively (Herborg et al., 2007). Assuming the individual probability of survival is small and that propagule pressure is proportional to ballast water, I estimate the relative waiting time for invasion. Table 2.6 provides relative waiting times for all ports relative to the mostly likely invaded, Norfolk, VA.

An example of interpreting the relative waiting time formulation is that we should expect to wait, on average, 5.6 times longer for Portland to be invaded than Norfolk. However, this does not guarantee that Norfolk will be invaded before Portland by chance. Using the order formulation (Equation 2.15), we can estimate the probability of Norfolk being invaded before Portland as 0.85 while the probability of Portland being invaded before Norfolk is the complement, 0.15. Table 2.6 also reports probabilities of the each port being invaded before the most likely port, Norfolk.

2.7 Discussion

Two approaches have emerged to predict the arrival and successful establishment of NIS. The first approach emphasizes characteristics of the invader or the location allowing for the NIS to survive and establish (Richardson and Rejmanek, 2004). The second approach argues that propagule pressure alone is a sufficient indicator to predict successful invasions (Lockwood et al., 2005). However, neither approach is general enough to move beyond a primarily case–by–case treatment of NIS arrival (Gilpin, 1990). The hierarchical probability model highlights the limitation of these approaches for predicting invasions. Consider two species entering a location. The first has $p_s = 0.01$ and $\phi = 10$ while the second has $p_s = 0.0001$ and $\phi = 1000$. Strict adherence to either approach would identify one invader as more likely to be successful than the other, yet γ, and the resulting waiting time for both species, is the same. Therefore, a strategic approach to predicting successful invasions should consider both the propagule pressure and the mechanisms influencing the individual probability of survival.

When there is very little propagule pressure and a small probability of survival, the invasion risk is negligible, and when there is overwhelming propagule pressure, the invasion dynamics are essentially deterministic. My results regarding invasion waiting time distributions and order of invasions are most applicable to levels of intermediate propagule pressure where there is appreciable invasion risk but uncertainty as to whether or when the invasion will occur. Explicit incorporation of survival, p_s, in the expected surviving propagule pressure means that the hierarchical framework allows for investigation of the influences of changes in survival on NIS establishment and has the advantage of being connected to the influences of propagule pressure. For example, Simberloff and Von Holle (1999) proposed the concept of 'invasional meltdown' whereby the presence of NIS in a location increases the p_s for an arriving NIS. Similarly, hypotheses such as the role of fluctuating resources (Davis et al., 2000), enemy release (Keane and Crawley, 2002), or biotic resistance (Levine et al., 2004) may also be formulated as a change in p_s leading to increases or decreases in the probabilities of establishment, γ, and the resulting waiting time distributions. Likewise, the natural history of the NIS or the location being invaded may change the individual probability of survival.

My modeling approach suggests strategies for controlling invasion even when a location is overwhelmed by propagule pressure (Von Holle and Simberloff, 2005). Indeed, the arrival of NIS in such abundances that γ approaches one essentially guarantees that invasion will occur in the near future. In such cases, attempting to reduce invasion success will consist of evaluating the contribution of parameters λ, p_t, p_i and p_s to the probability of establishment, γ. It may be that because the individual probability of survival, p_s, is large that even a few individuals are enough to overwhelm the system. Alternatively, if p_s is small, then it may take orders of magnitude more propagule pressure to overwhelm the system. In this situation, evaluating the parameters of the arrival process may lead to more effective targeting of management actions such as reducing the mean abundance of individuals available for transport, λ. Additionally, when comparing the risk of establishment for multiple locations, if p_s is sensitive to difference or changes in biotic and abiotic conditions, then the amount of propagule pressure needed to overwhelm the system will also be sensitive to the biotic and abiotic conditions.

The invasion waiting time is a pattern of the invasion process resulting from the arrival and establishment stages. Strategically, both the propagule pressure and the individual probability of survival will determine the distribution of the waiting time and the order of invaded locations through the parameter γ. However, many tactical approaches to predicting invasion have used relative measures to rank locations most likely to be invaded. Two assumptions should be emphasized that are used to estimate the relative waiting time, R, and the order of invasion for two locations $(R/(1+R), 1/(1+R))$. First, $\overline{N}_E$ is assumed to be small. If $\overline{N}_E$ is moderately large, then both locations are likely overwhelmed by propagule pressure and both locations are deterministically invaded at the next time step. However, in systems such as the Great Lakes, zebra mussel where it is known that individuals have been introduced to inland lakes that remain uninvaded, this assumption seems reasonable (Johnson et al., 2001; Bossenbroek et al., 2001). Second, p_s is assumed to be the same between locations, though unknown. A violation of this assumption would potentially result in a similar misleading conclusion as demonstrated at the beginning of this discussion, and would not only influence the relative waiting time but also the ranks that are commonly present in the invasion literature.

Tactical approaches to biological invasion are undoubtedly useful for performing risk assessments and making management decisions. But a strategic approach may help us reveal the mechanisms of biological invasions that more generally allow NIS to eventually establish or fail. I believe this framework is a skeleton on which to model the NIS arrival process and strategically identify patterns of successful invasion.

Bibliography

Bossenbroek, J., C. Kraft, and J. Nekola. 2001. Prediction of long-distance dispersal using gravity models: Zebra Mussel invasion of inland lakes. Ecological Applications **11**:1778–1788.

Carlton, J. 1996. Pattern, process, and prediction in marine invasion ecology. Biological Conservation **78**:97–106.

Casella, G., and R. Berger. 2002. Statistical Inference. 2nd edition. Duxbury, Pacific Grove, CA.

Clark, J. 2003. Uncertainty and variability in demography and population growth: a hierarchical approach. Ecology **84**:1370–1381.

Cohen, A., and J. Carlton. 1998. Accelerating invasion rate in a highly invaded estuary. Science **279**:555–558.

Davis, M., J. Grime, and K. Thompson. 2000. Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology **88**:528–534.

Drake, J., P. Baggenstos, and D. Lodge. 2005. Propagule pressure and persistence in experimental populations. Biology Letters **1**:480–483.

Drake, J., and J. Bossenbroek. 2004. The potential distribution of Zebra Mussels in the US. BioScience **54**:931–941.

Drake, J., and D. Lodge. 2004. Global hot spots of biological invasions: evaluating options for ballast-water management. Proceedings of the Royal Society of London, Series B **271**:575–580.

Efron, B., and R. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall / CRC, Boca Raton, Florida.

Gilpin, M. 1990. Review: ecological prediction. Science **248**:88–89.

Godfray, H., and M. Rees. 2002. Population growth rates: issues and an application. Philosophical Transactions of the Royal Society of London, B **357**:1307–1319.

Herborg, L., C. Jerde, D. Lodge, G. Ruiz, and H. MacIsaac. 2007. Predicting the North American distribution of Chinese mitten crabs (*Eriocheir sinensis*) using measures of propagule pressure and environmental niche models. Ecological Applications **17**:663–674.

Johnson, L., A. Ricciardi, and J. Carlton. 2001. Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. Ecological Applications **11**:1789–1799.

Jules, E., M. Kauffman, W. Ritts, and A. Carroll. 2002. Spread of an invasive pathogen over a variable landscape: a nonnative root rot on Port Orford cedar. Ecology **83**:3167–3181.

Keane, R., and M. Crawley. 2002. Exotic plant invasions and the enemy release hypothesis. Trends in Ecology and Evolution **17**:164–170.

Kolar, C., and D. Lodge. 2001. Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution **16**:199–204.

Lande, R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. American Naturalist **142**:911–927.

Leung, B., J. Drake, and D. Lodge. 2004. Predicting invasions: propagule pressure and the gravity of Allee effects. Ecology **85**:1651–1660.

Levine, J., P. Adler, and S. Yelenik. 2004. A meta–analysis of biotic resistance to exotic plant invasions. Ecology Letters **7**:975–989.

Lockwood, J., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. Trends in Ecology and Evolution **20**:223–228.

Longino, J., and R. Colwell. 1997. Biodiversity assessment using structured inventory: Capturing the ant fauna of a tropical rain forest. Ecological Applications **7**:1263–1277.

Muirhead, J., and H. MacIsaac. 2005. Development of inland lakes as hubs in an invasion network. Journal of Applied Ecology **42**:80–90.

Parker, I., D. Simberloff, W. Lonsdale, K. Goodell, M. Wonham, P. Kareiva, M. Williamson, B. Von Holle, P. Moyle, J. Byers, and L. Goldwasser. 1999. Impact: Toward a framework for understanding the ecological effects of invaders. Biological Invasions **1**:3–19.

Puth, L., and D. Post. 2005. Studying invasion: have we missed the boat? Ecology Letters **8**:715–721.

Richardson, D., and M. Rejmanek. 2004. Conifers as invasive aliens: a global survey and predictive framework. Diversity and Distributions **10**:321–331.

Ruiz, G., L. Fegley, P. Fofonoff, Y. Cheng, and R. Lemaitre. 2006. First records of *Eriocheir sinensis* H. Milne Edwards, 1853 (Crustacea: Brachyura: Varunidae) for Chesapeake Bay and the mid-Atlantic coast of North America. Aquatic Invasions **1**:137–142.

Sakai, A., F. Allendorf, J. Holt, D. Lodge, J. Molofsky, K. With, S. Baughman, R. Cabin, J. Cohen, N. Ellstrand, D. McCauley, P. O'Neil, I. Parker, J. Thompson, and S. Weller. 2001. The population biology of invasive species. Annual Reviews of Ecology and Systematics **32**:305–32.

Schneider, D., C. Ellis, and K. Cummings. 1998. A transportation model assessment of the risk to native mussel communities from Zebra Mussel spread. Conservation Biology **12**:788–800.

Shaffer, M. 1981. Minimum population sizes for species conservation. BioScience **31**:131–134.

Simberloff, D. 2003. How much information on population biology is needed to manage introduced species? Conservation Biology **17**:83–92.

Simberloff, D., and B. Von Holle. 1999. Positive interactions of nonindigenous species: Invasional meltdown? Biological Invasions **1**:21–32.

Smith, L., M. Wonham, L. McCann, G. Ruiz, A. Hines, and J. Carlton. 1999. Invasion pressure to a ballast–flooded estuary and an assessment of inoculant survival. Biological Invasions **1**:67–87.

Spidle, A., E. Mills, and B. May. 1995. Limits to tolerance of temperature and salinity in the Quagga Mussel (*Dreissena bugensis*) and the Zebra Mussel (*Dreissena polymorpha*). Canadian Journal of Fisheries and Aquatic Sciences **52**:2108–2119.

Suter, G. 1993. Ecological Risk Assessment. Lewis Publishers, Chelsea, Michigan.

Taylor, H., and S. Karlin. 1998. An Introduction to stochastic modeling. 3rd edition. Academic Press, San Diego, Calf.

Vermeij, G. 1996. An agenda for invasion biology. Biological Conservation **78**:3–9.

Von Holle, B. 2005. Biotic resistance to invader establishment of a southern Appalachian plant community is determined by environmental conditions. Journal of Ecology **93**:16–26.

Von Holle, B., and D. Simberloff. 2005. Ecological resistance to biological invasion overwhelmed by propagule pressure. Ecology **86**:3212–3218.

Williamson, M. 1996. Biological Invasions. Chapman and Hall, London, United Kingdom.

Wonham, M., J. Carlton, G. Ruiz, and L. Smith. 2000. Fish and ships: relating dispersal frequency to success in biological invasions. Marine Biology **136**:1111–1121.

Chapter 3

The invasion waiting time of scentless chamomile

3.1 Introduction

Uninvaded locations can be overwhelmed by incoming invaders such that predicting invasion essentially becomes deterministic (Von Holle, 2005). Alternatively, some locations receive a large number of introductions but remain uninvaded (Williamson, 1996). Between these two extremes are intermediately susceptible locations where species are regularly introduced, but fail, only to later become established. The variable success of some invaders is presumably due to chance events in the pathways of introduction, life history, and environmental conditions (Carlton, 1996) and is likely why patterns of successful invaders continue to undermine prediction (Gilpin, 1990). Determining which species and locations are likely to invade or to become invaded, and when, are critical for predicting invasions (Kolar, 2004), and we are motivated to accurately forecast biological invasions in order to design monitoring strategies, prioritize control efforts, and ultimately prevent the impacts, both ecologically and economically, due to invasion (Vermeij, 1996; Parker et al., 1999; Pimentel et al., 2005).

In Chapter 2, I proposed a framework to evaluate the risk of invasion for a single species based on the number of individuals introduced into a location (ϕ) and the survival (p_s) of these individuals. For a single time step, the number of establishing individuals (N_E) is a Poisson random variable,

$$\text{Prob}(N_E = n_E) = \frac{e^{-p_s \phi}\ p_s \phi^{n_E}}{n_E!}. \tag{3.1}$$

Establishment is defined as the probability at least one individual establishes (γ),

$$\gamma_P = \text{Prob}(N_E \geq 1) = 1 - \text{Prob}(N_E = 0) = 1 - e^{-p_s\ \phi}, \tag{3.2}$$

but could be modified to capture the probability of a specified level of individuals establishing. Extending the introduction of individuals at the same rate and with the same survival in discrete time leads to invasion waiting time distributions that follow a geometric distribution (T),

$$\text{Prob}(T = t) = \gamma(1 - \gamma)^{t-1}, \tag{3.3}$$

where t=1,2,3,..∞, and is the time at which establishment occurs. A flat waiting time distribution ($\gamma \rightarrow 0$) implies a low risk of invasion and a right skewed distribution ($\gamma \rightarrow 1$) of the invasion waiting time implies a high risk of invasion. Waiting time distributions also provide predictions regarding the expected frequency of invasions into multiple locations and the order of invasion into different locations (Chapter 2; Jerde and Lewis, 2007). What remains unresolved is whether biological invasions follow these waiting time distributions.

Using the invader scentless chamomile (*Matricaria perforata*), I spatially replicate the invasion process to estimate the probability of establishment and validate the proposed waiting time formulation. Because count data are susceptible to greater than expected variation in the response variable (N_E), I test for overdispersion and discuss the sources and consequence of an overdispersed invasion process (see, for example Pielou, 1969).

3.2 The invasion of central Canada by *Matricaria perforata*

Scentless chamomile is a widespread, short–lived perennial that was first introduced to Canada from its native European range in the early twentieth-century either through contaminated livestock feed or for commercial horticulture. Although

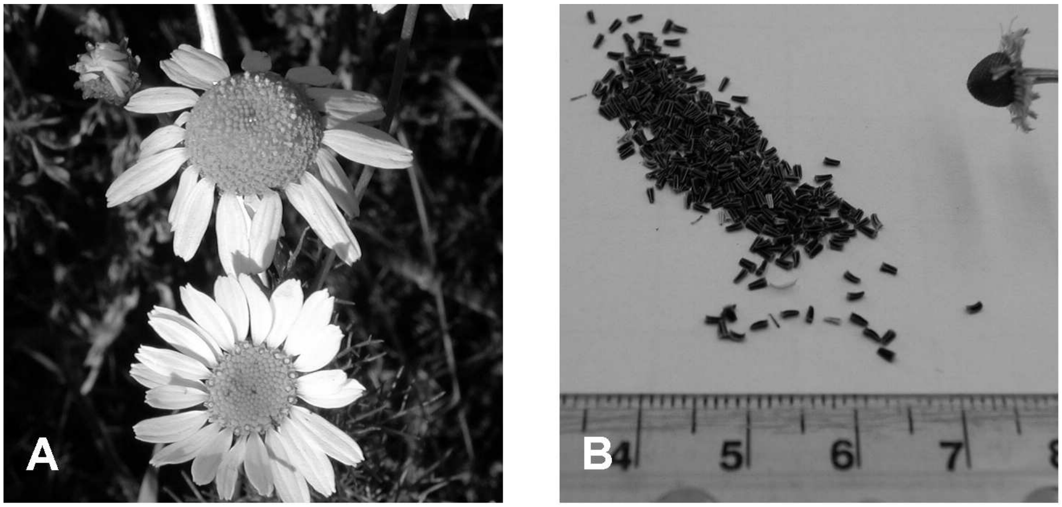


Figure 3.1: Mature scentless chamomile plants (Panel A) can have a single or potentially hundreds of flowering heads. The introduction of scentless chamomile may be due in part to its attractive daisy–like appearance. The head of any flower produces hundreds of achenes (Panel B) with the seed protected within.

scentless chamomile has a perennial life history, most plants found in Manitoba, Saskatchewan, and Alberta, are annuals that germinate in early spring and mature by mid to late summer. Perennial specimens have an intermediate stage of a rosette form (Woo et al., 1991).

There is considerable variation in the number of heads found on a single plant (fig. 3.1A) as well as variation in the number of achenes (seeds) per head (fig. 3.1B). The variability in survival and seed production is significantly influenced by spring precipitation, soil disturbance, and interspecific competition. Above average spring precipitation leads to increased survival, and disturbed soil where competitors are removed allows for increased densities of plants (Bowes et al., 1994). Under average, abiotic conditions across Canada, scentless chamomile can increase to densities of more than 25 plants m^{-2} when left unchecked, which can lead to wheat yield losses of greater than 55% (Douglas et al., 1991).

When a flowering scentless chamomile head matures, it produces hundreds of achenes that fall to the ground. The seeds generally do not disperse far from the adult plant. de Camino-Beck (2006) showed that 99% of seeds fell within 5m of the adult. However, scentless chamomile is often mixed with agricultural crops, such as wheat (Douglas et al., 1991) or hay harvested for livestock, and is transported by human activities rather than natural dispersal. The anthropogenic transport of individuals is characteristic of many invaders (Carlton, 1996), but the threshold as to how many invaders are required to overwhelm a system and produce a new invaded location is species and location specific (Chapter 2; Von Holle and Simberloff, 2005; Jerde and Lewis, 2007).

Controlling scentless chamomile has proven difficult. Although multiple biological control agents have been released, none have proven successful in eradicating scentless chamomile from test plots (McClay, 2003). de Camino-Beck (2006) studied the life history and found the net reproductive rate (R_0) was much greater than one and identified the necessary levels of damage a biological control agent would be required to inflict to reduce the R_0 below one. When R_0 is greater than one, the establishment of just one individual is sufficient to allow a population to grow and invade a location. He concluded none of the biological control agents proposed or released are capable of achieving this objective. Herbicide and physical removal have been the only treatments proven to eradicate scentless chamomile from agricultural plots (McClay, 2003).

3.3 Methods

3.3.1 Site and experimental design

All experiments were conducted at the Alberta Research Council's research station in Vegreville, Alberta, Canada between 1 June 2005 and 1 October 2005. The tilled, hay field selected for the mensurative experiments was previously used for scentless chamomile dispersal and life history studies (de Camino-Beck, 2006). The experimental field (approximately 7m by 30m) was divided in half, here after referred to as location A and location B. Each location contained 10 plots with 110 cells per plot with an area of $100cm^2$ per cell. Ten cells within each plot were randomly selected as controls to monitor background scentless chamomile abundances. Location A received 10 scentless chamomile seeds per cell that were pressed into the ground, and location B received 10 seeds per cell that were not pressed into the ground. This was not a test of treatment effects (i.e. pressed versus non-pressed) as this experimental layout would be pseudoreplicated (Hurlbert, 1984). The pressed versus non-pressed application was to differentiate the individual probability of survival and produce two different waiting time predictions and mimicked tilled and wind dispersed achenes, which is known to influence survival (Woo et al., 1991). The plots were not weeded or watered, but few other species established within the plots. Establishment was determined at the end of the growing season by monitoring the number of adult (rosette or flowering form) individuals per cell.

3.3.2 Estimating survival and testing for overdispersion

Each cell provided an observation ranging from 0 to 10 individuals establishing, N_E. To estimate the individual probability of survival, p_s, for each location, I used maximum likelihood estimation applied to the Poisson (equation 3.1) and negative binomial distributions with $\phi = 10$,

$$\text{Prob}(N_E = n_E) = \binom{k + n_E - 1}{k - 1} \left(\frac{p_s \phi}{k}\right)^{n_E} \left(1 + \frac{p_s \phi}{k}\right)^{-(k+n_E)}. \qquad (3.4)$$

The parameter k of the negative binomial is a measure of overdispersion. Overdispersion increases as $k \rightarrow 0$, and the negative binomial reduces to a Poisson distribution as $k \rightarrow \infty$ (Bliss and Fisher, 1953; White and Bennetts, 1996). Overdispersion was assessed using the Parametric Bootstrap Likelihood Ratio Test (PBLRT). The standard likelihood ratio test is invalid because the parameter k has a boundary and the resulting test statistic is not χ^2 distributed (Dennis and Taper, 1994). Confidence intervals on all parameter estimates were produced using parametric bootstrap (Efron and Tibshirani, 1993).

3.3.3 The influence of overdispersion on establishment

The probability of at least one individual establishing from the Poisson distribution is defined by γ_P (equation 3.2), and for overdispersed data following a negative binomial distribution, this probability is

$$\gamma_{NB} = 1 - \left(1 + \frac{p_s\phi}{k}\right)^{-k}, \tag{3.5}$$

where the subscripts P and NB are used to identify γ from Poisson and negative binomial distributions, respectively. As overdispersion increases, the values of γ from equations 3.2 and 3.5 diverge although the expectation for both the Poisson and negative binomial distribution is $p_s\phi$ (White and Bennetts, 1996). To emphasize the consequences of increased overdispersion, I evaluate the difference in γ estimates,

$$\Delta\gamma = \gamma_P - \gamma_{NB} = -e^{-p_s\phi} + \left(\frac{k + p_s\phi}{k}\right)^{-k}, \tag{3.6}$$

as a function of the overdispersion (k) and expectation ($p_s\phi$) for the two locations in this study.

3.3.4 Evaluation of waiting time

The waiting time for an invasion, as proposed in Chapter 2, is a geometric distribution (equation 3.3) for a discrete time process with parameter γ. In the absence of overdispersion, γ_P is used for the waiting time distribution. When overdispersion is detected γ_{NB} is used for producing the waiting time distribution.

The data used for estimating the probability of survival can also be used to validate the invasion waiting time. If I treat the cells as time step replications of the establishment process, then the number of zeros encountered until a non-zero observation is made becomes an empirical result of the waiting time distribution. For example, a series of 10 observations used to estimate the survival probability would read n_E = 2,0,0,1,0,1,3,0,1, and would produce waiting time data t = 1,3,2,1,2. Given the predicted waiting time distribution from the estimated γ (equations 3.2 or 3.5, and 3.3) and the empirical waiting time observations (t), Pearson's χ^2 goodness–of–fit tests were conducted on each location (Bain and Engelhardt, 1992).

3.3.5 Permutation invariance

The waiting times realizations, t, depend on the starting position and path taken within the grids. Consequently the conclusions reached about the goodness–of–fit between the observed and predicted waiting times may be sensitive to the starting position and path. In order to eliminate these influences, I conducted 200 permutations of the data for each location. This was archived by creating a vector of the observations, randomizing the order of the observations in the vector, and then reevaluating the waiting time. Because the data are only reorganized, the probability of at least one individual establishing, γ, will remain unchanged. On each permutation, I applied Pearson's χ^2 goodness–of–fit test and recorded the p-value. If fewer than 10 permutations have a p-values of less than 0.05, then I will conclude the goodness–of–fit results for the two locations are permutation invariant and are consequently not sensitive to the initial starting position within the grid nor the path taken through the grid.

3.4 Results

3.4.1 Exploratory data analysis

Two plots within each location (4 total) were destroyed by small mammal activity and were eliminated from the study. Of the 80 control cells at location A, there were no occurrences of scentless chamomile, while location B had one cell with one established adult. Only three plants were found in the surrounding tilled area, and because of the low background levels of scentless chamomile, I hereafter disregard systematic error effects (Regan et al., 2002).

As expected, the low levels of propagule pressure ($\phi = 10$) produced many observations with no establishing individuals (fig. 3.2). The variance-to-mean ratio of location A is 1.135 and 1.28 for location B which suggests some overdispersion in the data.

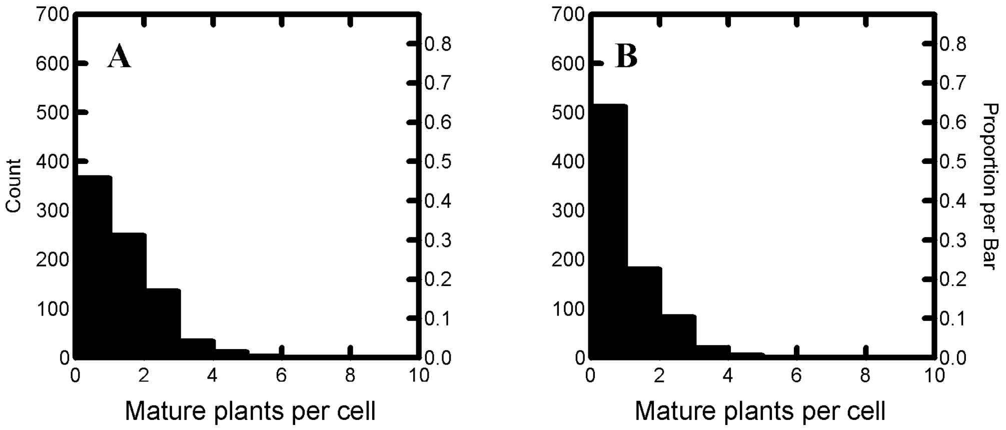


Figure 3.2: Histograms of the number of adult scentless chamomile established. Location A had seeds pressed into the soil and Location B had seeds not pressed into the soil. Both locations had greater than 350 of 800 cells with no adults present.

3.4.2 Parameter estimation and testing for overdispersion

The estimated individual probability of survival for location A, ($\hat{p}_{s,\mathrm{A}}$) was 0.086 (90%C.I: [0.081, 0.091]) with dispersion parameter, $\hat{k}_{\mathrm{A}}$ =5.91 (90%C.I: [3.52, 17.23]). Location B had an estimated individual probability of survival of $\hat{p}_{s,\mathrm{B}} =$ 0.053 (90%C.I: [0.048, 0.058]) with dispersion parameter, $\hat{k}_{\mathrm{B}}$ =1.56 (90%C.I: [1.12, 2.38]). From the PBLRT using 2000 replications to produce the test distribution, I conclude there is significant overdispersion for location A ($PBLRT_{test} =$ 6.78, $PBLRT_{crit} =$ 1.60, $p - value =$ 0.005) and location B ($PBLRT_{test} =$ 27.75, $PBLRT_{crit} =$ 1.49, $p - value <$ 0.001). Additionally from the bootstrapped confidence intervals, I can conclude that locations A and B had significantly different individual probabilities of survival and dispersion.

3.4.3 The influence of overdispersion

The probability of at least one individual surviving, γ, is different for Poisson (γ_P, equation 3.2) and negative binomially (γ_{NB}, equation 3.5) distributed established individuals. The changes in γ as a function of overdispersion (k) for locations A

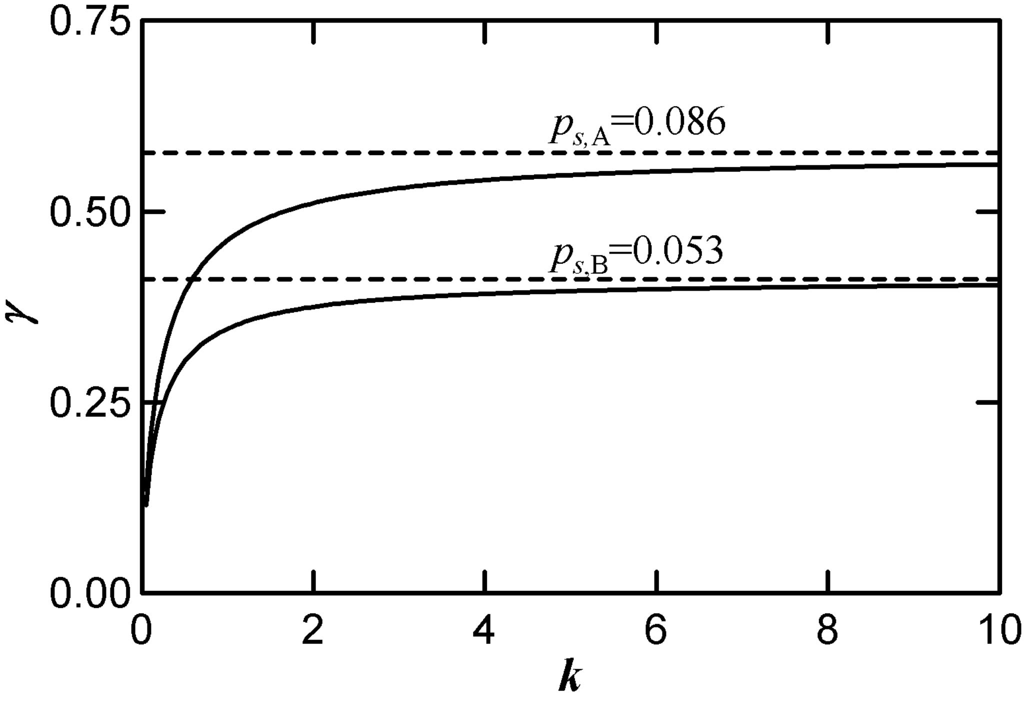


Figure 3.3: As $k \rightarrow \infty$, establishment (N_E) is Poisson distributed. The dashed lines represent γ from the Poisson distribution. The solid lines are the negative binomial estimate of γ as a function of k for locations A ($p_{s,\mathrm{A}} = 0.086$) and location B ($p_{s,\mathrm{B}} = 0.053$). As $k \rightarrow 0$ overdispersion in the establishment of the invader reduces γ.

and B are shown in Figure 3.

The probability of at least one individual establishing, γ, is a constant for the Poisson distribution. As overdispersion increases ($k \rightarrow 0$) γ_{NB} becomes less than γ_P. Thus, failure to account for overdispersion leads to overestimates of γ and consequently increases the right skew in the waiting time distribution. However, $\Delta\gamma \rightarrow 0$ as p_s $\phi \rightarrow 0$. A consequence of this result is that for invaders with a low risk of invasion ($\gamma \rightarrow 0$), accounting for overdispersion may be unnecessary.

3.4.4 Waiting time goodness–of– fit

The probability of at least one establishment from the negative binomial, γ_{NB}, for location A is 0.552, and γ_{NB} for location B is 0.366. The distributions of the spatially replicated observed and expected geometric waiting time distributions for locations A and B are shown in figure 4.

Pearson's χ^2 goodness–of–fit test for location A ($n = 435$, $d.f. = 6$, $p - value = 0.26$) and location B ($n = 280$, $d.f. = 12$, $p - value = 0.15$) failed to detect a significant difference between the empirical observations of the waiting time and the expected geometric distribution. In contrast, the Pearson's χ^2 goodness–of–fit tests following the γ_P for both locations, and that ignores the influence of overdispersion, rejects the geometric distribution ($p - value_A = 0.046$ and $p - value_B < 0.001$).

Of the 200 permutations of the waiting time data in location A, none of the p-values were less than 0.05. Similarly, of the 200 permutations of the waiting time data in location B only two resulted in p-values of less than 0.05. This indicates the starting location in the grids and the path taken through the grid do not change the distribution of the waiting times. I conclude the spatial replication of the establishment process supports the waiting time formulation in Chapter 2.

3.5 Discussion

Preventing invasions requires a pro–active management approach. This will entail predicting the species likely to invade and locations susceptible to invasion so that monitoring and control measures can be implemented before the invader persists in the new environment and impacts ecological and economic interests (Sakai et al., 2001; Parker et al., 1999). Additionally, evaluating the economic costs and benefits to preventing invasions must consider the probability that an invasion will occur (Leung et al., 2002). The invasion waiting time framework provides a means to predict invasions and estimate the probability of establishment based on the survival and propagule pressure (Chapter 2; Jerde and Lewis, 2007), and the invasion waiting time framework is supported by this scentless chamomile study.

The waiting time distribution of scentless chamomile is right skewed (figure 3.4) with a high probability of establishment at the next time step even with a propagule pressure of only $\phi = 10$ individuals. Therefore, pathways that transport even a single flowering head could easily overwhelm a location. The relative ease

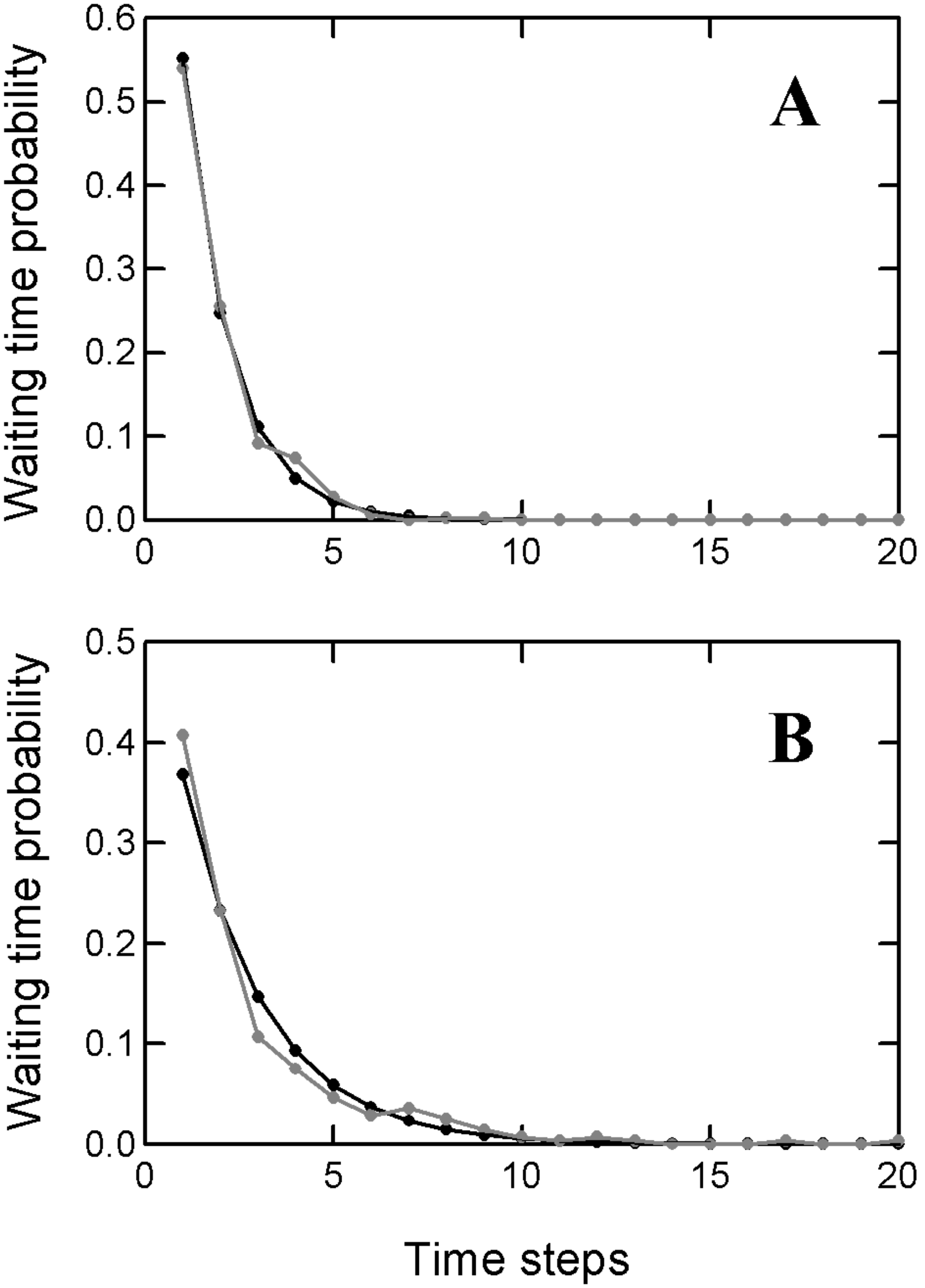


Figure 3.4: Waiting time distribution for location A (Panel A) and location B (Panel B). Gray circles with connecting gray line are the empirical waiting time distribution and the black circles with connecting black line are the predicted waiting time distributions. Pearson's χ^2 goodness–of–fit tests fail to distinguish a significant difference between predicted and observed distributions for both locations.

	Location A			**Location B**		
t	Obs.	*Geo*(γ_A)	Exp.	Obs.	*Geo*(γ_B)	Exp.
1	235	0.551	239.7	114	0.368	103.0
2	111	0.247	107.6	65	0.233	65.1
3	40	0.111	48.3	30	0.147	41.2
4	32	0.050	21.7	21	0.093	26.0
5	12	0.002	9.7	13	0.059	16.4
6	3	0.001	4.4	8	0.037	10.4
7		+0.008	+4.564	10	0.023	6.6
8	1			7	0.015	4.2
9	1			4	0.009	2.6
10				2	0.006	1.7
11				1	0.004	1.0
12				2	0.002	0.7
13				1	+0.004	+1.1
14						
15						
16						
17				1		
18						
19						
20				1		
N,P(χ^2)	435,	0.26		280,	0.15	

Table 3.1: Goodness–of–fit test for geometrically distributed waiting time data of locations A and B. Observed distribution of waiting time (Obs.) of scentless chamomile, *Geo*γ is the geometric expectations, and Exp. is the expected frequencies computed from the geometric distribution. The number of observations (N) and the χ^2 distributed p-value are provided.

by which a location can be overwhelmed is likely why scentless chamomile is a successful invader that has persisted in central Canada for almost a century. Because of the limited effectiveness of biological control and the short distance that scentless chamomile naturally disperses (de Camino-Beck, 2006), focussing control efforts on anthropogenically driven arrival pathways appears to be the best strategy for preventing future invasions.

The spatial replication of the arrival and establishment of scentless chamomile indicated significant levels of overdispersion. The source of overdispersion comes from slightly different expected values for the establishment random variable ($\mathrm{E}[N_E] = p_s\ \phi$) in each cell (Quenouille, 1949), and there are many mechanisms by which overdispersion may occur (Pielou, 1969). Small differences in the probability of survival between cells (p_s) would change the expectation and result in overdispersion. Although propagule pressure (ϕ) was controlled in this experiment, variation in arrival of individuals would also result in overdispersion. Additionally, biotic factors, such as interspecific and intraspecific competition that change p_s may contribute to overdispersion (Skellam, 1952). Biotic and abiotic mechanisms leading to spatial and temporal differences in p_s form many of the hypotheses of invasion biology such as biotic resistance (Elton, 1958), fluctuating resources (Davis et al., 2000), and invasional meltdown (Simberloff and Von Holle, 1999). However, untangling the mechanisms of overdispersion only from observing the number of established individuals is impossible (Pielou, 1969), and requires carefully manipulated experiments with controlled and measurable propagule pressure or survival to infer the mechanisms of invasion (i.e. Von Holle and Simberloff, 2005).

Overdispersion significantly influenced predictions of the invasion waiting time for scentless chamomile. Indeed, overdispersion continues to be the rule rather than the exception for ecological processes that rely on count data (McCullagh and Nelder, 1989; White and Bennetts, 1996). However, it should be noted that with 435 and 280 observations for locations A and B, the statistical power to detect overdispersion is high. Accounting for overdispersion may not always be a necessary consideration for risk assessments of invasive species. For locations that have a high risk of invasion ($p_s\phi >> 0$ which implies γ_P and $\gamma_{NB} \rightarrow 1$), then irrespective of the overdispersion, invasion is almost certain. In contrast, when invasion risk is low ($\gamma_P \rightarrow 0$ and $\gamma_{NB} \rightarrow 0$), only extremely overdispersed observations ($k \rightarrow 0$) can give rise to differences in Poisson and negative binomial estimates of establishment ($\Delta\gamma$, equation 3.6).

Spatial replication, while useful for validating the waiting time framework, is not a replacement for temporal replication. Spatial processes and temporal processes that allow for the successful establishment of invasive species, through the parameters of survival (p_s) and propagule pressure (ϕ), are not the same (Kolasa and Rollo, 1991; Milne, 1991). However the temporal data needed to validate the waiting times would require substantial monitoring effort to overcome errors in detection (Costello and Solow, 2003), and inaction in preventing the invasion while data is collected to validate the waiting time defeats the purpose of a pro–active risk assessment approach to biological invasions (Simberloff, 2003).

Bibliography

Bain, L., and M. Engelhardt. 1992. Introduction to probability and mathematical statistics. Duxbury Press, Belmont, California.

Bliss, C., and R. Fisher. 1953. Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial. Biometrics **9**:176–200.

Bowes, G., D. Spurr, A. Thomas, D. Peschken, and D. Douglas. 1994. Habitats occupied by scentless chamomile *(Matricaria perforata Merat)* in Saskatchewan. Canadian Journal of Plant Science **76**:383–386.

Carlton, J. 1996. Pattern, process, and prediction in marine invasion ecology. Biological Conservation **78**:97–106.

Costello, C., and A. Solow. 2003. On the pattern of discovery of introduced species. Proceedings of the National Academy of Sciences **100**:3321–3323.

Davis, M., J. Grime, and K. Thompson. 2000. Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology **88**:528–534.

de Camino-Beck, T., 2006. Theoretical considerations for biological control: a case study with scentless chamomile. Ph.D. thesis, The University of Alberta, Edmonton, Alberta, Canada.

Dennis, B., and M. Taper. 1994. Denisty dependence in time series observations of natural populations: estimation and testing. Ecological Monographs **64**:205–224.

Douglas, D., A. Thomas, D. Peschken, G. Bowes, and D. Derksen. 1991. Effects of summer and winter annual scentless chamomile (*Matricaria-perforata-Merat*) interference on spring wheat yield. Canadian Journal of Plant Science **71**:841–850.

Efron, B., and R. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall / CRC, Boca Raton, Florida.

Elton, C. 1958. The ecology of invasion by animals and plants. Methuen, London, UK.

Gilpin, M. 1990. Review: ecological prediction. Science **248**:88–89.

Hurlbert, S. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs **77**:187–211.

Jerde, C., and M. Lewis. 2007. Waiting for invasions: a framework for the arrival of non–indigenous species. American Naturalist **170**:1–9.

Kolar, C. 2004. Risk assessment and screening for potentially invasive fishes. New Zealand Journal of Marine and Freshwater Research **38**:391–397.

Kolasa, J., and C. Rollo, 1991. Ecological Heterogeneity, Chapter introduction: The heterogeneity of hterogeneity:a glossary, pages 1–23 . Springer–Verlag.

Leung, B., D. Lodge, D. Finnoff, J. Shogren, M. Lewis, and G. Lamberti. 2002. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proceedings of the Royal Society of London, B **269**:2407–2413.

McClay, A., 2003. Field releases and evaluation of biological control agents for scentless chamomile. Final Report 99M554, Alberta Agriculture Research Institute, Vegreville, Alberta: Alberta Research Council.

McCullagh, P., and J. Nelder. 1989. Generalized linear models. 2nd edition. Chapman and Hall, London, U.K.

Milne, B., 1991. Ecological Heterogeneity, Chapter heterogeneity as a multiscale characteristic of landscapes, pages 70–84 . Springer–Verlag.

Parker, I., D. Simberloff, W. Lonsdale, K. Goodell, M. Wonham, P. Kareiva, M. Williamson, B. Von Holle, P. Moyle, J. Byers, and L. Goldwasser. 1999. Impact: Toward a framework for understanding the ecological effects of invaders. Biological Invasions **1**:3–19.

Pielou, E. 1969. An introduction to mathematical ecology. John Wiley and Sons, Inc., New York.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien–invasive species in the United States. Ecological Economics **52**:273–288.

Quenouille, M. 1949. A relation between the logarithmic, Poisson, and negative binomial series. Biometrics **5**:162–164.

Regan, H., M. Colyvan, and M. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications **12**:618–628.

Sakai, A., F. Allendorf, J. Holt, D. Lodge, J. Molofsky, K. With, S. Baughman, R. Cabin, J. Cohen, N. Ellstrand, D. McCauley, P. O'Neil, I. Parker, J. Thompson, and S. Weller. 2001. The population biology of invasive species. Annual Reviews of Ecology and Systematics **32**:305–32.

Simberloff, D. 2003. How much information on population biology is needed to manage introduced species? Conservation Biology **17**:83–92.

Simberloff, D., and B. Von Holle. 1999. Positive interactions of nonindigenous species: Invasional meltdown? Biological Invasions **1**:21–32.

Skellam, J. 1952. Studies in statistical ecology. Biometrika **39**:346–362.

Vermeij, G. 1996. An agenda for invasion biology. Biological Conservation **78**:3–9.

Von Holle, B. 2005. Biotic resistance to invader establishment of a southern Appalachian plant community is determined by environmental conditions. Journal of Ecology **93**:16–26.

Von Holle, B., and D. Simberloff. 2005. Ecological resistance to biological invasion overwhelmed by propagule pressure. Ecology **86**:3212–3218.

White, G., and R. Bennetts. 1996. Analysis of frequency count data using the negative binomial distribution. Ecology **77**:2546–2557.

Williamson, M. 1996. Biological Invasions. Chapman and Hall, London, United Kingdom.

Woo, S., A. Thomas, D. Peschken, G. Bowes, D. Douglas, V. Harms, and A. McClay. 1991. The biology of Canadian weeds. 99. Matricaria perforata Merat (Asteraceae). Canadian Journal of Plant Science **71**:1101–1119.

Chapter 4

Uncertain invasions: A biological perspective

[1]

4.1 Introduction

In the absence of perfect knowledge about how a system or process works, there is unexplained variability in observations and predictions, or uncertainty. Uncertainty is the reason biologists experience difficulty identifying the properties of successful invaders and predicting invasions (Kolar and Lodge, 2001). Prior to and following Elton's treatise (Elton, 1958), invasion biology received mainly disparate scientific interest, and has only recently gained general ecological popularity (Davis et al., 2001; Puth and Post, 2005) because of threats to biodiversity and financial losses (Pimentel et al., 2000, 2005). Now scientists are faced with the challenge of predicting and ideally preventing invasions without fully understanding the mechanisms that lead to successful invasion. Some have argued that predicting invasions in light of such uncertainty is futile (Gilpin, 1990). Nevertheless, there has been considerable progress in predictive methods to identify species likely to invaded and the locations likely to become invaded, which have emerged from the study of biological invasions (Schneider et al., 1998; Kolar and Lodge, 2001;

---

[1] This chapter appears as: Jerde, C. and J. Bossenbroek. Uncertain Invasions: A biological perspective . In Bioeconomics of Invasive Species: Integrating Ecology, Economics, Policy, and Management. R. Keller, M. Lewis, D. Lodge, and J. Shogren Editors. To appear 2008. Peer reviewed.

Rouget et al., 2004) and the uncertainty of invasions has been reduced.

Uncertainty is an interdisciplinary subject and consequently has varied usage. In economics, uncertainty is regularly mentioned in the same context of risk, where risk is a product of the probability of an event occurring and the impact of that event (Knight, 1921). As long as the probability of the event is not zero or one, there is uncertainty. Biologists generally associate uncertainty with unexplained processes and mechanisms by which variability in observations arise (Taper and Lele, 2004). This uncertainty is often presented as confidence intervals, standard errors, or posterior distributions on model parameters (Lewin-Koh et al., 2004; Cumming et al., 2007). However, there is considerable overlap between these two generalized perspectives of uncertainty (e.g. Dovers and Norton, 1996). With an interdisciplinary approach to biological invasions, it is therefore unsurprising to find multiple perspectives regarding the importance and role of uncertainty. Because of the breadth of perspectives regarding uncertainty, I provide, in this chapter, a biological perspective of uncertainty and invasions. A complementary bioeconomic perspective of biological invasions, risk, and uncertainty can be found in Shogren and Finnoff (In Press).

My biological perspective of uncertainty begins by reviewing a general taxonomy of uncertainty (Regan et al., 2002) and applying this taxonomy to biological invasions. My list of sources of uncertainty is not comprehensive, but I believe it highlights some areas of invasion biology that, if emphasized in future studies, will strengthen explanatory and predictive capabilities. I then focus my study of uncertainty and biological invasions to deconstructing a national gravity model of zebra mussel invasion. Specifically, I demonstrate a bootstrapping method of survey data used to parameterize the national gravity model and assess the influence of uncertainty on the risk of invasion for two environmentally and economically valuable lakes, Lake Mead (AZ and NV, USA) and Lake Roosevelt (WA, USA). I then evaluate the risk of invasion for 13 uninvaded lakes and evaluate the predictive performance of the gravity model on 15 lakes recently invaded. In closing, I discuss the challenges of prediction and validation of predictions in the face of uncertain invasions.

4.2 Sources of uncertainty

Even in the discipline of biology there are multiple perspectives on sources of uncertainty (e.g. Shaffer, 1981; Burnham and Anderson, 2002). Many of these treatments focus on a particular type of uncertainty (e.g. measurement error: Thomas et al., 1993) or a particular methodology of assessing uncertainty (e.g. Harwood and Stokes, 2003; Cumming et al., 2007). However, a broad perspective of classifying uncertainty was taken by Regan et al. (2002) and is generally consistent with biological and statistical perspectives of sources of uncertainty. I adopt this classification for its completeness as a taxonomy on which to discuss the uncertainty related to biological invasions.

4.2.1 Linguistic uncertainty

The general classification system starts by separating uncertainty into linguistic and epistemic uncertainty (Regan et al., 2002). Linguistic uncertainty is associated with communicating ideas and definitions. Chapter 6 contains a discussion about the linguistic uncertainty associated with defining when an invader is said to have established. Establishment is defined both by an abundance of the invader and a persistence of the invader in a new landscape. In theoretical treatments of the invasion process, establishment is rarely defined with precise, numerical thresholds, but for empirical applications, these thresholds are required to define whether a species has established in a system and are necessary to perform statistical analyses. The variability in thresholds used to define establishment represents one form of linguistic uncertainty.

Within invasion biology, linguistic uncertainty has received some attention. Richardson et al. (2000) discuss the vagueness of the terms ‘naturalized’ and ‘invasive.’ Additionally they attempt to provide a clear vocabulary for discussing the invasion process. The vagueness, context dependence, and ambiguity (Regan et al., 2002) of invasion biology terms is pervasive throughout the entire invasion process from transport and arrival of invaders to the establishment, spread, and impact of those invasions (Colautti and MacIsaac, 2004; Sakai et al., 2001).

Another example of linguistic uncertainty is in the assessment of whether an invasive species has had an impact. Invasion biologists have variably defined impact as the presence of any non-indigenous species (NIS), a NIS that has spread, and a NIS that has produced harmful environmental changes, in particular to native species (Ricciardi and Cohen, 2007). Although attempts have been made to provide clear working definitions, such as the Parker et al. (1999) formulation stating impact is the product of species range (area), density, and biomass. These formulations may be entirely appropriate for considering impact in an ecological context, but this ignores the economic impact.

The simplest solution to linguistic uncertainty is to provide precise definitions that can be agreed upon by the scientific community (Richardson et al., 2000; Regan et al., 2002). However, finding consistent terminology has been, and continues to be, a problem for exploring the patterns and processes of biological invasion (Shrader-Frechette, 2001) and also for communicating invasion biology and risk to the public and policy makers (Bossenbroek et al., 2005). The evolution of invasion biology into an objective discipline will likely be tied to the preciseness of the definitions employed in its description and application (Colautti and MacIsaac, 2004). Another solution to the linguistic uncertainty of invasion biology is to provide syntheses of work on a particular subject, such as propagule pressure (Lockwood et al., 2005), that makes connections across multiple definitions and inferences.

4.2.2 Epistemic uncertainty

Epistemic uncertainty is more closely related to data, models, and the methods of scientific inquiry, which can be secondarily separated into uncertainty due to measurement error, systematic error, natural variation, inherent randomness, and subjective judgment (Regan et al., 2002). Epistemic uncertainty is often accounted for by reporting quantitative measures such as confidence intervals, prediction intervals, probability distributions, or p-values.

Measurement error and systematic error are associated with errors in the recording of data by either human observation or errors from measurement devices, where measurement error is an unbiased measurement and systematic error is a biased measurement (Thomas et al., 1993). Natural variation is uncertainty due to spatial or temporal differences in the values of model parameters (e.g. Lele et al., 1998). Some ecological processes are inherently (effectively) random because it is infeasible to account for all the processes that influence model outcomes, nor is it possible to know the initial conditions with such precision as to make deterministic predictions (Gilpin, 1990). Model uncertainty occurs when a biological process is described using a mathematical representation (Jonzen et al., 2002). Lastly, subjective judgment emerges as a form of uncertainty throughout scientific inquiry, from evaluating the quality of data and choosing a modeling approach, to interpreting results and making decisions (Harwood and Stokes, 2003). Table 1 provides the taxonomy of epistemic uncertainty with examples from biological invasions.

Many of the epistemic uncertainty issues in invasion biology are common problems in ecology. For example, some species are able to invade at low population densities and remain undetected at low densities only later to experience population growth and spread (Christian and Wilson, 1999). Detecting species at low abundance is the same problem community ecologists face when attempting to detect the presence of rare species (Longino and Colwell, 1997; Costello and Solow, 2003). Similarly in conservation biology, demographic stochasticity and minimum viable population size (Lande, 1993) are directly related to propagule pressure in invasion biology (Lockwood et al., 2005; Drake and Lodge, 2006).

Invasion biology, however, faces some unique sources of uncertainty. In community ecology, the rare species are usually known to occur within the area being searched and the observer has some search recognition pattern. This is often not true in invasion biology, where many invaders are surprise discoveries (Solow and Costello, 2004). Similarly, population parameters such as growth rate, survival, fecundity, and reproductive value may be known for a species in its native range, but unless the destination has very similar habitat characteristics, these parameters will be different and will remain unknown until the invasion has occurred and population data collected. As a consequence and in part to sidestep added uncertainty due to

Type of uncertainty	Mechanism
Measurement error (ME)	Estimates of propagule pressure (species abundance) Estimates of boater registrations Estimates of lake attractiveness
Systematic error (SE)	Measurement of failed invasions Detection of invasive species Ignorance of important vectors and pathways of introduction
Natural variability (NV)	Fluctuations in donor region populations Changes in pathways of introduction Changes in recipient region environment Seasonality of a species life cycle
Model uncertainty (MU)	Presence of Allee effects Population growth models Choosing parameters in an ecological niche model Fitness change as a function of distance traveled
Inherent randomness (IR)	Demographic stochasticity of introduction and survival Genetic bottlenecks due to small founder populations Predator avoidance in new locations
Subjective judgment (SJ)	Choice of species to study Including expert and public opinion into risk management and policy Use of survey data designed for other purposes to validate model

Table 4.1: The uncertainty of biological invasions. Biological invasions have a mixture of common ecological and unique sources of uncertainty. Each source of uncertainty could potentially hinder the ability to accurately predict successful invasions. Some examples are specific to dispersal of NIS by recreation boaters.

population dynamics, there has been a call to develop robust methods and models of invasive species risk assessment that do not require specific details regarding population dynamics (Simberloff, 2003).

One solution to both epistemic and linguistic uncertainty is through mathematical modeling of biological (Taper and Lele, 2004) and economic processes (Leung et al., 2002). For epistemic uncertainty, the modeling solution is straight forward. Models represent hypotheses about how a system or process works. Confronting models with data allows us to perform hypothesis testing and model selection among competing hypotheses (Hilborn and Mangel, 1997; Burnham and Anderson, 2002; Lewin-Koh et al., 2004). Alternatively, models of these process can be evaluated on their predictive performance. Both strategies require accounting for uncertainty, that is quantifying the explanatory (e.g. goodness–of–fit) and predictive (e.g. Receiver Operating Characteristic (ROC) curves) capabilities (Hosmer and Lemeshow, 2000). Methods to assess the explanatory and predictive performance of models are demonstrated in the next section.

Less obvious is the role mathematical modeling plays in reducing linguistic uncertainty. When theory and experimentation meet, there is necessarily a measurable quantity to evaluate from a model that is evaluated with data (hypothesis testing or model selection: Lewin-Koh et al., 2004). Returning to the example of defining establishment (Chapter 6), empiricists must specify thresholds of abundance and/or persistence that above which establishment is said to have occurred. It is possible, although not likely, that invasion biologists, mathematical modelers, and economists could reach a consilience regarding the threshold of establishment. More likely, however, is that we will continue to have a variety of thresholds. Arguably, the most common thresholds will be determined by the data available (e.g. presence or absence versus count data) and the models used (Chapter 6) to assess the process of establishment. Although some linguistic uncertainty will remain depending on the data collected and modeling approach, mathematical models will force clearly delineated (at least mathematically) definitions of establishment that theory likely would not evolve to if left in a conceptual form.

I now transition to an example of estimating the risk of invasion and accounting for uncertainty by modeling the transport of zebra mussels through the use of a gravity model produced by Bossenbroek et al. (2007). The purpose of this specific example is to demonstrate how uncertainty in explanation and prediction can be accounted for. While gravity models are regularly applied to aquatic invasions, and in particular to zebra mussels, previous studies have only haphazardly quantified uncertainty (but see Bossenbroek et al., 2001; Leung et al., 2006). Here I show a bootstrapping routine that accounts for uncertainty in parameter estimates used in the gravity model and I evaluate the predictive power of the gravity model on the order of lake invasions.

4.3 Zebra Mussels and a gravity model of arrival

The dreissenid mussel invasion of the Great Lakes began about 1986, with the successful establishment of the zebra mussel, *D. polymorpha*, in Lake St. Clair from ships' ballast water (Hebert et al., 1989), causing extensive ecological and economic impacts. The present North American range of *D. polymorpha* includes much of northeastern and north central North America, including over 400 inland lakes. A second species of dreissenid, the quagga mussel (*D. bugensis*), was found in the Erie Canal and Lake Ontario in 1991 (May and Marsden, 1992), and is now common in Lakes Erie and Ontario (Diggins et al., 2004). This species was also discovered in Lake Mead on the Colorado River in January 2007.

The range expansion of dreissenid mussels in North America to date has resulted from a combination of processes, involving the dispersal within and between water bodies. The primary pathways of dreissenid dispersal include shipping routes in the United States, natural downstream dispersal, and overland dispersal by human vectors, such as recreational boaters. Gravity models of recreational boater movement patterns have been used to forecast the overland dispersal of *D. polymorpha* (Schneider et al., 1998; Bossenbroek et al., 2001; Leung et al., 2006; Bossenbroek et al., 2007).

Symbol	Description	Examples of uncertainty
Gravity model input		
O_i	Number of boaters at the source	Temporal changes in the number of boaters (NV) Only licensed boaters (SE)
W_j	Attractiveness	Alternative measure to area (SJ,MU)
$c_{i,j}$	Distance	Euclidean versus travel distance (SJ, MU)
δ α	Distance multiplier Distance Coefficient	Change in gas prices (RV) Selective Sampling (SE) Sampling Error (ME)
Gravity model output		
Q	Number of boaters entering a destination with NIS	Functional relationship between Q and the probability of establishment (ME)

Table 4.2: The uncertainty of gravity models. The parameters α and δ are subject to uncertainty common to estimation methods and sampling. The variables (O, W, c, and Q) are also subject to uncertainty. One concern is whether attractiveness is rightly associated with the area of the lake. The abbreviations of the type of uncertainty (ME, SE, NV, MU, IR, and SJ) are defined in Table 1.

4.3.1 Gravity model formulation

Gravity models use formulations analogous to Newton's laws of attraction, to estimate the arrival of invaders into discrete patches. The variables and parameters susceptible to epistemic uncertainty are listed in Table 2 and an overview of the mathematics of gravity models can be found in Bossenbroek et al. (2001) or Leung et al. (2006). The formulation used here follows the national gravity model for zebra mussel dispersal (Bossenbroek et al., 2007). The critical result useful for predicting invasions from the gravity model is the number of arriving boaters that previously visited zebra mussel infested waters (Q)[2].

Estimating the number of arriving boaters starts by first modeling the distribution of boaters from source watershed, i, to destination, j, between N watersheds,

$$T_{i,j} = \frac{O_i \, W_j \, c_{ij}^{-\alpha}}{\sum_{j=1}^{n} W_j \, c_{ij}^{-\alpha}}, \quad \text{for all } i, \tag{4.1}$$

where O_i is the number of licensed boaters at the source watershed, W_j is the area of lakes in the destination watershed, c_{ij} is the Euclidean distance between the source and destination watersheds, and $c_{ii} = \delta \min_{j \neq i}(c_{ij})$. The sum in the denominator is a balancing factor that ensures all boaters that leave a source arrive at a destination. The national gravity model for zebra mussels (Bossenbroek et al., 2007) has two notable differences from previous zebra mussel studies (Bossenbroek et al., 2001; Leung et al., 2006). First, the locations are delineated by watersheds (Figure 1) rather than counties, and second, it is possible for boaters to redistribute themselves within the same watershed ($i = j$). This leads to an added parameter δ that is the distance traveled within a watershed as a proportion of the distance to the next nearest possible destination. The parameter α is a distance coefficient that describes the deterrent effect of distance upon a boater. The variables O, W, and c are properties of each watershed while the parameters α and δ are estimated from data.

Survey data for the 100th Meridian Initiative were used to parameterize the

---

[2] Q in Bossenbroek et al. (2007) is a proportion of boaters traveling. Here Q is the number of infested boaters arriving at a destination

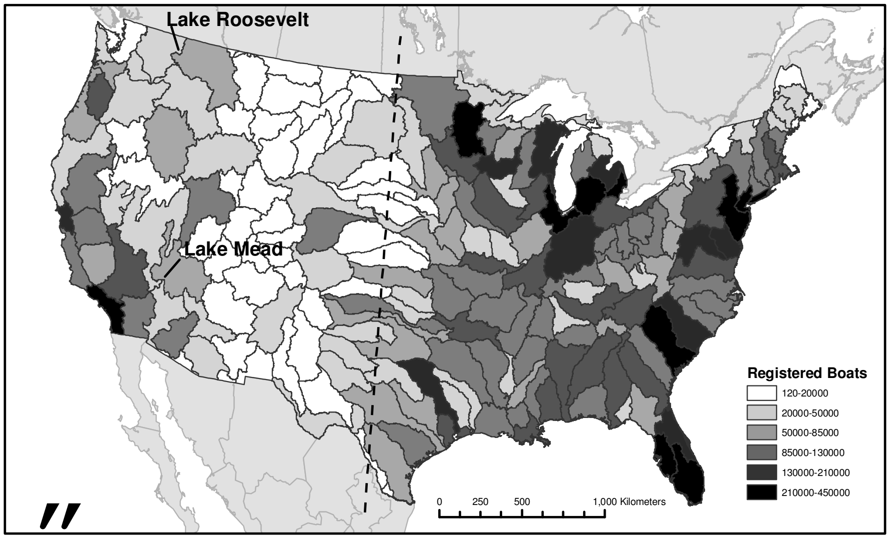


Figure 4.1: Reservoirs and watersheds of the United States. Lakes Roosevelt and Mead receive boater traffic from watersheds East of the 100th meridian (dashed line) that have zebra mussels present. These lakes have been the focus of managers as likely locations for zebra mussels to be introduced in the Western U.S. Each watershed is gray-scale colored by the number of licensed boats found within the watershed. The Great Lakes region, which has the highest density of zebra mussel invaded waters, also has a high density of licensed recreational boaters. This figure is adapted from Bossenbroek et al. (2007)

national gravity model. The 100th Meridian Initiative is a cooperative effort by state, federal, and provincial agencies to prevent the spread of zebra mussels and other aquatic nuisance species into western North America. Surveys were conducted at 20 reservoirs throughout the Great Plains region (i.e., roughly along the 100th Meridian) that recorded the distance traveled by recreational boaters between sources and destinations. The 13 reservoirs that had more than 50 completed surveys were used for the parameterization and is consistent with the analysis performed in Bossenbroek et al. (2007).

In the national gravity model, the number of boaters carrying zebra mussels (Z_i) is assumed to be proportional to the area of zebra mussel infested lakes found within watershed i. This results in the redistribution of infested boaters,

$$R_{ij} = \frac{Z_i}{O_i} T_{ij}, \tag{4.2}$$

where $\frac{Z_i}{O_i}$ is the proportion of boater carrying zebra mussels. The number of infested boaters arriving to a watershed j is

$$Q_j = \sum_{i=1}^{n} R_{ij}. \tag{4.3}$$

The number of boaters visiting a watershed that previously visited watersheds with invaded lakes (Q) is purportedly an indicator of invasion probability of establishment, and can be interpreted as the dose in a dose-response application of biological invasions (Chapter 6; Bossenbroek et al., 2001; Leung et al., 2004; Lockwood et al., 2005).

The 100th Meridian Initiative was initially established to stop or slow the spread of zebra mussels into the western U.S., but the funding was grossly inadequate relative to what would be required to meet this goal. No guidance currently exists about how best to allocate funding to alternative methods of prevention and control. The national gravity model was, in part, developed to inform this management dilemma (Bossenbroek et al., 2007). I focus my analysis, hereafter, on accounting for the uncertainty in the gravity model, as this is a necessary component for implementing management actions (Leung et al., 2002). How the measured uncertainty, dealt with here, feeds into economic considerations and resulting management decisions is more thoroughly treated in Shogren and Finnoff (In Press). I begin my treatment of uncertainty in gravity models by considering the invasion of two popular destinations for boaters across the United States, Lakes Mead and Roosevelt.

4.3.2 Order of invasion

Which of these two lakes is most likely to be invaded first? Jerde and Lewis (Chapter 2; 2007), using invasion waiting times, formulated the order of invasion for two locations. This ordering, applied to outputs from gravity models, requires three assumptions: the survival of individuals is small, the survival at both locations is approximately the same, and the gravity scores (Q_m: Lake Mead and Q_r: Lake Roosevelt) are proportional to the actual propagule pressure. Under these conditions, the probability of Lake Mead being invaded before Lake Roosevelt is $(1+\frac{Q_r}{Q_m})^{-1}$, and the probability of Lake Roosevelt being invaded before Lake Mead is the complement, $(1+\frac{Q_m}{Q_r})^{-1}$.

Based on (Bossenbroek et al., 2007), the probability of Lake Mead being invaded before Lake Roosevelt is 0.797 and the probability of Lake Roosevelt being invaded before Lake Mead is 0.203. These probabilities are based on the gravity model structure, the measured variables, the estimated parameters, and the model assumptions – all of which are subject to the influences of uncertainty. I now turn to quantifying the uncertainty in the parameters α and δ that are estimated from the survey data regarding the distance boaters traveled to get to reservoirs in the Midwest of the United States.

4.3.3 Bootstrapping confidence intervals

To investigate the change in the probabilities of ordered invasion, due to uncertainty in the estimates of α and δ, I performed a Monte Carlo simulation by:

1. For a single reservoir, I estimated the probability an arriving boater would be from a specific watershed using the survey data used to parameterize the gravity model

2. I then drew from a multinomial distribution with these probabilities as parameters and recorded the distance this boater traveled to arrive at the destination (c_{ij})

3. This was repeated at a single reservoir for the number of surveys recorded at that reservoir

4. Steps 1-3 were then repeated for each reservoir

5. From steps 1-4, the parameters α and δ where estimated following Bossenbroek et al. (2007), and recorded

6. Q_m, Q_r, and the estimated probabilities of ordered invasion were recorded

7. Lastly, steps 2-6 were repeated 1000 times

This procedure is a bootstrapping routine that accounts for the uncertainty in and

Parameter	Lower 95% C.I.	Upper 95% C.I.
α	2.49	2.63
δ	0.70	1.35
Pr(Mead before Roosevelt)	0.795	0.799
Pr(Roosevelt before Mead)	0.201	0.205

Table 4.3: Confidence intervals on parameters and probability estimate. There appears to be very little variation in the probability estimates due to variability in the parameters estimated from the survey data. However the parameter estimates may have some temporal variability from changes in driving habits, such as increased gas prices, that are not reflected in the confidence intervals.

from the survey samples. The uncertainty arises not from the number of boaters arriving to the reservoir but from the variability in the contribution of observed sources to the estimates. The list of replicates resulting from the bootstrapping routine can be used to build 95% confidence intervals on α, δ, and the resulting uncertainty in the ordered probabilities of invasion (Efron and Tibshirani, 1993).

The relatively small range of the confidence intervals on the parameters and small change in the probabilities of ordered invasion indicates the uncertainty in the boater surveys due to variability in the source of boater movements is negligible. The distance coefficient (α) estimate has fairly tight confidence intervals, while the distance multiplier (δ) has broader confidence intervals (Table 3). This is expected after inspecting the sums-of-squares surface provided in Figure 1 of Bossenbroek et al. (2007). A subsequent sensitivity analysis, performed in Bossenbroek et al. (2007), indicated that a 25% reduction in α could lead to approximately an 8% decrease in the proportion of boats arriving to a location. However, as demonstrated here, this does little to change the predicted ordered probabilities of invasion. Taken together, the reduction in α would likely increase the expected invasion waiting time for both lakes but does not change the order in which the invasion would likely occur (Chapter 2; Jerde and Lewis, 2007).

Bootstrapped confidence intervals account for uncertainty in the parameters that can be quantified from the survey data and this encompasses many of the common sources of uncertainty surrounding surveys, such as sample size, randomness, and completeness (Barnett, 2002). Bootstrapping does not account for any bias, such as interviewing boaters with only out-of-state plates, nor does this bootstrapping account for any temporal variability in the values of α or δ.

4.3.4 Gravity models, evaluating variables, and model selection

Uncertainty may also influence the variables in a model (such as the gravity variables O, W, and c) and the model structure. Assessing model structure and the inclusion or exclusion of different variables is generally referred to as model selection. In statistical practice, there are multiple ways to perform model selection such as likelihood ratio tests and Akaike's Information Criterion (AIC: Burnham and Anderson, 2002). But these methods have yet to be applied to invasion gravity models because the data are insufficient and, due to the expenses involved in continuously monitoring boater traffic at multiple uninvaded locations, unlikely to ever be collected.

As an alternative to applying a model selection approach on Q, the estimated number of boaters arriving is used as an explanatory variable and then the probability of establishment is modeled from presence or absence of an invasion into a lake. This may be accomplished using logistic regression or a functional form of the response curve (Bossenbroek et al., 2001; Leung et al., 2004, 2006). Statistically, this approach is problematic because the explanatory variable, in this case Q, is usually assumed fixed and known (Hosmer and Lemeshow, 2000), but here the number of arriving boaters is an estimate from the gravity model and therefore contains some variability or uncertainty. Specifically, there is measurement error and/or systematic error in the explanatory variable. Measurement error can change the observed mean structure, the variance structure, and obscure significant covariates (Thomas et al., 1993). This in turn will lead to poor predictions and model–fit from a dose-response curve. Proper model selection using this approach would require accounting for the measurement error. Assessing the gravity model structure and accounting for measurement error in gravity scores are important future directions for invasion biology research.

4.4 Uncertainty and prediction

Predicting establishment with accuracy is an aspiration of invasion biology (Kolar and Lodge, 2001), and of gravity models in particular (Bossenbroek et al., 2001; Leung et al., 2006). Predictive accuracy is critical for proactive management to prevent invasions (Stohlgren and Schnase, 2006). However, it is still unclear if models of species arrival and spread developed thus far are sufficient to offer reliable guidance about when and where to intervene against invasive species (Gilpin, 1990). I demonstrate three related predictive insights related to gravity models. First I look at the pair–wise probabilities of 13 uninvaded lakes. Second, I estimate the pair–wise probabilities of 15 recently invaded lakes and show graphically the relationship between Q and the pair–wise probabilities of invasion given the distribution of Q from the invaded lakes. And third, I apply logistic regression, knowing there is uncertainty in the explanatory variable Q, to demonstrate how uncertainty confounds predictive performance.

4.4.1 Making predictions and gravity models

Gravity scores, Q, are known to be positively correlated with successful invasions (MacIsaac et al., 2004). Moreover, invasion order probability, as demonstrated for Lakes Mead and Roosevelt, can be calculated. For the 13 lakes of interest presented in Bossenbroek et al. (2007), the pair–wise probabilities are provided in Table 4. All lakes in the table were uninvaded by zebra mussels at the onset of this project. Since then, however, Lake Mead has been invaded with quagga mussels and the Lake of the Ozarks, just downstream from H.S. Truman Reservoir, was reported to contain zebra mussels in June 2006.

As apparent from gravity score, Q, H.S. Truman Reservoir is the most likely to become invaded, and each paired probability is greater than 0.75. This later result indicates that I should not expect many, if any, lakes in this group to become invaded before H.S. Truman Reservoir. Analytically, the probability of r lakes becoming invaded before some time, t, while H.S. Truman Reservoir remains uninvaded is,

	Q	H.S. Truman Reservoir	Lake Oahe	Barren River Lake	Martin Lake	Austin Lake	Lake Mead	Upper Saranac Lake	Perry Lake	Chesuncook Lake	Roosevelt Lake	Amistad Reservoir	Goose Lake	Conchas Lake
H.S. Truman Reservoir	3145.4	–	0.78	0.78	0.82	0.84	0.90	0.91	0.96	0.96	0.97	0.99	0.99	0.99
Lake Oahe	899.1	0.22	–	0.51	0.57	0.60	0.72	0.74	0.87	0.88	0.91	0.95	0.96	0.98
Barren River Lake	876.5	0.22	0.49	–	0.56	0.59	0.72	0.74	0.87	0.88	0.90	0.95	0.95	0.98
Martin Lake	691.1	0.18	0.44	0.44	–	0.53	0.67	0.69	0.83	0.85	0.89	0.94	0.94	0.98
Austin Lake	602.2	0.16	0.40	0.41	0.47	–	0.64	0.66	0.81	0.83	0.87	0.93	0.94	0.97
Lake Mead	342.6	0.10	0.28	0.28	0.33	0.36	–	0.53	0.71	0.74	0.80	0.89	0.89	0.95
Upper Saranac Lake	309.6	0.09	0.26	0.26	0.31	0.34	0.48	–	0.69	0.72	0.78	0.88	0.88	0.95
Perry Lake	137.4	0.04	0.13	0.14	0.17	0.19	0.29	0.30	–	0.53	0.61	0.76	0.77	0.89
Chesuncook Lake	120.9	0.04	0.12	0.12	0.15	0.17	0.26	0.28	0.47	–	0.58	0.73	0.74	0.88
Roosevelt Lake	89.3	0.03	0.09	0.09	0.11	0.13	0.20	0.22	0.40	0.430	–	0.67	0.68	0.84
Amistad Reservoir	44.0	0.01	0.05	0.05	0.06	0.07	0.11	0.12	0.24	0.27	0.33	–	0.51	0.72
Goose Lake	42.0	0.01	0.05	0.05	0.05	0.07	0.11	0.12	0.23	0.26	0.32	0.49	–	0.72
Conchas Lake	16.8	0.01	0.02	0.02	0.02	0.03	0.05	0.05	0.10	0.12	0.16	0.28	0.29	–

Table 4.4: Pair-wise order of invasion for currently uninvaded lakes. Each lake has a probability, estimated from the gravity score (Q), of being invaded before a different lake. For example the probability of H.S. Truman Reservoir being invaded before Lake Oahe is 0.778. The complement, the probability of Lake Oahe being invaded before H.S. Truman Reservoir is 0.222. Probabilities near 0.5 indicate lake pairs with similar invasion risk based on the gravity score. For example, Lake Mead and Upper Saranac Lake have similar gravity scores (Q=342.6 and 309.6 respectively). These lakes are the largest lakes within their watershed and are considered the most likely to become invade.

$$\underbrace{\Pr(R=r)}_{\substack{\text{Probability of}\\ \text{r lakes}\\ \text{invaded}}} \;=\; \underbrace{\binom{N}{r}}_{\substack{\text{No. of}\\ \text{ways}\\ \text{r arises}}} \; \underbrace{(1-(1-p_u)^t)^r}_{\substack{\text{Probability of}\\ \text{r invaded}\\ \text{lakes}}} \; \underbrace{((1-p_u)^t)^{N-r}}_{\substack{\text{Probability of}\\ \text{N}-\text{r uninvaded}\\ \text{lakes}}} \; \underbrace{(1-p_i)^t}_{\substack{\text{Probability of}\\ \text{H.S.Truman}\\ \text{uninvaded}}} \tag{4.4}$$

where there are N lakes in the group of uninvaded lakes (Table 3; N=12 when excluding H.S. Truman Reservoir). The details of this formulation are found in Appendix C. The parameter p_u is the probability at each discrete time step that an uninvaded lake transitions to become invaded, and p_i is the probability that H.S. Truman Reservoir becomes invaded. This formulation rests on the assumptions that each lake, in the group, has the same p_u, and that both p_i and p_u are known. Unfortunately neither of these assumptions are easily justified or met for the gravity scores presented in Table 3. I know the gravity scores, Q, are very different between the group of uninvaded lakes and this can reflect different probabilities of invasion (Von Holle and Simberloff, 2005, Table 3). The pair-wise probabilities (Table 3) of Lake Oahe, Barren River Lake, Martin Lake, and Austin Lake all have order pairings close to 0.5 implying that either lake could be invaded before the other, and these lakes form a group with similar likelihoods of invasion. Given a larger sample than just four lakes with similar pairing scores may be more useful for producing the group of reference lakes.

For an example, I could assume that the group of uninvaded lakes were similar to Lake Oahe, the next most likely invaded lake based on the gravity scores. Then using the relative probability formulation, presented in Jerde and Lewis (Chapter 2; 2007), the relative probability of a H.S. Truman Reservoir transitioning to invaded is $p_i = \frac{Q_{H.S.Truman}}{Q_{Oahe}}\, p_u = \frac{3145.4}{899.1}\, p_u = 3.5\, p_u$. This can be inserted into the equation and I am left with one parameter to estimate, p_u. Alternatively, I can insert $\frac{899.1}{3145.4}\, p_i$ for p_u and then I am similarly left with p_i to estimate. However, neither of these parameters are known.

Figure 4.2A is a plot of the probability of observing one or more of the 12 reference lakes becoming invaded (given $p_u = \frac{899.1}{3145.4}\, p_i$) as a function of probability that H.S. Truman Reservoir becomes invaded, and Figure 4.2B is a plot of the

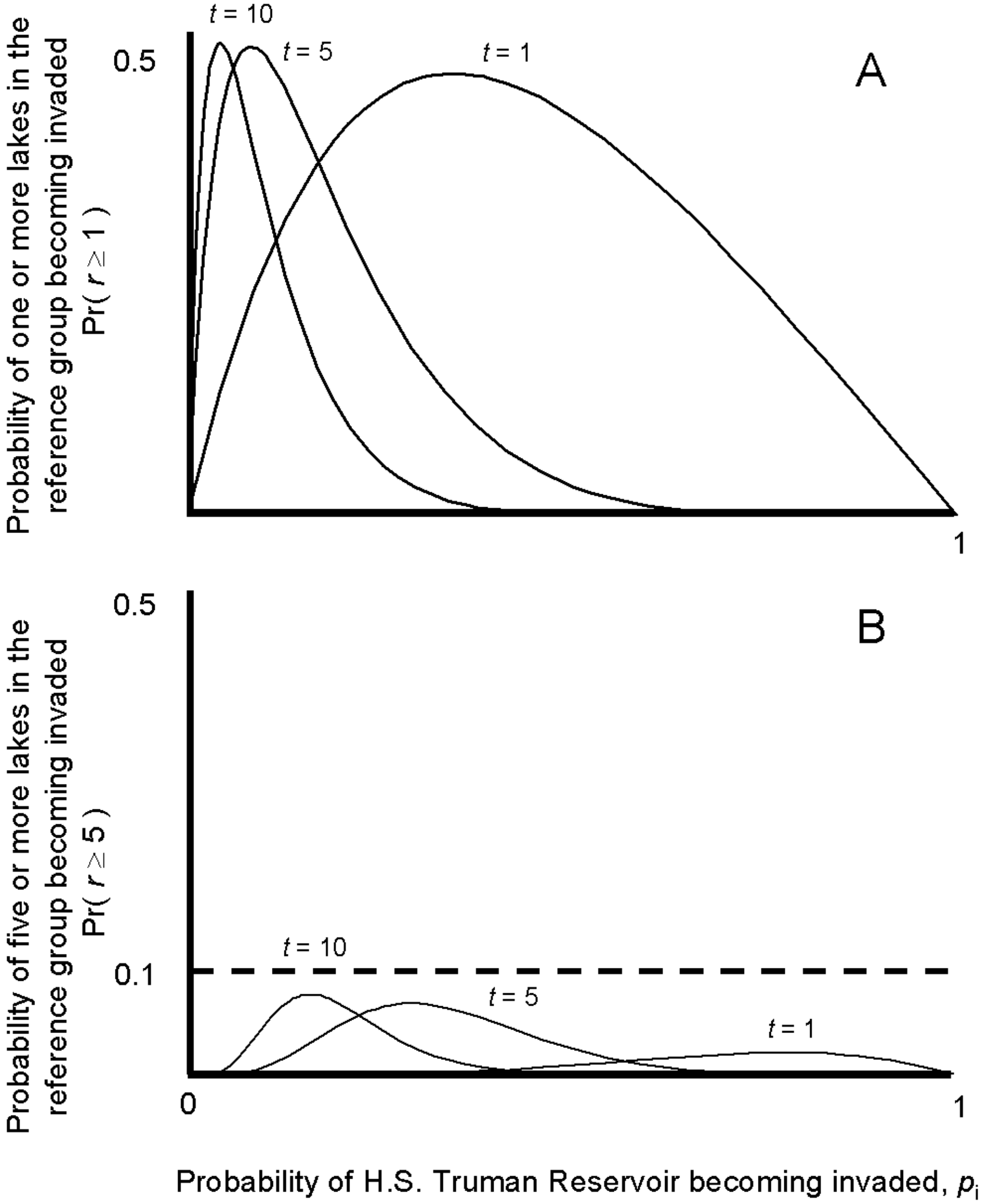


Figure 4.2: The probability that one or more (Panel A) or five or more (Panel B) of the reference lakes ($N = 12$) becoming invaded by time t as a function of the probability that H.S. Truman Reservoir becomes invaded, p_i. Because p_i is unknown there is considerable uncertainty about the probability of one or more lakes becoming invaded while H.S. Truman remains uninvaded, even over ten time steps. However, for all values of p_i and $t = 1, 5,$ and 10, the probability of observing five or more lakes invaded is uncommon (Pr$(r \geq 5) < 0.1$).

probability of observing five or more of the 12 reference lakes becoming invaded under the same conditions. The probability of observing one or more lakes invaded ranges from zero to approximately 0.5 for the possible values of p_i and for t=1, 5, and 10 time step periods of observation. With almost no better than a fair coin's chance of one or more of the reference lakes becoming invaded and the uncertainty due to p_i being unknown, observing a few of the reference lakes becoming invaded before H.S. Truman would not invalidate the predictions of the gravity model. However, observing five or more lakes invaded before H.S. Truman (Figure 4.2B) has only a 0.1 or less probability of being observed by chance for all time periods and all values of p_i. This observation provides a robust rule of thumb for evaluating the performance of predictions gleaned from the gravity scores. If I observe five or more lakes in the reference group become invaded before H.S. Truman Reservoir becomes invaded, I should be skeptical of the gravity model's predictive capabilities.

4.4.2 Evaluating predictions from gravity models

Validating gravity models through testing of predictions is needed. Gravity models for invasive species have a relatively young history compared to other models of species spread (Schneider et al., 1998; Bossenbroek et al., 2001; MacIsaac et al., 2004; Leung et al., 2004, 2006; Bossenbroek et al., 2007). One evaluation of predictive performance is to compare the gravity scores of recently invaded locations to uninvaded locations (MacIsaac et al., 2004; Leung et al., 2004, 2006). As mentioned previously, logistic regression is used with presence absence data to test the significance of the gravity score as the explanatory variable. In the studies thus far, the gravity score is shown to be positively correlated with successful invasions and is significant explanatory variable. This is the usual extent to which diagnostics of model fit and predictive power are performed (but see MacIsaac et al., 2004). However, more diagnostics exists, and some of these diagnostics are more useful for evaluating the predictive capability of logistic regression models.

The Receiver Operating Characteristic curve (ROC curve) and the resulting Area Under this Curve (AUC) is one such diagnostic. The purpose of such curves is to evaluate how well the logistic model is able to discriminate invaded lakes and uninvaded lakes based on the gravity score. The curve is a reflection of the sensitivity and specificity. Sensitivity accounts for the correct discrimination of successful invasions – meaning it is a proportion of lakes that are predicted and observed to be invaded. Similarly specificity accounts for correctly discriminating lakes predicted and observed to be uninvaded. Both sensitivity and specificity are proportions and range from 0 to 1. The ROC uses 1-specificity. This reflects the proportion of misclassified lakes that are predicted to be invaded but are actually uninvaded. A more thorough discussion of ROC curves, sensitivity, and specificity can be found in Hosmer and Lemeshow (2000).

The heuristic measure of the ROC is the AUC. When the AUC is between .9 and 1, the model does an excellent job of discriminating between invaded and uninvaded. In contrast when the AUC is close to 0.5, discriminating between invaded and uninvaded lakes is really no better than flipping a fair coin to predict whether it is invaded or not.

Table 5 provides the gravity scores for 15 lakes invaded just prior to construction of the national gravity model (Bossenbroek et al., 2007). Applying logistic regression to these lakes and the gravity scores of the uninvaded lakes found in Table 4 produces the ROC and AUC found in figure 3. With an AUC of 0.63, I can conclude the logistic regression model with the gravity score as an explanatory variable does a relatively poor job of discriminating, and hence predicting successful invasions. However a few caveats are in order. First this is a small sub-sample of the lakes in the national gravity model and it would be unfair to say the failure of this subset to provide a larger AUC is evidence for a failure of the entire system of lakes and the gravity model. Second, the logistic regression showed Q was not a significant covariate. Using a likelihood ratio test between a constant only and constant with parameter for the Q variable, the p-value was 0.57. This leads us to select the constant only model even though I used the model with the added parameter for estimating the AUC. Poor model fit is often associated, although not necessarily, with poor discrimination (Hosmer and Lemeshow, 2000). As mentioned earlier, one of the consequence of measurement error is the failure to

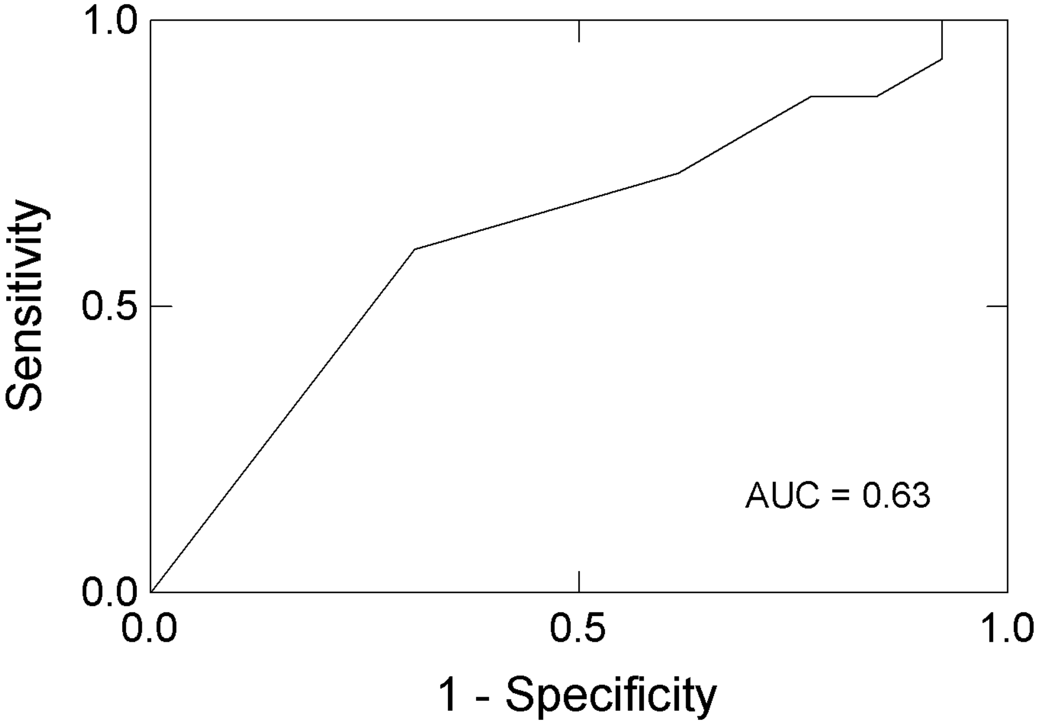


Figure 4.3: The Receiver Operator Characteristic (ROC) curve for the logistic regression of invaded and uninvaded lakes based on the gravity scores. The area under this curve (AUC) is 0.63 and indicates there is poor discrimination between invaded and uninvaded lakes. As a diagnostic, the ROC and resulting AUC indicate the gravity score for this subsample of lakes has poor predictive capabilities.

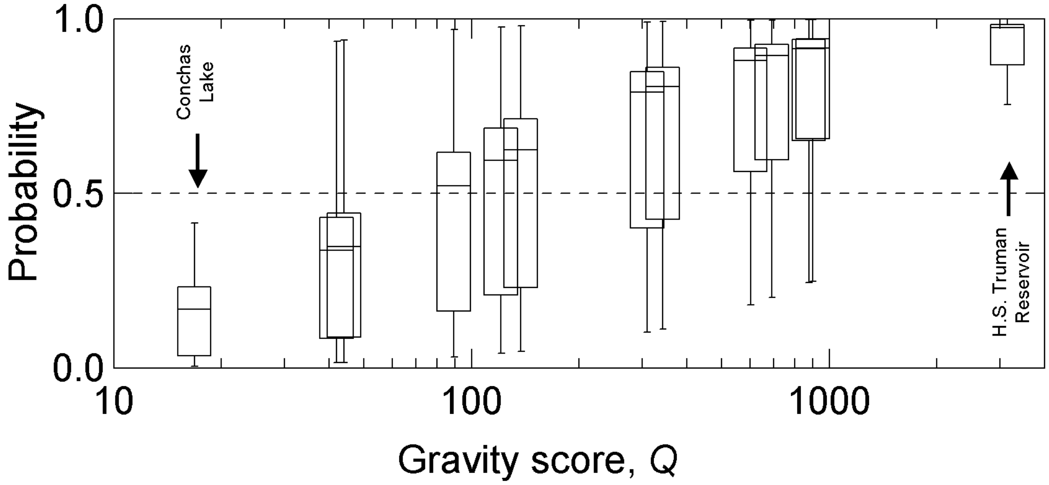


Figure 4.4: The pair–wise probability distributions of invaded lakes for each uninvaded lake. Each box plot is one of the uninvaded lakes from Table 4, positioned at its respective gravity score Q. The distribution of the box plot is all of the pair–wise probabilities with invaded lakes. Most uninvaded lakes have expectations greater than 0.5 which implies they are more susceptible, based on the gravity score, to invasion than lakes already invaded.

detect significant covariates (Thomas et al., 1993). Future studies may consider the approach and subsequent improvements on this concern suggested by Wacholder et al. (1993) or Lele and Allen (2006).

The purpose of discussing ROC and AUC as diagnostics is not to call into question the national gravity model for zebra mussels. Rather ROC and AUC diagnostics represent a tool to evaluate the predictive performance of the model. ROC and AUC are not exclusive to logistic regression and may be useful with other applications, and therefore should be the subject of future studies that seek to demonstrate the predictive capabilities of invasion models.

Pair–wise probabilities can also be used as a visual diagnostic to assess uncertainty in model predictions. Table 5 contains the pair–wise probabilities of invaded (rows) and uninvaded (columns) lakes. The table itself is difficult to decipher as would be the ROC and AUC diagnostics without some familiarity. However, Figure 3 contains 13 box plots, one for each uninvaded lake, that shows the distribution of the pair–wise probabilities to the group of invaded lakes. The

	Q	H.S. Truman Reservoir	Lake Oahe	Barren River Lake	Martin Lake	Austin Lake	Lake Mead	Upper Saranac Lake	Perry Lake	Chesuncook Lake	Roosevelt Lake	Amistad Reservoir	Goose Lake	Conchas Lake
Lake of the Ozarks	2717.6	0.46	0.75	0.76	0.80	0.82	0.89	0.90	0.95	0.96	0.97	0.98	0.99	0.99
Grand Lake O'the Cherokees	1024.5	0.25	0.53	0.54	0.60	0.63	0.75	0.77	0.88	0.90	0.92	0.96	0.96	0.98
Hamilton Lake	629.09	0.17	0.41	0.42	0.48	0.51	0.65	0.67	0.82	0.84	0.88	0.94	0.94	0.97
Lake Margarethe	608.92	0.16	0.40	0.41	0.47	0.50	0.64	0.66	0.82	0.83	0.87	0.93	0.94	0.97
Clear Lake	356.32	0.10	0.28	0.29	0.34	0.37	0.51	0.54	0.72	0.75	0.80	0.89	0.90	0.96
Rock Lake	214.22	0.06	0.19	0.20	0.24	0.26	0.39	0.41	0.61	0.64	0.71	0.83	0.84	0.93
Prairie River Lake	85.83	0.03	0.09	0.09	0.11	0.13	0.20	0.22	0.39	0.42	0.49	0.66	0.67	0.84
Horicon NWR	82.27	0.03	0.08	0.09	0.11	0.12	0.19	0.21	0.38	0.41	0.48	0.65	0.66	0.83
Lake George	76.86	0.02	0.08	0.08	0.10	0.11	0.18	0.20	0.36	0.39	0.46	0.64	0.65	0.82
Winfield City Lake	68.63	0.02	0.07	0.07	0.09	0.10	0.17	0.18	0.33	0.36	0.43	0.61	0.62	0.80
Cass Lake	56.19	0.02	0.06	0.06	0.08	0.09	0.14	0.15	0.29	0.32	0.39	0.56	0.57	0.77
Big Bradford Lake	54.46	0.02	0.06	0.06	0.07	0.08	0.14	0.15	0.28	0.31	0.38	0.55	0.57	0.77
Hartwick Lake	52.92	0.02	0.06	0.06	0.07	0.08	0.13	0.15	0.28	0.30	0.37	0.55	0.56	0.76
Ess Lake	23.62	0.01	0.03	0.03	0.03	0.04	0.07	0.07	0.15	0.16	0.21	0.35	0.36	0.59
Base Lake⋆	2.90	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.03	0.06	0.07	0.15

Table 4.5: Probabilities of invaded lakes being invaded before currently uninvaded lakes. Light gray cells identify pairing that have a greater than 0.5 probability of being invaded before the uninvaded locations. The regular invasion of lakes with small paired probabilities of invasion relative to more susceptible lakes would suggest the model is not adequate to predict invasions. ⋆ Base Lake is inside the confines of a military base and was likely invaded through a different mechanism or pathway than is modeled by a gravity model based on recreational boaters.

box plots are ordered by the gravity score, Q, such that H.S. Truman reservoir, with the largest gravity score is the right most box plot and Conchas Lake, with the smallest gravity score, is the left most plot. The spread of each box represents the uncertainty associated with each uninvaded lake given the observed group of invaded lakes. Interestingly, 10 of the 13 uninvaded lakes have means greater than 0.5 indicating, that relative to the values of Q for the recently invaded lakes, there is a reasonable chance of invasion. Of the three remaining uninvaded lakes, there is considerable variability, as demonstrated by the whiskers of the box plots. I conclude, similarly to the ROC and AUC plots, but by visual inspection, that there is considerable uncertainty in the likelihood of invasion and the predictive power resulting from the gravity model is questionable, again with the same caveats.

I have offered a few prescriptions to account for uncertainty in gravity models, including building and reporting confidence intervals on parameter estimates, applying probability theory to the order of invasions, ROC plots and AUC metrics, and building box plots of relative probabilities. Undoubtedly, other diagnostics exist and should be explored. I very much encourage the development of diagnostics for gravity models of invasive species dispersal, and believe this will be an avenue of future invasive species research.

From the small collection of invaded and uninvaded lakes and the analyses performed here, one may conclude that gravity models are left wanting. I believe this is not a fair conclusion. Indeed, there appears to be predictive performance issues with gravity models. However, this is less an indictment of gravity models than more of a guidepost of where research needs to go. For example, there are likely groups of lakes with similar susceptibility of being invaded where there is substantial differences in susceptibility between groups. This phenomena was observe in the suitability of United States shipping ports to the potential invasion of Chinese Mitten Crab (*Eriocheir sinensis*) when coupling relative measure of propagule pressure and habitat suitability mapping (Herborg et al., 2007). The result of comparing two locations with different susceptibility is that the probability of invasion will be different even though the gravity scores are similar (Chapter 2; Jerde and Lewis, 2007). A likely next objective in gravity model research will be to include susceptibility in evaluating the model predictions (Muirhead, 2007). In the case of the uninvaded lakes in table 4, only Lakes Mead and Roosevelt have been evaluated for susceptibility of zebra mussel invasion (Bossenbroek et al., 2007).

4.5 Discussion

The study of biological invasions is plagued by uncertainty. From identifying the characteristics of successful invaders (Goodwin et al., 1999) or detecting new invaders at a location (Costello et al., 2007), to predicting when and where the next lake in the Western United States is going to be invaded by zebra mussels (Bossenbroek et al., 2007), there are few topics that are deterministic (Gilpin, 1990). This includes not only the epistemic uncertainty emphasized in this chapter, but also the linguistic uncertainty of the terminology used in the biological and bioeconomic research of invasive species (Colautti and MacIsaac, 2004; Shrader-Frechette, 2001). Here I have provided a biological and statistical perspective of uncertainty with emphasis on the role of making and evaluating predictions, in particular to a subset of invaded and uninvaded lakes with scores from a gravity model of zebra mussel dispersal (Bossenbroek et al., 2007). I have offered a few prescriptions for evaluating the predictive capability of gravity models, but much more needs to be done to account for the uncertainty in invasion biology. Listing all the ways we may account for uncertainty would be an arduous task indeed. Instead, I will offer examples from the literature that I believe will guide researchers in future studies and investigations of biological invasions.

Identifying sources of uncertainty for any biological process is a critical step that should be done in concert with formulating hypotheses and models (Hilborn and Mangel, 1997; Lewin-Koh et al., 2004). Carlton (1996) provides an exemplary overview of the sources of uncertainty for predicting the arrival of invasive species into new locations. He identified six important sources of uncertainty: changes in donor regions, new donor regions, changes in recipient region, invasion windows, stochastic inoculation events, and dispersal vector changes. These processes have become the subject of empirical investigations (e.g. Wonham et al., 2000) and theoretical frameworks (e.g. Chapter 2; Jerde and Lewis, 2007). More generally, overviews of the study of biological invasions (e.g. Vermeij, 1996; Puth and Post, 2005) highlight areas in the field that need more attention, due in large part to the uncertainty of specific processes, such as predicting invasions (Kolar and Lodge, 2001).

Reducing linguistic uncertainty has been accomplished, so far, by evaluating the terminology used in the literature. To this end there are many notable examples of how to proceed (e.g. Colautti and MacIsaac, 2004). When recognizable differences between definitions for the same term are detected, it may be time for a critical review that attempts to bridge and clarify the discrepancy. One such term, from a bio-economic perspective, is the term 'risk'. For biologists risk is often associated with a probability of some unwanted event occurring, such as invasion (Chapter 2; Suter, 1993; Jerde and Lewis, 2007). However, economists generally associate risk with the probability of an event occurring times the loss accrued because of that event occurring (Knight, 1921). Undoubtedly interdisciplinary approaches to biological invasions will uncover similar disparities and will require some attention.

Reporting parameters with some measure of variability is a common method for dealing with uncertainty (Cumming et al., 2007). One approach to mathematical modeling is to produce a model and then search the literature for the parameter values of a particular species and/or system from which to make predictions. This is actually quite difficult because only point estimates (i.e. means), and not measures of variability, are often reported, especially for parameters such as growth and predation rates (see Chapter 5). A measure of variability in parameter estimates can be use to analyze qualitative differences between model predictions (e.g. Wonham et al., 2006) from perturbation analysis (Caswell, 2001). This problem can easily overcome with diligence in the reporting of descriptive statistics for point estimates (such as the variance or standard deviation) in future empirical studies of biological invasions. A good review of the appropriate error bars to produce for point estimates is presented by Cumming et al. (2007).

Assessing gravity model structure and the variables to include in a gravity model has received limited attention. The differences in models is usually determined by the available data, such as production constrained or doubly constrained gravity models (Leung et al., 2006). That said, Leung et al. (2004) provided an example of how to detect Allee effects from gravity scores. Yet many questions remain for gravity model research. For example, the attraction coefficient, W_j (Table 2), is the area of the destination lake or the area of lakes within a water shed (Bossenbroek et al., 2007), but is area a reasonable measure of how attractive a lake is to boaters (e.g. Reed-Andersen et al., 2000) or should measures of recreational fishing opportunities, water quality, or water skiing be used? Similarly, watersheds that have big reservoirs, such as Lakes Mead, Roosevelt, and Oahe can probably be assumed to be the main attractor of these watershed and it seems reasonable that the proportion of boaters coming to these big reservoirs is related to the overall proportion of water in the watershed. For watersheds in, for example, Michigan that have a lot of small lakes, the spatial interactions within the watershed, including the distribution of people living within the watershed, may be a critical consideration not currently captured in the gravity model dynamics. With the predictive performance observed within this chapter, it is likely that assessing the model structure of gravity models will be a fruitful area for reducing the uncertainty of invasions.

Likely the most pressing issue for invasion biologists with respect to epistemic uncertainty is that of predicting successful invasions – successfully. This is why I focused much of my attention on gravity model diagnostics of predictive performance. To date, little has been presented regarding the predictive performance of gravity models other than to show significant correlation between the gravity scores and observed invasions (MacIsaac et al., 2004; Leung et al., 2004). However, it bears repeating that correlation is not necessarily an indicator of good predictive performance (Hosmer and Lemeshow, 2000).

Applying the probability theory of waiting times (Chapter 2; Jerde and Lewis, 2007; Drake et al., 2005) appears to be one avenue for evaluating predictions from gravity models with a lot of potential. With the specific question, ‘what is the probability of observing r lakes invaded before H.S. Truman Reservoir,’ I was able to find a robust threshold, that is to say there is a probability of less than 0.1 of observing five lakes invaded within 1, 5, and 10 year time periods. Similarly, there appears to be some usefulness of formulating relative waiting times (Chapter 2; Jerde and Lewis, 2007), but this likely needs to be expanded into formulations that include multiple lakes or groups of lakes with similar invasion susceptibility (Herborg et al., 2007).

One of the drawbacks of predictive formulations based on waiting times is that in order to validate or invalidate the predictions, we must wait for invasions to occur. This is likely a problem for the impatient and more importantly for managing the spread of invasive species. Alternatively there are diagnostics like ROC and AUC that assess predictive performance based on the model's ability to discriminate between invaded and uninvaded lakes based on the gravity score, and can be conducted on existing data without having to wait for future invasions. However, the ability to discriminate between invaded and uninvaded can be sensitive to measurement error (Thomas et al., 1993). Therefore it appears gravity scores will need to be calibrated with census data about the number of boaters arriving to lakes that are infested with zebra mussels, but some of the necessary information may already be available (e.g. Johnson et al., 2001).

I have focused on the biological processes of invasive species and the uncertainty in understanding these processes. The next step is to translate these measures into decision making frameworks based on the economic realities of managing invasive species. One last uncertainty issue that should be raised is the uncertainty of uncertainty. As mentioned earlier, risk, in economics, is the product of the probability of an event occurring and the loss accrued due to that event occurring. Certainty is defined by the probability of an event occurring equal to 0 or 1 and uncertainty in the outcome occurs everywhere in between. Much of the uncertainty I have emphasized is in the estimate of the probability of that event occurring. As a result, the biological uncertainty in a processes will in large part also influence the ability to make any economic decision.

Bibliography

Barnett, V. 2002. Sample survey: principles and methods. 3rd edition. Hodder Arnold Press.

Bossenbroek, J., C. Kraft, and J. Nekola. 2001. Prediction of long-distance dispersal using gravity models: Zebra Mussel invasion of inland lakes. Ecological Applications **11**:1778–1788.

Bossenbroek, J., J. Ladd, B. Peter, and D. Lodge. 2007. Forecasting the expansion of Zebra Mussels in the United States. Conservation Biology **21**:800–810.

Bossenbroek, J., J. McNulty, and R. Keller. 2005. Can ecologists heat up the discussion on invasive species risk? Risk Analysis **25**:1595–1597.

Burnham, K., and D. Anderson. 2002. Model selection and multimodel inference: a practical information–theoretic approach. Second edition. Springer, New York.

Carlton, J. 1996. Pattern, process, and prediction in marine invasion ecology. Biological Conservation **78**:97–106.

Caswell, H. 2001. Matrix Population Models. 2nd edition. Sinauer.

Christian, J., and S. Wilson. 1999. Long–term ecosystem impacts of an introduced grass in the northern Great Plains. Ecology **80**:2397–2407.

Colautti, R., and H. MacIsaac. 2004. A neutral terminology to define invasive species. Diversity and Distributions **10**:135–141.

Costello, C., J. Drake, and D. Lodge. 2007. Evaluating an invasive species policy: ballast water exchange in the Great Lakes. Ecological Applications **17**:655–662.

Costello, C., and A. Solow. 2003. On the pattern of discovery of introduced species. Proceedings of the National Academy of Sciences **100**:3321–3323.

Cumming, G., F. Fidler, and D. Vaux. 2007. Error bars in experimental biology. Cell Biology **177**:7–11.

Davis, M., K. Thompson, and J. Grime. 2001. Charles S. Elton and the dissociation of invasion ecology from the rest of ecology. Diversity and Distributions **7**:97–102.

Diggins, T., M. Weimer, K. Stewart, R. Baier, A. Meyer, R. Forsberg, and M. Goehle. 2004. Epiphytic refugium: are two species of invading freshwater bivalves partitioning spatial resources? Biological Invasions **6**:83–88.

Dovers, S., and T. Norton. 1996. Uncertainty, ecology, sustainability and policy. Biodiversity and Conservation **5**:1143–1167.

Drake, J., P. Baggenstos, and D. Lodge. 2005. Propagule pressure and persistence in experimental populations. Biology Letters **1**:480–483.

Drake, J., and D. Lodge. 2006. Allee effects, propagule pressure and the probability of establishment: risk analysis for biological invastions. Biological Invasions **8**:365–375.

Efron, B., and R. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall / CRC, Boca Raton, Florida.

Elton, C. 1958. The ecology of invasion by animals and plants. Methuen, London, UK.

Gilpin, M. 1990. Review: ecological prediction. Science **248**:88–89.

Goodwin, B., A. McAllister, and L. Fahrig. 1999. Predicting invasiveness of plant species based on biological information. Conservation Biology **13**:422–426.

Harwood, J., and K. Stokes. 2003. Coping with uncertainty in ecological advice: lessons from fisheries. Trends in Ecology and Evolution **18**:617–622.

Hebert, P., B. Muncaster, and G. Mackie. 1989. Ecological and genetic studies on *Dreissena polymorpha* (Pallas): a new mollusc in the Great Lakes. Canadian Journal of Fisheries and Aquatic Sciences **46**:1587–1591.

Herborg, L., C. Jerde, D. Lodge, G. Ruiz, and H. MacIsaac. 2007. Predicting the North American distribution of Chinese mitten crabs (*Eriocheir sinensis*) using measures of propagule pressure and environmental niche models. Ecological Applications **17**:663–674.

Hilborn, R., and M. Mangel. 1997. The Ecological Detective: confronting models with data. Princeton University Press, Princeton, New Jersey.

Hosmer, D., and S. Lemeshow. 2000. Applied Logistic Regression. 2nd edition. Wiley Interscience, New York, NY.

Jerde, C., and M. Lewis. 2007. Waiting for invasions: a framework for the arrival of non–indigenous species. American Naturalist **170**:1–9.

Johnson, L., A. Ricciardi, and J. Carlton. 2001. Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. Ecological Applications **11**:1789–1799.

Jonzen, N., P. Lundberg, E. Ranta, and V. Kaitala. 2002. The irreducible uncertainty of the demography-environment interaction in ecology. Proceedings of the Royal Society of London, B **269**:221–225.

Knight, F. 1921. Risk, Uncertainty, and Profit. Houghton Mifflin, Boston, Mass.

Kolar, C., and D. Lodge. 2001. Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution **16**:199–204.

Lande, R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. American Naturalist **142**:911–927.

Lele, S., and K. Allen. 2006. On using expert opinion in ecological analyses: a frequentist approach. Environmetrics **17**:683–704.

Lele, S., M. Taper, and S. Gage. 1998. Statistical analysis of population dynamics in space and time using estimating functions. Ecology **79**:1489–1502.

Leung, B., J. Bossenbroek, and D. Lodge. 2006. Boats, pathways, and aquatic biological invasions: estimating dispersal potential with gravity models. Biological Invasions **8**:241–254.

Leung, B., J. Drake, and D. Lodge. 2004. Predicting invasions: propagule pressure and the gravity of Allee effects. Ecology **85**:1651–1660.

Leung, B., D. Lodge, D. Finnoff, J. Shogren, M. Lewis, and G. Lamberti. 2002. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proceedings of the Royal Society of London, B **269**:2407–2413.

Lewin-Koh, N., M. Taper, and S. Lele, 2004. The nature of scientific evidence: statistical, philosophical, and empirical considerations, Chapter a brief tour of statistical concepts, pages 3–16 . University of Chicago Press, Chicago, Ill.

Lockwood, J., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. Trends in Ecology and Evolution **20**:223–228.

Longino, J., and R. Colwell. 1997. Biodiversity assessment using structured inventory: Capturing the ant fauna of a tropical rain forest. Ecological Applications **7**:1263–1277.

MacIsaac, H., J. Borbely, J. Muirhead, and P. Graniero. 2004. Backcasting and forecasting biological invasions of inland lakes. Ecological Applications **14**:773–783.

May, B., and J. Marsden. 1992. Genetic identification and implications of another invasive species of dreissenid mussel in the Great Lakes. Canadian Journal of Fisheries and Aquatic Science **49**:1501–1506.

Muirhead, J., 2007. Forecasting and validating the establishment of aquatic nonindigenous species. Ph.D. thesis, The University of Windsor, Windsor, Ontario, Canada.

Parker, I., D. Simberloff, W. Lonsdale, K. Goodell, M. Wonham, P. Kareiva, M. Williamson, B. Von Holle, P. Moyle, J. Byers, and L. Goldwasser. 1999. Impact: Toward a framework for understanding the ecological effects of invaders. Biological Invasions **1**:3–19.

Pimentel, D., L. Lach, R. Zunign, and D. Morrison. 2000. Environmental and economic costs of non-indiginous species in the United States. BioScience **50**:53–67.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien–invasive species in the United States. Ecological Economics **52**:273–288.

Puth, L., and D. Post. 2005. Studying invasion: have we missed the boat? Ecology Letters **8**:715–721.

Reed-Andersen, T., E. Bennett, B. Jorgensen, G. Lauster, D. Lewis, D. Nowacek, J. Riera, B. Sanderson, and R. Stedman. 2000. Distribution of recreational boating across lakes: do landscape variables affect recreational use? Freshwater Biology **43**:439–448.

Regan, H., M. Colyvan, and M. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications **12**:618–628.

Ricciardi, A., and J. Cohen. 2007. The invasiveness of an introduced species does not predict its impact. Biological Invasions **9**:309–315.

Richardson, D., M. Pysek, M. Rejmanek, M. Barbour, F. Panetta, and C. West. 2000. Naturalization and invasion of alien plants: concepts and definitions. Diversity and Distributions **6**:93–107.

Rouget, M., D. Richardson, S. Milton, and D. Polakow. 2004. Predicting invasion dynamics of four alien *Pinus* species in a highly fragmented semi–arid shrubland in South Africa. Pland Ecology **152**:79–92.

Sakai, A., F. Allendorf, J. Holt, D. Lodge, J. Molofsky, K. With, S. Baughman, R. Cabin, J. Cohen, N. Ellstrand, D. McCauley, P. O'Neil, I. Parker, J. Thompson, and S. Weller. 2001. The population biology of invasive species. Annual Reviews of Ecology and Systematics **32**:305–32.

Schneider, D., C. Ellis, and K. Cummings. 1998. A transportation model assessment of the risk to native mussel communities from Zebra Mussel spread. Conservation Biology **12**:788–800.

Shaffer, M. 1981. Minimum population sizes for species conservation. BioScience **31**:131–134.

Shogren, J., and D. Finnoff, In Press. Bioeconomics of Invasive Species: Integrating Ecology, Economics, Policy, and Management, Chapter uncertain invasions: An economic perspective . Oxford University Press.

Shrader-Frechette, K. 2001. Non–indigenous species and ecological explanation. Biology and Philosophy **16**:507–519.

Simberloff, D. 2003. How much information on population biology is needed to manage introduced species? Conservation Biology **17**:83–92.

Solow, A., and C. Costello. 2004. Estimating the rate of species introductions from the discovery record. Ecology **85**:1822–1825.

Stohlgren, T., and J. Schnase. 2006. Risk Analysis for biological hazards: what we need to know about invasive species. Risk Analysis **26**:163–173.

Suter, G. 1993. Ecological Risk Assessment. Lewis Publishers, Chelsea, Michigan.

Taper, M., and S. Lele, 2004. The nature of scientific evidence: statistical, philosophical, and empirical considerations, Chapter dynamical models as paths to evidence in ecology, pages 275–297 . University of Chicago Press, Chicago, Ill.

Thomas, D., D. Stram, and J. Dwyer. 1993. Exposure measurement error: influence on exposure-disease relationships and methods for correction. Annual Reviews of Public Health **14**:69–93.

Vermeij, G. 1996. An agenda for invasion biology. Biological Conservation **78**:3–9.

Von Holle, B., and D. Simberloff. 2005. Ecological resistance to biological invasion overwhelmed by propagule pressure. Ecology **86**:3212–3218.

Wacholder, S., B. Armstrong, and P. Hartge. 1993. Validation studies using an alloyed gold standard. American Journal of Epidemiology **137**:1251–1258.

Wonham, M., J. Carlton, G. Ruiz, and L. Smith. 2000. Fish and ships: relating dispersal frequency to success in biological invasions. Marine Biology **136**:1111–1121.

Wonham, M., M. Lewis, J. Renclawowicz, and P. van den Driessche. 2006. Transmission assumptions generate conflicting predictions in hostvector disease models: a case study in West Nile virus. Ecology Letters **9**:706–725.

Chapter 5

Chance establishment for sexual, semelparous species: overcoming the Allee effect

5.1 Introduction

Populations at low density are potentially prone to declining or negative per capita growth rates as the population density decreases (Allee effect) (see, for example, Allee, 1938; Odum and Allee, 1954; Dennis, 1989). This effect has been invoked to explain why many repeatedly introduced, non-indigenous species (NIS) fail to establish (Sakai et al., 2001; Lockwood et al., 2005). For some NIS, such as zebra mussels (*D. polymorpha*), there is evidence the Allee effect is a plausible explanation for the frequency of unsuccessful invasions at locations receiving low levels of propagule pressure (Leung et al., 2004).

In the absence of an Allee effect, repeated introductions will eventually lead to successful establishment, although this may take considerable time to occur (Chapter 2; Jerde and Lewis, 2007). The only management action that will negate the risk of invasion is to eliminate all propagule pressure. However, due to the costs of searching for and eliminating the last few individuals in a population (Regan et al., 2006) or in the arrival pathway, the elimination of propagule pressure may be difficult or impossible (Simberloff, 2003*a*). In contrast, the presence of Allee effects allow alternative management: if propagule pressure can be reduced below a threshold, then repeated introductions will continually fail to establish (Drake and

Lodge, 2006; Drury et al., In Press). However, testing and assessing the strength of Allee effects generally requires collecting population data after the species establishes and while it spreads (Veit and Lewis, 1996; Leung et al., 2004), whereas proactive management of NIS to prevent invasions would require identifying the presence and strength of Allee effects before data are actually available to fit population models. This has led some to question the utility of population models for invasive species risk assessment and proactive management (Simberloff, 2003*b*). Therefore, developing a methodology for proactively assessing the strength of an Allee effect based on known life history traits from other locations would be useful for informing decisions regarding the management of potential invaders.

Separate from Allee effects, demographic stochasticity will also influence the persistence or extinction of a species (Goodman, 1987). Demographic stochasticity arises because of chance events resulting from individual birth and death processes (Lande, 1993) and through fluctuations in the sex ratio (Bessa-Gomes et al., 2004). At large population abundances, the variability due to demographic stochasticity becomes negligible (Engen et al., 1989), but at small population abundances, demographic stochasticity can be a dominant force that will determine the fate of a population (Lande, 1993), which in turn will have consequences on conservation management (Goodman, 1987). Because NIS are often introduced at low abundances (Lockwood et al., 2005), the influence of demographic stochasticity is an additional, critical consideration for evaluating the risk of invasion (Drake and Jerde, In Press).

Using life history traits to predict invasions has a mixed record of success but some generalizations have been identified (Williamson, 1999; Kolar and Lodge, 2001). For example, Keller et al. (2007) found that fecundity of freshwater molluscs was a significant predictor of whether a species would cause ecological or economic damage. Similarly, statistics of life history, such as the net reproductive rate (R_0), which is the lifetime production of offspring per female, can be estimated for a wide variety of life histories (de Camino Beck and Lewis, 2007) and is related to the spread rate of invading organisms (Kot et al., 1996). But can life history traits be used to predict the presence and strength of Allee effects?

In this chapter, I propose a stochastic model for semelparous, two-sex species establishment that has an emerging Allee effect and demographic stochasticity. The model employs life history parameters (survival, sex ratio, fertilization probability, and fecundity) from which it is possible to calculate the probability of extinction in the next generation, the expected net per capita growth rate, and the expected net population change for an initial population size introduced into a location. From the model, I identify whether a weak or strong Allee effect is expected in the population using independently estimated parameters from other locations or laboratory studies. Additionally, introduced populations have the potential to persist for a few generations and then go extinct due to the Allee effect and demographic stochasticity. Therefore, I show how the model can be formulated as a stochastic process and demonstrate the resulting changes in the distribution of population abundance.

The model is applied to two semelparous organisms, the apple snail and Chinese mitten crab, both identified as some of the world's worst invaders by The World Conservation Union's Invasive Species Specialist Group (http://www.issg.org). The apple snail demonstrates a weak Allee effect, and the Chinese mitten crab has a strong Allee effect. However, because the expected fecundity of these two organisms is so large, the Allee effect, while present, does not significantly influence the probability of population establishment. Instead, the probability of at least one female becoming fertilized and surviving to give rise to the next generation appears to be the critical consideration for predicting successful invasions.

5.2 Hierarchical Allee effect model

I take the approach of modeling population dynamics with discrete probability events formulated in a hierarchy of processes (Figure 5.1). All random variables are defined in Table 5.1, and all parameters are defined in Table 5.2. First, a fixed number of individuals, n_0, are introduced into a new landscape. These individuals must then survive to reproductive maturity with a survival probability (survival) denoted as p. These surviving individuals, S, are then separated into females, F,

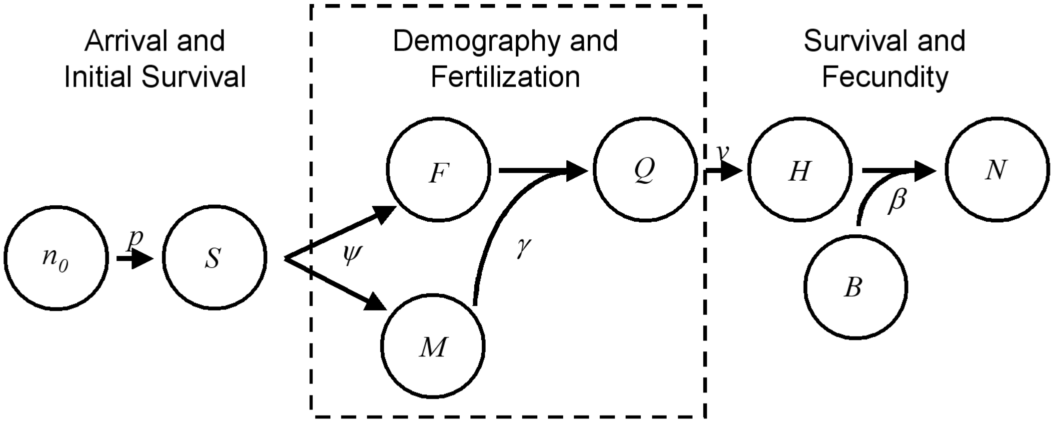


Figure 5.1: Flow diagram of the hierarchical population model. Random variables are defined in Table 5.1 and parameters are defined in Table 5.2.

and males, M, with the probability of being female, ψ. A female is fertilized by any given male with probability, γ. Assuming promiscuous mating and no mate choice, the probability of a female becoming fertilized increases monotonically, $1 - e^{-n_0\gamma p(1-\psi)}$, with the number of males. Each fertilized female has a survival, v, to birthing. The model is of a semelparous organism, so the number of offspring, B, produced by each fertilized female, H, become the population at the next generation.

Reducing the model through conditional probability results in a compound Poisson process (CPP) that is conditional on the random variables that describe the number of fertilized females that survive to birthing, H, and the number of offspring per surviving fertilized female, B,

$$N \mid H, B = \sum_{i=1}^{H} B_i, \tag{5.1}$$

where $H \sim \text{Poisson}(n_0 v p \psi(1 - e^{-n_0\gamma p(1-\psi)}))$ and $B \sim \text{Poisson}(\beta)$. The random variable, N, is the population size in the next generation. The model formulation and details of the simplifications used to arrive at Equation 5.1 are found in Appendix H: Formulation of the hierarchical probability model of establishment.

Random Variable	Description	Support
S	Number of initially surviving individuals in the population	$\{0, 1, 2, ...n_0\}$
F	Number of females in the population	$\{0, 1, 2, ...S\}$
M	Number of males in the population	$\{0, 1, 2, ...S\}$
D	Number of individuals that do not initially survive (see appendix)	$\{0, 1, 2, ...n_0\}$
K	Success of one male attempting to fertilize one female (see appendix)	$\{0, 1\}$
Y	Number of males that successfully fertilize a female (see appendix)	$\{0, 1, 2, ...M\}$
Q	Number of successfully fertilized females	$\{0, 1, 2...F\}$
H	Number of successfully fertilized females that survive to reproduce	$\{0, 1, 2, ...Q\}$
B	Fecundity	$\{0, 1, 2, ...\}$
N	The total number of offspring (Next generation population size)	$\{0, 1, 2, ...\}$
R	The *net* geometric growth rate	$\{-1 < r < \infty\}$

Table 5.1: Random variables of the hierarchical model.

Parameter	Description	Support
n_0	Initial population size	$\{1, 2, \ldots\}$
p	Probability of surviving from introduction to mating	$\{0 \leq p \leq 1\}$
ψ	Probability of being female	$\{0 \leq \psi \leq 1\}$
γ	Probability of a male fertilizing a female	$\{0 \leq \gamma \leq 1\}$
v	Probability of surviving from mating to birthing	$\{0 \leq v \leq 1\}$
β	Expected number of offspring	$\{\beta > 0\}$

Table 5.2: Parameters of the hierarchical model.

5.3 Model properties

5.3.1 Mean, variance, and probability generating function of compound Poisson processes

Compound Poisson process random variables do not have convenient marginal distributions, and instead I rely on the probability generating function to describe the shape of a random variable's distribution. Probability generating functions are useful mathematical constructs that capture the random variable's distributional shape as a power series and allows for the calculation of the mean, variance, and probability mass.

The probability generating function for the CPP is,

$$G(t) = \mathrm{E}[t^N] = \sum_{n=0}^{\infty} \Pr(N = n) t^n = e^{(e^{(t-1)\beta} - 1)(n_0 v p \psi (1 - e^{-n_0 \gamma p (1-\psi)}))}, \tag{5.2}$$

and is used to calculate the mean and variance of N (see, for example Chatfield and Theobald, 1973),

$$\mathrm{E}[N] = \beta(n_0 v p \psi(1 - e^{-n_0 \gamma p(1-\psi)})) \tag{5.3}$$

$$\mathrm{VAR}[N] = (\beta + \beta^2)(n_0 v p \psi(1 - e^{-n_0 \gamma p(1-\psi)})) \tag{5.4}$$

$$= (1 + \beta)\mathrm{E}[N]. \tag{5.5}$$

The probability mass function (PMF) is recovered from the probability generating function by,

$$\Pr(N = n) = \frac{G^{(n)}(0)}{n!}. \tag{5.6}$$

This formulation of the PMF is computationally expensive because evaluating $\Pr(N = n)$ requires calculating the n^{th} derivative of the generation function and evaluating the derivative at $t = 0$. Alternatively, it is possible to write the probability of $N = n$ recursively such that,

$$\Pr(N = n) = \begin{cases} G(0) & \text{when } n = 0, \\ \frac{c(n_0)\,\beta^n}{n} \sum_{i=0}^{n-1} \frac{\Pr(N=i)}{\beta^i (n-1-i)!} & \text{when } n > 0 \end{cases} \tag{5.7}$$

where $c(n_0) = n_0 p v \psi(1 - e^{(-n_0 \gamma p(1-\psi))})e^{-\beta}$. This recursive formulation does not require the calculation of derivatives, thus easing the computational burden. A derivation and proof of the recursive formula are given in Appendix E.

5.3.2 Geometric growth rate

Because the model projects the population from n_0 to the random variable N in discrete time (with time step of one generation), I use the ***net*** *geometric per capita growth rate*, R, to assess population growth (see, for example, Case, 2000),

$$R = \frac{N}{n_0} - 1. \tag{5.8}$$

The population at the next generation, N, is a random variable, so R is also a random variable, where $\mathrm{E}[R] = \frac{\mathrm{E}[N]}{n_0} - 1$ and $\mathrm{VAR}[R] = \frac{\mathrm{VAR}[N]}{n_0^2}$. When n_0 becomes large, R converges to $p v \psi \beta - 1$ where the $\mathrm{VAR}[R] \to 0$. The *geometric per capita rate of growth*, $\lambda = R + 1$, is an alternative formulation. Positive R indicates population growth, negative R indicates population decline, and $R = 0$ results in no change from the initial population size to the next generation population size, $N = n_0$.

The asymptotic growth rate ($n_0 \to \infty$) is the product of the survival (vp), the probability of being female (ψ), and the fecundity (β). However, when the initial population size (n_0) is small, the expected geometric growth rate, E$[R]$, is discounted by the probability of at least one male fertilizing a female, $1 - e^{(-n_0 \gamma p(1-\psi))}$. This is one mechanism that leads to Allee effects (see, for example, Allee, 1938; Boukal and Berec, 2002; Bessa-Gomes et al., 2004; Hurford et al., 2006, and references therein). Two examples of the changes in the expected net per capita growth rate at low population abundance are provided in Figure 5.2B.

5.3.3 The Allee effect

As discussed in the introduction, the Allee effect arises because individuals in the population, on average, can not replace themselves, and can be observed by plotting the expected net population change, where E$[\Delta N]$ = E$[N] - n_0$. The characteristic shape indicative of an Allee effect for the expected net population change is demonstrated in Figure 5.2A and is consistent with a deterministic formulation of the Allee effect (Lewis and Kareiva, 1993; Wang and Kot, 2001).

Allee effects may be strong or weak (Wang and Kot, 2001). A weak Allee effect is said to occur when the per capita growth rate or expected net population change is positive but decreasing as the initial population size decreases (Figure 5.2, gray lines). In contrast, a strong Allee effect (Figure 5.2, black lines) is said to occur when the per capita growth rate or expected net population change becomes negative for low population density. Only strong Allee effects induce an Allee threshold (E$[\Delta N] = 0$).

5.3.4 Demographic stochasticity

The probability mass function for N (Equation 5.6) allows for investigating the influence of demographic stochasticity in the presence of the Allee effect. Figure 5.3 contains an interpolated surface of masses. The probability space can be separated into three distinct classes, the probability of extinction, the probability

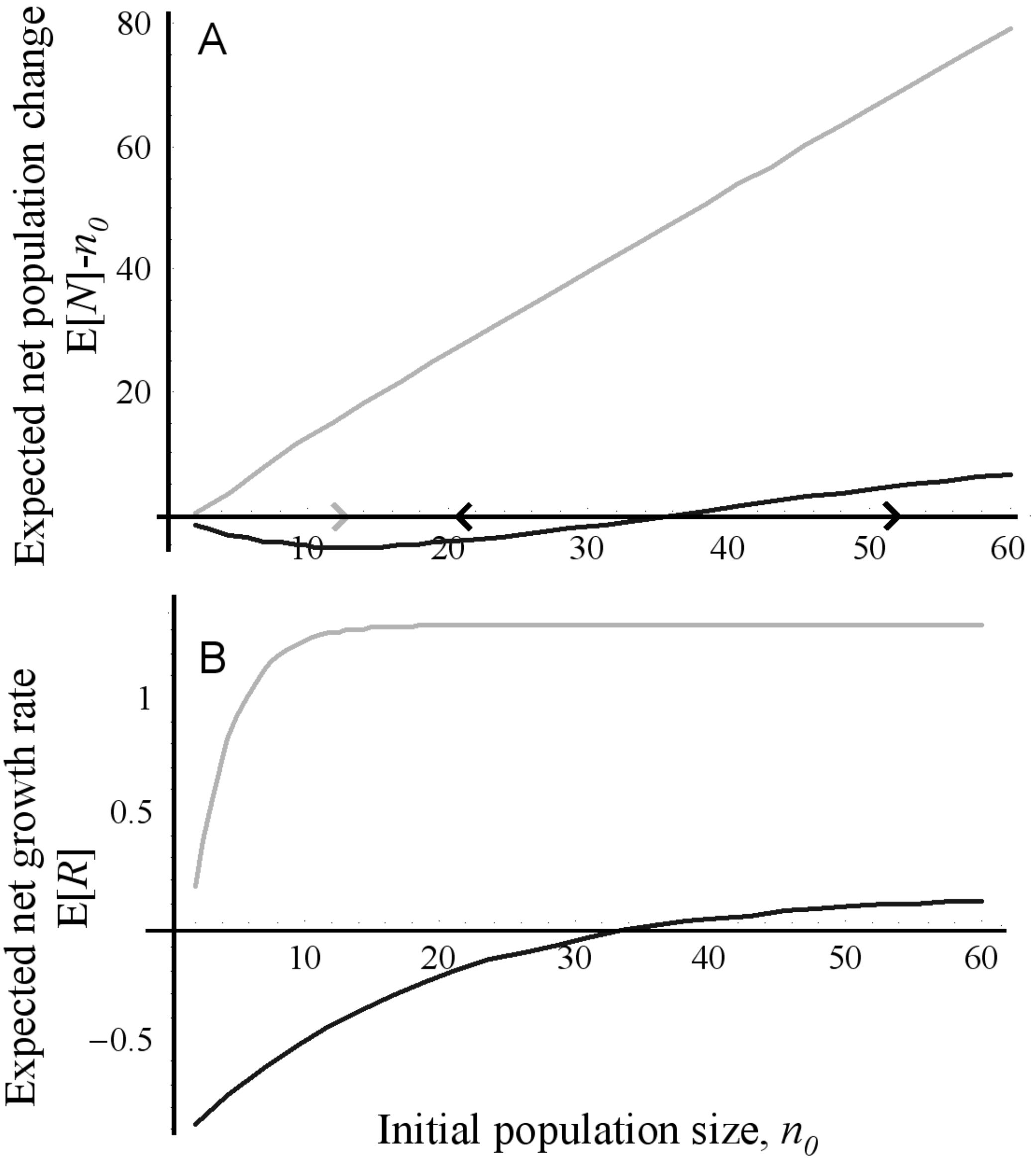


Figure 5.2: The expected net population change (Panel A) and the expected net geometric growth rate (Panel B), as a function of the initial population size, n_0, for two parameter sets. The black lines ($p = 0.8$,$v = 0.8$,$\psi = 0.3$,$\gamma = 0.1$, and $\beta = 6$) have negative values of E$[R]$ and E$[\Delta N]$, thus indicating a strong Allee effect. In deterministic models with Allee effects, n_0 values below E$[\Delta N] = 0$ (the Allee threshold) would go extinct. Values of n_0 above the threshold would lead to population persistence. In contrast, the gray lines ($p = 0.9$,$v = 0.9$,$\psi = 0.5$,$\gamma = 0.8$, and $\beta = 6$) demonstrate a weak Allee effect where the growth rate declines as n_0 decreases but is always positive. Since no Allee threshold exists, deterministic models with $n_0 \geq 2$ would imply population growth and persistence.

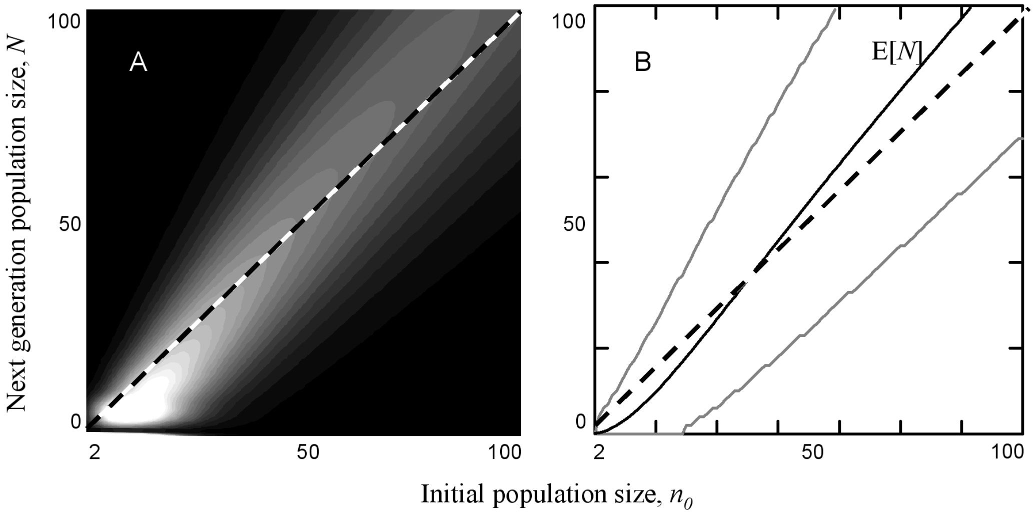


Figure 5.3: The interpolated probability surface of N (Panel A) and the expectation of N (Panel B) with $\approx 90\%$ prediction intervals (gray lines) as a function of the initial population size, n_0, for $p = 0.8$,$v = 0.8$,$\psi = 0.3$,$\gamma = 0.1$, and $\beta = 6$. The dashed line in both A and B is a 1 to 1 line ($N = n_0$). For small n_0 there is a high probability of individuals being unable to replace themselves ($N < n_0$). However, as n_0 increases the influence of the Allee effect is diminished, the expectation crosses the 1 to 1 line, and the probability that $N > n_0$ increases.

of persisting with N less than n_0 (below the dashed line in Figure 5.3A), and the probability of N being greater than n_0 (above the dashed line in Figure 5.3 A).

In deterministic settings with Allee effects, initial populations below an Allee threshold will go extinct while initial populations above the threshold will grow (Dennis, 1989; Wang and Kot, 2001). Stochastic models with Allee effects (Dennis, 2002), such as the one developed here, assign a probability to observing population growth due to chance events, such as demographic stochasticity in the region where deterministic models would indicate extinction.

5.4 Extending the system as a stochastic process

Previously, I have assumed the initial population size to be fixed and known; hence, in Equation 5.7, n_0 appears as a parameter. Projecting the distribution of subsequent generations of the population requires accounting for n_0 arising as random variable. This is a stochastic process of the population size, indexed by time. The probability model can be written as a dynamical system by formulating the distribution of the next generation through chains of conditional probability,

$$\text{Pr}(N_{t+1} = n_{t+1}) = \sum_{n_t=0}^{\infty} \text{Pr}(N_{t+1} = n_{t+1} \mid N_t = n_t)\, \text{Pr}(N_t = n_t). \tag{5.9}$$

The distribution of the next generation can be projected using a linear process,

$$\text{Pr}(N_{t+1}) = \mathbf{A}\ \text{Pr}(N_t). \tag{5.10}$$

In this linear process, the j^{th} element of the vector $\text{Pr}(N_t)$ is the probability of observing $N_t = j - 1$. The vector of probabilities for N_1 are calculated using Equation 5.7 for the initial population size, n_0. The length of this vector is set to $n_{max}+1$ to account for extinction, $n = 0$. The matrix $\mathbf{A}$ comprises $i = 1 \ldots n_{max}+1$ column vectors of conditional probability (Equation 5.9). Because I am projecting the population dynamics of a sexual species, columns $i = 1$ and $i = 2$ are absorbing states meaning that the probability of being extinct in the next time step is 1 if there are 0 or 1 individuals in the population (Taylor and Karlin, 1998). For the remaining

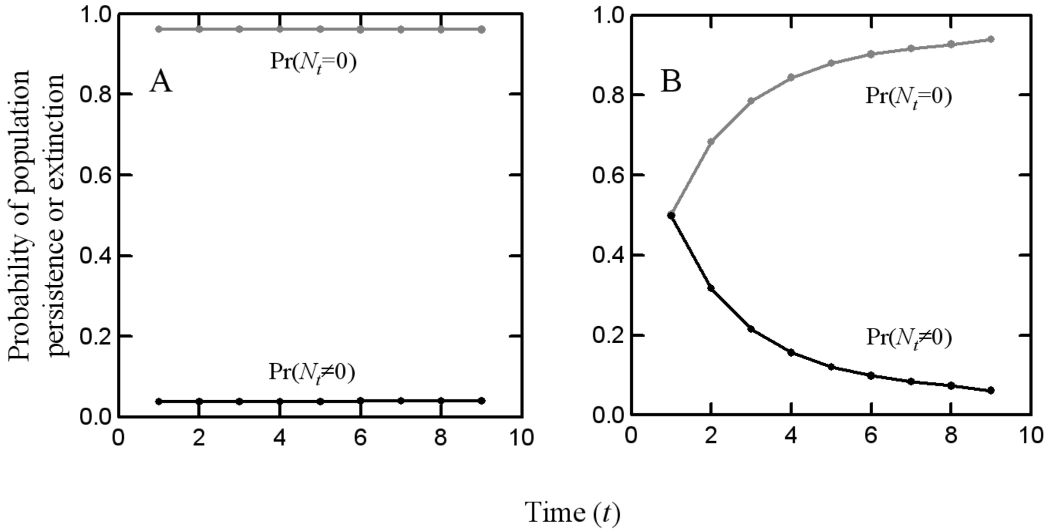


Figure 5.4: Two realizations of the stochastic process resulting from equation 5.10. Panel A is parameterized using the life history of the Chinese mitten crab (Table 5.3) and initialized at $n_0 = 400$. Panel B ($p = 0.8$,$v = 0.8$,$\psi = 0.3$,$\gamma = 0.1$, and $\beta = 6$) was initialized at $n_0 = 15$, approximately half the value of the Allee threshold (see Figure 5.2). For the mitten crab, the probability of establishing is very small (Pr($N_1 \neq 0$) $= 0.038$), but once established the population persists. In contrast the population in panel B experiences an increased probability of extinction through time.

$i = 2 \ldots n_{max} + 1$ columns, the elements, $a_{j,i}$, are calculated using Equation 5.7 where $n = j - 1$ and $n_0 = i - 1$, resulting in,

$$a_{j,i} = \begin{cases} G(0), n_0 = i - 1 & \text{when } j = 1, \\ \frac{c(i-1)}{j-1} \frac{\beta^{j-1}}{} \sum_{k=0}^{j-2} \frac{\Pr(N=j-2)}{\beta^k (j-2-k)!} & \text{when } j > 1. \end{cases} \tag{5.11}$$

The resulting conditional probability matrix is,

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & a_{1,3} & \cdots & a_{1,n_{max}+1} \\ 0 & 0 & a_{2,3} & \cdots & a_{2,n_{max}+1} \\ 0 & 0 & a_{3,3} & \cdots & a_{3,n_{max}+1} \\ 0 & 0 & a_{4,3} & \cdots & a_{4,n_{max}+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{n_{max}+1,3} & \cdots & a_{n_{max}+1,n_{max}+1} \end{bmatrix}. \tag{5.12}$$

The distribution of any future population at time t can be obtained by iterating Equation 5.10. Figure 5.4 provides two examples with different dynamics.

5.5 Examples

The Chinese mitten crab and the golden apple snail are two aquatic invertebrates listed on the World Conservation Union's (IUCN) "100 of the World's worst" list of invasive species (www.issg.org). Both species have invaded large regions of the world but are continuing to invade new locations and are of concern in North America. Both semelparous, sexually reproducing species have received attention because of their ecological and economical impacts and have a wealth of observational and laboratory research from which to obtain life history parameters.

5.5.1 Chinese mitten crab (*Eriocheir sinensis*)

The Chinese mitten crab (Figure 5.5 A), native to eastern and southern Asia, has successfully invaded parts of Europe, Japan, and the San Francisco and Chesapeake Bays of the United States (Rudnick et al., 2005; Herborg et al., 2006; Ruiz et al., 2006). Mitten crabs are imported as a food source, but live imports into the United States have been banned since 1989 and before their invasion into North America (Ruiz et al., 2006). The most likely introduction pathway is through the ballast water transport of larvae into estuaries (Herborg et al., 2007). Similar to other aquatic invaders, mitten crabs have changed estuary and stream food web dynamics, but the most noticeable ecological impact is to the erosion of stream banks by mitten crab burrows causing sedimentation into rivers and increased economic costs due to maintaining the structural integrity of levees (Rudnick et al., 2005).

Mitten crabs have a semelparous life history where the larvae are released in the brackish water of estuaries (Rudnick et al., 2005). The water temperature and salinity in which the larvae are released appear to be key determinants of survival (Zhang et al., 2001) and can produce juvenile survival of less than 0.1 across variable salinity and temperature. For analysis here, I use the probability of survival to megalopa stage under conditions of 12°C and salinity of 25 $\%S$ ($p = 0.01$: Anger, 1991). Individuals then migrate to freshwater streams where they remain until sexual maturity. Adults return to the estuaries and mate. Once in the estuary, mate finding can be difficult at low densities because mate recognition occurs by

Parameter	Value	Source
p	0.01	Anger (1991)
ψ	0.5	Zhang et al. (2001)
γ	0.01	Herborg et al. (2006)
v	≈ 1	Anger (2006)
β	250,000	Kobayashi (2001)

Table 5.3: Parameters values and sources used to parameterize the hierarchical probability model for the Chinese mitten crab, *Eriocheir sinensis*.

physical contact, as opposed to releasing a pheromone into the water which attracts males (Herborg et al., 2006). This would imply a very small probability of at least one male fertilizing a female, γ, for low population abundances but would almost certainly be determined by the geography of the estuary. I use $\gamma = 0.01$ for the analysis assuming that at low densities, males are unlikely to find a mate. Male mitten crabs die shortly after reproduction. However, females can brood the egg cluster for 1-2 months before release of the larvae, after which the female dies (Rudnick et al., 2005). The sex ratio appears to be 1.0 (Zhang et al., 2001) and the fecundity of Chinese mitten crabs is thought to be similar to the Japanese mitten crab with between 250,000 and 1 million eggs produced per female (Kobayashi, 2001). A generation for a Chinese mitten crab is between 2-3 years (Rudnick et al., 2005). The point estimates used for the model are provided in Table 5.3 along with the sources of information. Survival between mating and birthing is assumed to be approximately one because of the short time period (1-2 months) and the absence of information for this parameter value.

The parameter estimates for the Chinese mitten crab reveal a strong Allee effect (Figure 5.5 D; E$[\Delta N] = 0$ occurs when $n_0 \approx 16$, an Allee threshold). Maybe even more surprising is the range in initial population sizes (n_0) that result in high

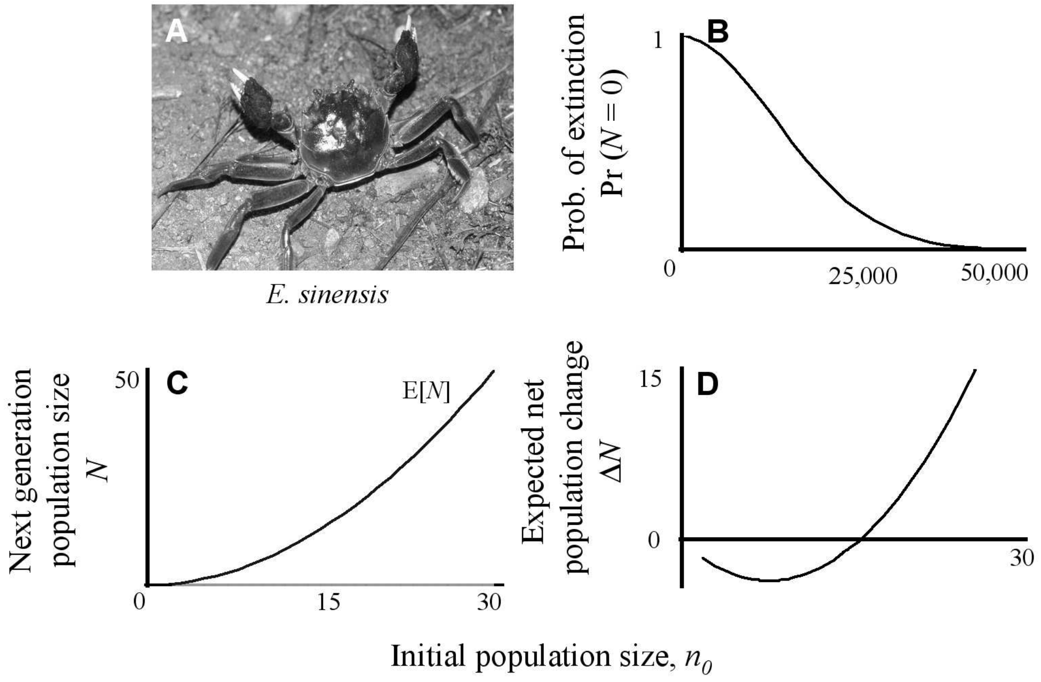


Figure 5.5: Trends in population statistics for the Chinese mitten crab. Adult mitten crab (Panel A) photo provided by Christian Fischer at Wikimedia Commons. The probability of extinction in the next generation is large for even thousands of individuals introduced (Panel B). However due to the large fecundity and increased probability that at least one fertilized female will survive, the expectation $\mathrm{E}[N]$ increases (Panel C, black line),and the lower 95% prediction boundary (gray line) is at $N = 0$. However, the expected net population change (Panel D) indicates the presence of a strong Allee effect.

probabilities of the population becoming extinct in the next generation (Figure 5.5B and C), a point I discuss further in an application of the stochastic process (Section 5.5.3).

Chinese mitten crabs have an interesting detection history. Individuals have been observed in some United States ports but only the San Francisco and Chesapeake Bays have established persistent populations (Ruiz et al., 2006). Although there have been individuals reported in several harbors, the possibility of Allee effects influencing the detection and establishment of mitten crabs has not been raised. The results from my model formulation indicates the Allee effect may be a reasonable explanation for these observations.

The parameter estimate of survival from birth to reproductive maturity, p, used for this analysis comes from laboratory experiments and is likely much lower in natural systems. As $p \rightarrow 0$, the Allee threshold expectedly increases. Of the parameters used in the analysis, the estimate of the probability of at least one male fertilizing a female, γ, is the most tenuous. Although Herborg et al. (2006) provided a detailed study of the biology of mitten crab mating, the factors affecting this probability are practically unknown.

5.5.2 Apple snail (*Pomacea canaliculata*)

The apple snail is native to South America and has established invasive populations in southern Asia, Japan, and the southern United States. Many of these establishments occurred because the snail was intentionally introduced to rice fields as a secondary crop for the escargot industry. However, apple snails are voracious herbivores, causing substantial loss in rice yield and changes in the state and function of freshwater lakes. In some freshwater systems, the native vegetation is completely extirpated resulting in lakes with clear water and food webs regulated by macrophytes changing states to turbid lakes dominated by phytoplankton (Carlsson et al., 2004, and references therein).

In the United States, apple snails were introduced in some locations as a biological control of invasive plants, such as *Hydrilla verticillata*. However, many of the introductions are thought to be a result of the aquarium trade where apple snails where released haphazardly. Although the aquarium industry has stopped the import and sales of apple snails, many small populations persist in private aquariums, and sightings of apple snails, such as in Lake Wawasee, Indiana where three individuals were found, are attributed to this pathway of introduction (Anon., 2005).

Apple snails have a variable life history including both semelparous and iteroparous strategies. Semelparous life histories are typically associated with the relatively constant temperatures in tropical and subtropical environments, taking approximately one year for a generation (Estebenet and Cazzaniga, 1992). Apple snails are sexual organisms with equal proportions of males and females in the population, although there appears to be a large amount of variability in the sex ratio that can be attributed to genetic relatedness of populations (Yusa, 2007). In laboratory experiments, survival from egg to sexual maturity and from sexual maturity to birthing is approximately 0.8 for each stage (Estebenet and Cazzaniga, 1992). In natural populations, the survival in both stages is likely substantially lower because snails are a common prey species for a wide array of predators, but snail egg masses are thought to be unpalatable, resulting in naturally high survival until the juveniles enter the water. Egg masses (Figure 5.6A) can vary from 100 to more than 1000 eggs with multiple egg masses being produced in a reproductive bout. In one study, the average fecundity was 4,506 eggs and ranged from 1,136 to 10,869 eggs (Estebenet and Martin, 2002).

The fertilization process of apple snails pairs with the fertilization process of the proposed model. Apple snails are promiscuous with females able to store sperm, resulting in a probability of fertilization very close to one (Estebenet and Martin, 2002). It has been reported that males are able to find females at long distances, at low density, and in the dark, suggesting a possible chemotactic mechanism. Even at experimentally controlled low densities, almost all females are fertilized (Albrecht et al., 1996).

Although the values for all parameters are not available for a single study site, the estimates for survival, fecundity, sex ratio, and probability of being fertilized appear to be fairly consistent across studies. Table 5.4 contains the point estimates

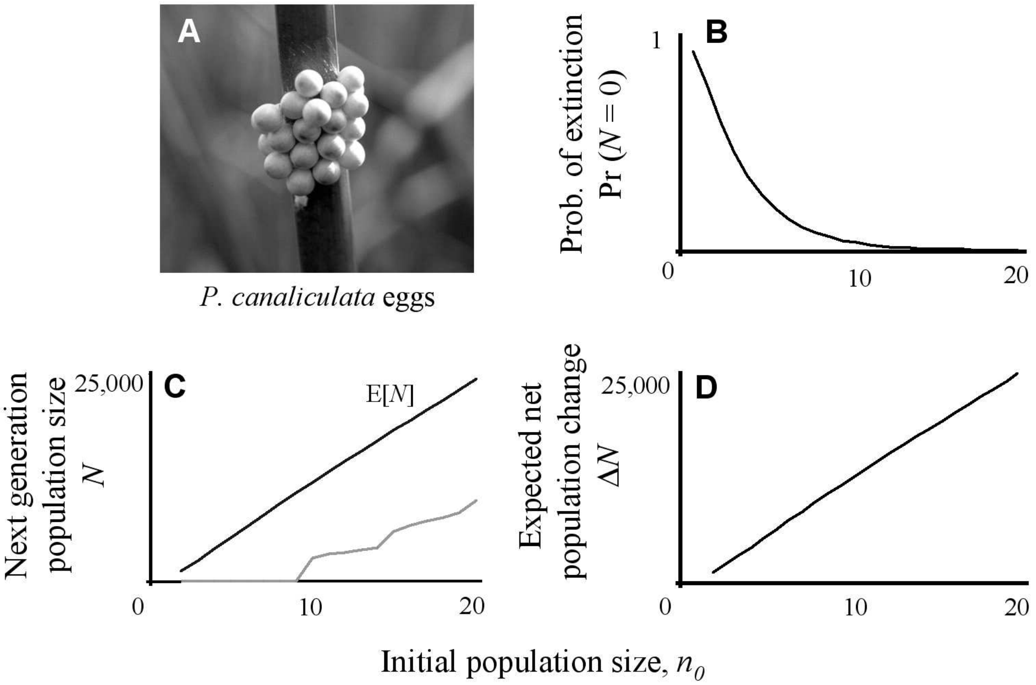
Parameter	Value	Source
p	0.8	Estebenet and Cazzaniga (1992) Estebenet and Martin (2002)
ψ	0.5	Yusa (2007)
γ	≈ 1	Albrecht et al. (1996)
v	0.8	Estebenet and Cazzaniga (1992) Estebenet and Martin (2002)
β	4,500	Estebenet and Martin (2002) Albrecht et al. (1996)

Table 5.4: Parameters values and sources used to parameterize the hierarchical probability model for the apple snail, *Pomacea canaliculata*.

used to parameterize the model.

The probability of extinction (B), the population size with lower approximately 95% prediction bound (C), and the net population change (D) at a single time step for the apple snail are plotted in Figure 5.6 as a function of the initial population size, n_0, using the parameters found in Table 5.4. For as few as 20 individuals released into a location, the probability of the population being extinct in the next generation is approximately zero. For all n_0, the expected growth rate is positive, although the expected net population change does decline and thus indicates week Allee effects.

For introductions of greater than 10 individuals, it is almost a certainty that establishment of an apple snail population will occur. This observation is of particular concern given the introduction pathway of private aquariums being exposed to freshwater lakes and likely explains the disparate observations in the locations where apple snails have successfully established. These systems were simply overwhelmed by propagule pressure (Von Holle and Simberloff, 2005). Apple snail population dynamics do not contain an Allee threshold, and prevention



P. canaliculata eggs

Figure 5.6: Trends in population statistics for the apple snail. Egg mass photo (Panel A) provided by Gary M. Stolz, U. S. Fish and Wildlife Service (Wikimedia Commons). In contrast to the mitten crab, the probability extinction for the apple snail is approximately zero by $n_0 = 14$ (Panel B). This is further reflected in the expected next generation population size (Panel C, black line). The approximately 95% lower prediction bound (Panel C, gray line) jumps to over $N = 3400$ when $n_0 = 10$, reflecting the low probability of observing $N = 0$ when $n_0 \geq 10$. The expected net population change (Panel D) indicates a weak Allee effect as it is positive everywhere, but declines as n_0 declines.

of apple snail invasion should focus on zero propagule pressure strategies rather than reactive, eradication efforts (Simberloff, 2003*a*).

Of the parameter estimates used, the survival is likely the most susceptible to changes, due to biotic interactions (predation). However, reducing $p = 0.1$ and $v = 0.1$ results in the same weak Allee effect, although the probability of extinction in the next time step is large (> 0.8) for values of $n_0 \leq 446$, a threshold that is not likely above the population size of apple snails in many personal aquariums. These results indicate that natural system and predation studies of apple snails are critical for assessing invasion risk. This has been noted previously (Estebenet and Martin, 2002).

5.5.3 Stochastic process of establishment

There is an interesting observation in the dynamics of both species, but in particular the mitten crab. That is, how can a species with an expected positive growth rate (Figure 5.5C) after the Allee threshold ($n_0 = 16$) have such a large probability of being extinct in the next time step (Figure 5.5B)? With the presence of a weak or strong Allee effect, the successful establishment of populations has been shown to largely depend on population dynamics occurring after the initial introduction (Drake et al., 2005). This appears not to be so for either the apple snail or Chinese mitten crab and can be observed in the stochastic process (Equation 5.9) and resulting probability distributions of future generations (Equation 5.10).

For illustration, consider an initial population size of Chinese mitten crab, $n_0 = 400$, has a probability of being extinct in the next generation, Pr$(N_1 = 0)$, of 0.962. For N_2 to N_9, this probability remains essentially unchanged (Figure 5.4). That is to say, there is a probability $1 - \text{Pr}(N_1 = 0)$ the population will persist at $t = 1 \ldots 9$, that is approximately constant after the introduction. This 'boom or bust' population dynamic occurs because of the small probability at n_0 of a surviving fertilized female, H, and the very large expected fecundity, β, where the number of offspring becoming the next generation overwhelm the system such that the probability of the N_2 generation being extinct is ≈ 0. This same dynamic is observed in the apple snail.

More generally, some NIS may have lower fecundities resulting in Allee effects that contribute to the probability of being extinct in future generations. For these situations, the probability distributions resulting from the stochastic process (Equation 5.10) are useful for estimating the probability of the population being extinct within some time, t. This analysis is equivalent to performing population viability analysis common in conservation biology (see, for example Brook et al., 2000) and will likely have the same predictive shortcomings (Ellner et al., 2002).

For many pathways of invasion, there are repeated introductions that contribute to the increased probability of successful invasion (Carlton, 1996; Cohen and Carlton, 1998). Even though apple snails and Chinese mitten crab have weak and strong Allee effects, respectively, the boom–bust dynamics of establishment fit well into the invasion waiting time formulation of invasion risk (Chapter 2; Jerde and Lewis, 2007). The invasion waiting time formulation assumes that each introduction event is independent of previous introduction events and has the same propagule pressure (n_0) leading to identically distributed probabilities of establishment, $1 - \text{Pr}(N = 0)$, where $\text{Pr}(N = 0) = G(0) \mid n_0$ from Equation 5.7 and n_0 is the propagule pressure. This implies that at each time step, n_0 individuals are introduced to a location. For a specified, regularly recurring propagule pressure, the probability of an invasion by time t is,

$$\text{Pr}(T = t) = 1 - (1 - G(0) \mid n_0)^t, \tag{5.13}$$

(Chapter 2; Jerde and Lewis, 2007).

Equation 5.13 is the cumulative mass function of a geometric waiting time distribution. The probability of Chinese mitten crab establishment when $n_0 = 400$ is $1 - 0.962 = 0.038$. Therefore the probability of establishment of a population after $t =$ 5, 10, and 100 generations is 0.18, 0.32, and 0.98, respectively. Consequently, the semelparous life history strategy with overwhelming fecundity further supports performing ecological risk assessment based on propagule pressure (Lockwood et al., 2005) and proactive risk management even in the absence of well understood population dynamics (Simberloff, 2003*b*).

5.6 Discussion

Demographic stochasticity and the Allee effect are known to decrease the persistence of populations at low population density (Shaffer, 1981; Dennis, 2002), be they managed, endangered, or invasive species. The presence of an Allee effect and stochasticity in a population has been shown to change management actions, optimal control decisions, and risk assessments (Liebhold and Bascompte, 2003; Taylor and Hastings, 2005; Drake and Lodge, 2006). Yet assessing the presence and strength of an Allee effect, particularly for invasive species, has a small literature (Taylor and Hastings, 2005, and references therein), these assessments have been conducted only after the species has successfully established into multiple locations under different propagule pressures (Leung et al., 2004) or by using population data of established species (Veit and Lewis, 1996). As a consequence, risk assessment and risk management is commonly practiced without a management plan where an Allee effect is considered (Simberloff, 2003*b*). I have proposed a hierarchical probability model of a semelparous, sexual species that can assess the strength of an Allee effect prior to invasion and that directly connects to ecological risk assessment of invasion.

Stephens et al. (1999) define a demographic Allee effect as the overall fitness being positively correlated with density such that the per capita growth rate decreases with reduced population density. Through the expected net per capita growth rate, $\mathrm{E}[R]$, and expected net population change, $\mathrm{E}[\Delta N]$, it is possible to observe these trends. However, the model will always contain an Allee effect because the expected net growth rate, $\mathrm{E}[R] = \frac{\beta(n_0 v p \psi(1-e^{-n_0 \gamma p(1-\psi)}))}{n_0} - 1$, necessarily decreases as $n_0 \rightarrow 0$. Therefore the question of interest is not whether an Allee effect exists, but rather what is the strength of the Allee effect and does it influence the dynamics of the population?

The examples of apple snails and Chinese mitten crabs show the difference between weak (Figure 5.6C) and strong (Figure 5.5C) Allee effects, respectively. However, the influence of the Allee effect, strong or weak, does not change the fate of establishing populations of these two species. The statistics used to identify the

Allee effect are expectations and in both examples the probability of extinction is very high (≈ 1 at small n_0). The expectations are therefore driven by the probability of at least one fertilized female surviving, H, and producing offspring with an average fecundity, β. Therefore, the presence of a strong or weak Allee effect, indicated by $\mathrm{E}[R]$ and $\mathrm{E}[\Delta N]$, may not be a critical consideration for predicting some NIS.

Indeed, large fecundity may be the characteristic that has made apple snails and Chinese mitten crabs successful invaders (Keller et al., 2007), but some invasive species may not benefit from such large fecundities. For these species, the dynamics may be critical for assessing the probability that the invasive species will persist (Dennis, 2002; Drake et al., 2005). As previously mentioned, projecting the population abundance with the stochastic process is essentially a population viability analysis approach to estimating the risk of NIS establishment. As with PVAs for conservation biology, the population abundance predictions for subsequent generations become increasingly dispersed such that prediction envelopes are very wide even after a few generations (Ellner et al., 2002). Predicting population abundance after even short periods (such as 4 or 5 generations) is difficult.

Risk management will seek to adjust parameters to induce a strong Allee effect from a weak Allee effect, increase the strength of a strong Allee effect if one is already present, and reduce the probability of successful establishment. In the mitten crab, decreasing survival will increase the strength of the Allee effect. Similarly, decreasing survival in the apple snail, for example through increased predation, can increase the probability of extinction. Performing sensitivity analysis (see, for example Caswell, 2001) on the model parameters or by evaluating the change in the probability of extinction provides an approach for exploratory risk management for potential invaders.

Bibliography

Albrecht, E., N. Carreno, and A. Castro-Vazquez. 1996. A quantitative study of copulation and spawning in the South American apple-snail *Pomacea canaliculata* Lamarck (Prosobranchia, Ampullariidae). Veliger **39**:142–147.

Allee, W. 1938. The social life of animals. W.W. Norton and Co., New York, NY, USA.

Anger, K. 1991. Effects of temperature and salinity on the larval development on the Chinese mitten crab *Eriocheir sinensis* (Decapoda: Grapsoidea). Marine Ecology Progress Series **72**:103–110.

Anger, K. 2006. Contributions of larval biology to crustacean research: a review. Invertebrate Reproduction and Development **49**:175–205.

Anon., 2005. Aquatic invasive species: channeled apple snail. Electronic, Indiana Department of Natural Resources.

Bessa-Gomes, C., S. Legendre, and J. Clobert. 2004. Allee effects, mating systems and the extinction risk in populations with two sexes. Ecology Letters **7**:802–812.

Boukal, D., and L. Berec. 2002. Single-species models of the Allee effect: extinction boundaries, sex ratios, and mate encounters. Journal of Theoretical Biology **218**:375–394.

Brook, B., J. O'Grady, A. Chapman, A. H. Burgman, M., and R. Frankham. 2000. Predictive accuracy of population viability analysis in conservation biology. Nature **404**:385–387.

Carlsson, N., C. Bronmark, and L. Hansson. 2004. Invading herbivory: the golden apple snail alters ecosystem functioning in Asian wetlands. Ecology **85**:1575–1580.

Carlton, J. 1996. Pattern, process, and prediction in marine invasion ecology. Biological Conservation **78**:97–106.

Case, T. 2000. An illustrated guide to theoretical ecology. Oxford University Press, New York, NY, USA.

Caswell, H. 2001. Matrix Population Models. 2nd edition. Sinauer.

Chatfield, C., and C. Theobald. 1973. Mixtures and random sums. The Statistician **22**:281–287.

Cohen, A., and J. Carlton. 1998. Accelerating invasion rate in a highly invaded estuary. Science **279**:555–558.

de Camino Beck, T., and M. Lewis. 2007. A new method for calculating net reproductive rate from graph reduction with applications to the control of invasive species. Bulletin of Mathematical Biology **69**:1341–1354.

Dennis, B. 1989. Allee effects: population growth, critical density, and the chance of extinction. Natural Resource Modeling **3**:481–538.

Dennis, B. 2002. Allee effects in stochastic populations. Oikos **96**:389–401.

Drake, J., P. Baggenstos, and D. Lodge. 2005. Propagule pressure and persistence in experimental populations. Biology Letters **1**:480–483.

Drake, J., and C. Jerde, In Press. Bioeconomics of Invasive Species: Integrating Ecology, Economics, Policy, and Management, Chapter stochastic models of propagule pressure and establishment . Oxford University Press.

Drake, J., and D. Lodge. 2006. Allee effects, propagule pressure and the probability of establishment: risk analysis for biological invastions. Biological Invasions **8**:365–375.

Drury, K., J. Drake, D. Lodge, and G. Dwyer. In Press. Immigration events dispersed in space and time: factors affecting invasion success. Ecological Modelling .

Ellner, S., J. Fieberg, D. Ludwig, and C. Wilcox. 2002. Precision of population viability analysis. Conservation Biology **16**:258–261.

Engen, S., O. Bakke, and A. Islam. 1989. Demographic and environmental stochasticity – concepts and definitions. Biometrics **54**:840–846.

Estebenet, A., and N. Cazzaniga. 1992. Growth and demography of *Pomacea canaliculata* (Gastropoda: Ampullariidae) under laboratory conditions. Malacological Review **25**:1–12.

Estebenet, A., and P. Martin. 2002. *Pomacea canaliculata* (Gastropoda: Ampullariidae): Life-history traits and their plasticity. Biocell **26**:83–89.

Goodman, D. 1987. Considerations of stochastic demography in the design and management of reserves. Natural Resource Modeling **1**:205–234.

Herborg, L., M. Bentley, A. Clare, and K. Last. 2006. Mating behavior and chemical communication in the invasive Chinese mitten crab *Eriocheir sinensis*. Journal of Experimental Marine Biology and Ecology **329**:1–10.

Herborg, L., C. Jerde, D. Lodge, G. Ruiz, and H. MacIsaac. 2007. Predicting the North American distribution of Chinese mitten crabs (*Eriocheir sinensis*) using measures of propagule pressure and environmental niche models. Ecological Applications **17**:663–674.

Hurford, A., M. Hebblewhite, and M. Lewis. 2006. A spatially explicit model for an Allee effect: why wolves recolonize so slowly in Greater Yellowstone. Theoretical Population Biology **70**:244–254.

Jerde, C., and M. Lewis. 2007. Waiting for invasions: a framework for the arrival of non–indigenous species. American Naturalist **170**:1–9.

Keller, R., J. Drake, and D. Lodge. 2007. Fecundity as a basis for risk assessment of nonindigenous freshwater molluscs. Conservation Biology **21**:191–200.

Kobayashi, S. 2001. Fecundity of the Japanese mitten crab *Eriocheir japonica* (de Haan). Benthos Research **56**:1–7.

Kolar, C., and D. Lodge. 2001. Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution **16**:199–204.

Kot, M., M. Lewis, and P. van den Driessche. 1996. Dispersal data and the spread of invading organisms. Ecology **77**:2027–2042.

Lande, R. 1993. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. American Naturalist **142**:911–927.

Leung, B., J. Drake, and D. Lodge. 2004. Predicting invasions: propagule pressure and the gravity of Allee effects. Ecology **85**:1651–1660.

Lewis, M., and P. Kareiva. 1993. Allee dynamics and the spread of invading organisms. Theoretical Population Biology **43**:141–158.

Liebhold, A., and J. Bascompte. 2003. The Allee effect, stochastic dynamics and the eradication of alien species. Ecology Letters **6**:133–140.

Lockwood, J., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. Trends in Ecology and Evolution **20**:223–228.

Odum, H., and W. Allee. 1954. A note on the stable point of populations showing both intraspecific cooperation and disoperation. Ecology **35**:95–97.

Regan, T., M. McCarthy, P. Baxter, F. Panetta, and H. Possingham. 2006. Optimal eradication: when to stop looking for an invasive plant. Ecology Letters **9**:759–766.

Rudnick, D., T. Veldhuizen, R. Tullis, C. Culver, K. Hieb, and B. Tsukimura. 2005. A life history model for the San Francisco estuary population of the Chinese mitten crab, *Eriocheir sinensis* (Decapoda: Grapsoidea). Biological Invasions **7**:333–350.

Ruiz, G., L. Fegley, P. Fofonoff, Y. Cheng, and R. Lemaitre. 2006. First records of *Eriocheir sinensis* H. Milne Edwards, 1853 (Crustacea: Brachyura: Varunidae) for Chesapeake Bay and the mid-Atlantic coast of North America. Aquatic Invasions **1**:137–142.

Sakai, A., F. Allendorf, J. Holt, D. Lodge, J. Molofsky, K. With, S. Baughman, R. Cabin, J. Cohen, N. Ellstrand, D. McCauley, P. O'Neil, I. Parker, J. Thompson, and S. Weller. 2001. The population biology of invasive species. Annual Reviews of Ecology and Systematics **32**:305–32.

Shaffer, M. 1981. Minimum population sizes for species conservation. BioScience **31**:131–134.

Simberloff, D. 2003*a*. Eradication – preventing invasions at the outset. Weed Science **51**:247–253.

Simberloff, D. 2003*b*. How much information on population biology is needed to manage introduced species? Conservation Biology **17**:83–92.

Stephens, P., W. Sutherland, and F. R. 1999. What is the Allee effect? Oikos **87**:185–190.

Taylor, C., and A. Hastings. 2005. Allee effects in biological invasions. Ecology Letters **8**:895–908.

Taylor, H., and S. Karlin. 1998. An Introduction to stochastic modeling. 3rd edition. Academic Press, San Diego, Calf.

Veit, R., and M. Lewis. 1996. Dispersal, population growth, and the Allee effect: dynamics of the house finch invasion of eastern North America. American Naturalist **148**:255–274.

Von Holle, B., and D. Simberloff. 2005. Ecological resistance to biological invasion overwhelmed by propagule pressure. Ecology **86**:3212–3218.

Wang, M., and M. Kot. 2001. Speeds of invasion in a model with strong and weak Allee effects. Mathematical Biosciences **171**:83–97.

Williamson, M. 1999. Invasions. Ecography **22**:5–12.

Yusa, Y. 2007. Nuclear sex-determining genes cause large sex-ratio variation in the apple snail, *pomacea canaliculata*. Genetics **175**:179–184.

Zhang, T., Z. Li, and Y. Cui. 2001. Survival, growth, sex ratio, and maturity of the Chinese mitten crab (*Eriocheir sinensis* reared in a Chinese pond.). Journal of Freshwater Ecology **16**:633–640.

Chapter 6

Synthesis

6.1 Summary

In the preceding four chapters, I have proposed a stochastic modeling approach for ecological risk assessment of non-indigenous species (NIS). In Chapter 2, I formulated a hierarchical model of NIS arrival to new locations and assess the risk from arrival under repeated introductions as a stochastic process resulting in invasion waiting times. Additionally, I showed how invasion waiting times can be reduced to relative waiting times, and I applied the relative formulation to scentless chamomile and Chinese mitten crabs. In Chapter 3, I manipulated the arrival of scentless chamomile to two locations, and I showed that waiting time distributions predicted from the estimated survival of the population matched the observed, spatially–replicated waiting time distributions. This study supports a stochastic modeling approach for ecological risk assessment. Chapter 4 connected invasion waiting times, using the relative formulation, to the gravity model approach commonly used for approximating the propagule pressure. Along with uncovering sources of uncertainty in the national zebra mussel gravity model (Bossenbroek et al., 2007), I demonstrated established methods for evaluating predictive performance (ROC and AUC), and I proposed model validation techniques that arise from the relative invasion waiting time. In Chapter 5, I developed a hierarchical probability model for a semelparous, sexual species being introduced into a new location, and I demonstrated the emergence of an Allee effect in the expected net geometric per capita growth rate. Further I showed that, although the dynamics indicated the existence of strong and weak Allee effects for the Chinese mitten crab and apple snail, respectively, NIS risk assessment can be simplified to calculations of the probability that one or more fertilized females survive to give birth. As a consequence, the invasion waiting time formulation, presented in Chapter 2, can also be used for some populations with Allee effects.

In what follows, I take a broader, critical, and sometimes speculative, approach to the ideas and work presented in the previous four chapters. The purpose of this discourse is to: 1. demonstrate how my dissertation merges or complements existing approaches of NIS risk assessment, 2. distinguish where my dissertation advances NIS risk assessment, and 3. identify where future invasive species research needs to proceed. I structure this discussion by dissecting the arrival and establishment stages of the invasion process (Sakai et al., 2001), critiquing the application of risk assessment for NIS (Suter, 1993), and connecting my work to applications in risk management and conservation biology (D'Antonio et al., 2001; Lodge et al., 2006).

6.2 The invasion process and NIS risk assessment

As discussed in the introduction, the invasion process is often discritized into arrival, establishment, and spread stages (Vermeij, 1996; Sakai et al., 2001). My dissertation advances our understanding of the interaction between arrival and establishment for assessing invasion risk and contributes to addressing the concern that little theoretical or empirical work in invasion biology has been conducted on arrival (Puth and Post, 2005). I discuss the invasion process in reverse order for reasons following shortly, starting with establishment followed by arrival.

6.2.1 Establishment

Establishment has many definitions and suffers from linguistic uncertainty (Table 6.2). That is to say, establishment definitions contain debatable thresholds as to how many individuals and how long those individuals are present in a new landscape (Regan et al., 2002). The differences in abundance thresholds are noticeable in experimental studies (E in Table 6.2). For example, Bossenbroek et al. (2001)

Definition	Citation	Overview or Experiment
The presence of at least one individual in a location for one discrete time step	Chapter 2 and 3, Jerde and Lewis (2007), Levine et al. (2004)	O and E
A measured population having more than 10 individuals present	Drake and Lodge (2006)	E
The detected presence of the species	Chapter 4, Bossenbroek et al. (2001), Leung et al. (2004), MacIsaac et al. (2004)	E
A viable self-sustaining population	Sakai et al. (2001)	O
A species with a self-sustaining population outside of its native range	Kolar and Lodge (2001), Puth and Post (2005)	O
A new population that can sustain itself through local reproduction and recruitment, which thus augments or replaces dispersal from the donor region as a means for the population's persistence	Vermeij (1996)	O
Survival to form a reproducing population	Wonham et al. (2000), Carlton (1996)	O and E
A fixed period of time over which the invader is present	Chapter 5,Taylor and Hastings (2005)	O
Survival and growth of at least one individual	Von Holle and Simberloff (2005)	E

Table 6.1: The definition of establishment is variable. However there is a trend in that experimental applications (E) must define a population threshold to define what it means to establish. Alternatively, overviews (O) of invasion biology and conceptual frameworks emphasize population persistence.

built a gravity model for the spread of zebra mussel (*Dreissena polymorpha*) and use presence-absence data regarding the invasion of lakes. If zebra mussels are detected in a lake, then they are said to have established, as population abundances with positive growth are likely below the detection threshold. But this can lead to biased detection errors where lakes are invaded but not yet identified as invaded. Alternatively, Drake and Lodge (2004) used 281 populations of *Daphnia magna* to assess the role of environmental variation on establishment. Establishment in this study was defined as the presence of greater than 10 individuals in a population. Moreover, many theoretical considerations (O in Table 6.2) of biological invasions consider persistence of a population in defining establishment. Such is the case with the stochastic model presented in Chapter 5. But persistence faces the same threshold problems encountered with abundance based definitions of establishment in that the invader must be present for a defined period of time in order to be considered established. And again, this threshold is somewhat arbitrary. The uncertainty of establishment is further compounded in some definitions by population properties, such as viability, reproduction, and dispersal that must also occur while the invader is persisting in the new environment.

It is reasonable to believe that all of the Table 6.2 definitions are operationally useful for the study of biological invasions, and therefore it is not likely, nor particularly useful (Regan et al., 2002), that a consensus is formed on what exactly it means to establish. Additionally, in many of the studies, the definition of establishment is limited by the data available, such as in the case with the presence or absence of zebra mussels. Regan et al. (2002) suggest a number of approaches, such as supervaluations and fuzzy logic, to address the vagueness in threshold related problems, but the simplest suggestion is to provide clear definitions of establishment.

My dissertation has variable establishment definitions (Table 6.2). Chapters 2 and 3 have the strictest definition, that is the presence of one or more individuals surviving to reproductive maturity. In contrast, Chapter 4 uses a practical definition of establishment tied to detected presence or absence of an established zebra mussel population in a lake, and Chapter 5 allows for considering the probability of population persistence over a given time interval. I propose that any one of these definitions is reasonable for a particular study, species, or location. However, a risk assessment approach to biological invasions imposes rigor in that performing a risk assessment requires clearly defining an endpoint, a quantifiable expression of the environmental value at risk (Suter, 1993), on which to calculate a probability. The definition of establishment is this expression. As a consequence, invasive species risk assessment, either explicitly or implicitly, must consider establishment, even if the focus of the research is on propagule pressure.

6.2.2 Arrival

The study of arrival is essentially the study of propagule pressure (Lockwood et al., 2005), which is a critical consideration in estimating the risk of species invasions (Drake and Lodge, 2004; Von Holle and Simberloff, 2005). The framework I devised in Chapter 2 highlights that NIS management can intervene to reduce propagule pressure by targeting the uptake of individuals at source locations (Floerl and Inglis, 2005), expelling individuals during the transport (such as ballast water exchange Wonham et al., 2005*b*), or by reducing the release of individuals at the destination (following Carlton, 1996). My formulation of invasion risk is a simple model of the propagule pressure and survival, and does not explicitly consider the population dynamics after introduction (but see Chapter 5). This is also consistent with other invasion models based on alternative ecological theory (Tilman, 2004).

The simplest risk assessment approach, emphasizing arrival with an implicit assumption regarding establishment, is the generation of dose response curves (Lockwood et al., 2005; Drake and Lodge, 2006). From eco-toxicology, a dose is a concentration of a chemical released into the environment (Suter, 1993), but for invasive species, dose refers to the concentration of individuals, or propagule pressure, entering a new location. Given a dose of invaders, a logical response to measure is establishment (Drake and Lodge, 2006). However, as demonstrated in Table 6.2, establishment carries multiple definitions, which results in various dose-response curves for the same data (Figure 6.1). For example establishment may be tied to the number of individuals establishing (Figure 6.1A), the expected number of individuals establishing (Figure 6.1B), or the probability of a specified number of individuals establishing (Figures 6.1 C and D).

Often, dose response curves are phenomenological descriptions of data. Figures 6.1B and C show spline curves (black lines) applied to the establishment of scentless chamomile. Splines are descriptions of the observed data without a mechanistic basis (Suter, 1993) and do not provided a causal explanation between the propagule pressure and establishment. However, phenomenological dose-response curves may be useful for risk assessment of invasive species when details of the population dynamics are unknown but observational data are available. Other statistical methods such as logistic regression may provide useful descriptions of observed patterns in the data for testing the effect of control treatments, again when the underlying establishment dynamics are unknown (see, for example, Bossenbroek et al., 2001).

Mechanistic models imply a causal relationship between the propagule pressure and establishment and are preferred for two reasons. First, mechanistic formulations help us to identify the general mechanisms (or characteristics) that allow species to successfully establish or fail (Kolar and Lodge, 2001). Secondly, mechanistic models of establishment can be used to inform management about parameters that may be targeted to reduce the probability of establishment (e.g. Wonham et al., 2006). In contrast, phenomenological models are essentially limited to targeting the only input, propagule pressure. As an example, the gray lines in Figure 6.1 represent simple mechanistic models using the probability of survival for a single seed of scentless chamomile (p_s=0.085) estimated from all observations. Panels A and B have a model of the expectation of the number of establishing individuals (Y) as a function of the propagule pressure (ϕ) when I assume the seeds are independent and identically distributed. Using a Poisson error distribution, the expected number of establishing individuals is 0.085ϕ. Panels C and D (Figure 6.1) use the probability of at least one individual establishing, $\Pr(Y \geq 1) = 1 - \Pr(Y = 0) = 1 - e^{-0.085\phi}$, and the probability of at least five individuals establishing, $\Pr(Y \geq 5) = 1 - (\Pr(Y = 0) + \ldots + \Pr(Y = 5))$, respectively. Although the establishment of scentless chamomile can be formulated with a dose response

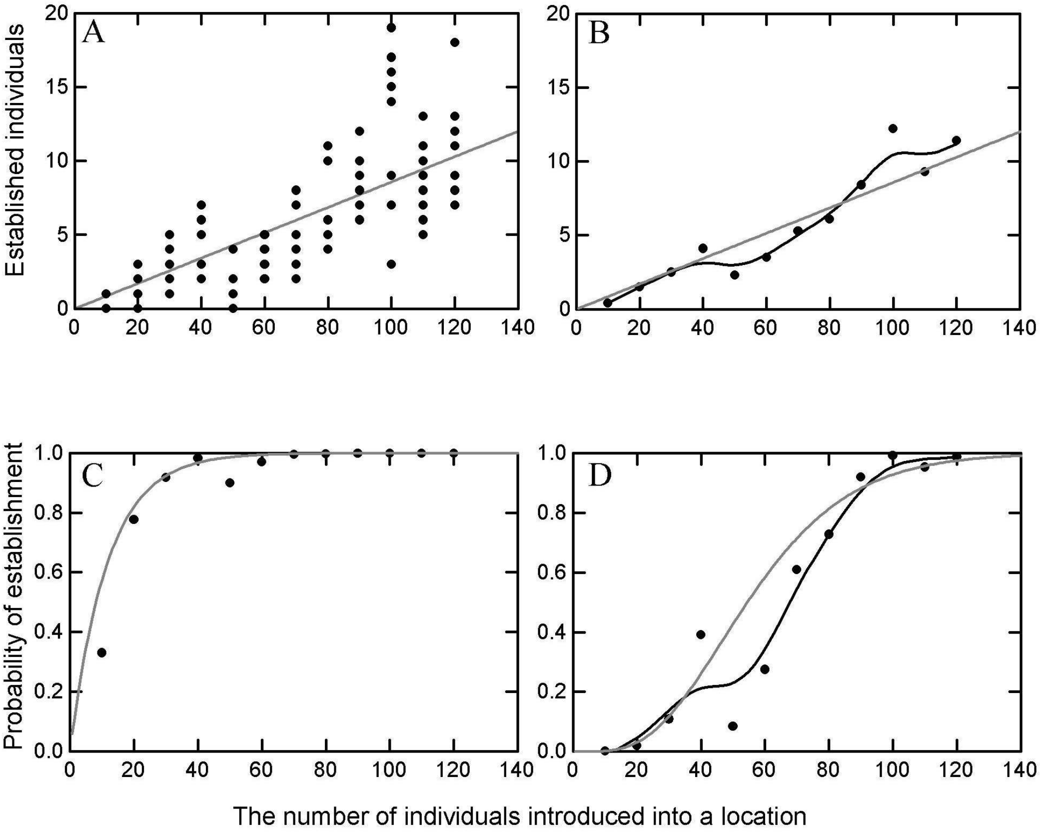


Figure 6.1: Four different dose-response curves for the establishment of the invader scentless chamomile (*Matricaria perforata*). (Panel A) The data of the number of established individuals given the propagule pressure. For each dose there are 10 observations. The individual probability of survival across all doses is 0.085. The response curve is the expectation given the propagule pressure and the individual probability of survival. (Panel B) The expectation of each dose shows has the same trend. (Panel C) Alternatively, if a Poisson distribution is fit to each dose, then the probability of at least one individual establishing can be estimated. The model fit comes from using the individual probability of survival for different doses of individuals. (Panel D) The probability of 5 or more individuals establishing produces a sigmoid dose response curve of the establishment process. The gray lines come from modeling the expectations and probabilities and the dark lines are spline fits to the data.

approach, the general role of survival in the establishment process is apparent, thus revealing a mechanistic bases for the formulation (Chapters 2 and 5).

6.2.3 Gravity models

The relative risk formulation in Chapter 2 was developed in part to link gravity model output to risk assessment and uncover the assumptions needed to make this link. Previously, gravity scores were assessed using logistic regression in a dose response curve framework (see, for example, Bossenbroek et al., 2001). Because the gravity scores significantly explained the presence or absence of successful lake invasions, it has been argued that gravity models are good predictors of invasions (Leung et al., 2004, 2006). However, explanation does not predicate accurate predictions (Breiman, 2001), and when I applied ROC and AUC diagnostics of predictive performance to the gravity scores, the national gravity model of zebra mussel dispersal (Bossenbroek et al., 2007) performed poorly. As a consequence, I am sceptical of gravity model estimates of propagule pressure and the use of logistic regression to justify the gravity model approach. In Chapter 4, I proposed an alternative approach for validating predictions using the relative risk formulation in Chapter 2 and the frequency of invasions for lakes with similar gravity scores, but this approach could not be tested because of the small population of lakes and gravity scores available to me for the analysis. However, the approach is motivating further developments (*Personal communication* J. Bossenbroek) for predicting and validating the dispersal of emerald ash borer (*Agrilus planipennis*) by the transport of firewood between campgrounds in Ohio and Indiana.

One idea for further research, initially proposed by Johnson et al. (2001) that remains largely ignored, is the potential loss in individual fitness as NIS are transported from source to destination (see also Wonham et al., 2005*b*). Johnson et al. (2001) showed adult zebra mussels attached to macrophytes, and that are exposed to the air as boats are transported from lake to lake, experience decreased survival. It is interesting that most of the produced gravity models are for the transport of zebra mussels (Schneider et al., 1998; Bossenbroek et al., 2001; Leung et al., 2004, 2006; Bossenbroek et al., 2007), but do not account for decreased survival as a function of distance traveled. This was identified as a source of gravity model uncertainty in Chapter 4. One explanation for ignoring the transport–fitness–loss aspect of the arrival process has developed because it is now generally held that juvenile, free swimming zebra mussels transported in live wells are causing invasions and not adults (Bossenbroek et al., 2001). Similarly, gravity models for spiny water flea do not appear to necessitate accounting for loss of fitness during transport because invasions likely occur because of environmentally resilient egg masses (Wonham et al., 2005*a*; Muirhead, 2007). However, this is not true for all aquatic NIS, and accounting for the loss of fitness due to transport mechanisms is a potentially fruitful direction of study.

6.2.4 Allee effects

In Chapter 5, I developed a population model for a semelparous, sexual species that used discrete probability distributions and first principles of survival, demography, fertilization probability, and fecundity. In contrast to previous probabilistic approaches (see, for example, Dennis, 1989), the Allee effect is not explicitly modeled, but instead emerges in the expected net per capita geometric growth rate. As a stochastic process, persistence can be evaluated in discrete time. Future research should develop the iteroparous life history because more NIS have this reproductive strategy, as exemplified by the IUCN's 100 worst invasive species list (http://www.issg.org/). The semelparous hierarchical probability model (Chapter 5) could be modified to account for males and females that survive to the next reproductive bout, resulting in a stage-structure model (Caswell, 2001).

Inducing a strong Allee effect (negative per capita growth rate with decreasing population size (see, for example, Allee, 1938; Dennis, 1989; Kot et al., 1996) or decreasing propagule pressure (Leung et al., 2004)) is often invoked as a management strategy for preventing invasions (see for example, Taylor and Hastings, 2005; Drury et al., In Press). However decreasing propagule pressure below an Allee threshold only results in certain extinction with deterministic models (see, for example, Lewis and Kareiva, 1993; Kot et al., 1996). In contrast, stochastic models with an Allee effect have populations persisting with some probability even when propagule pressure is below the Allee threshold (Chapter 5; Dennis, 2002; Drake and Lodge, 2006). Population dynamics, such as the Allee effect and demographic stochasticity, leading to population extinction or persistence is germane to defining establishment. However, of greater concern is that Allee effects are not misrepresented as offering absolute protection against biological invasions.

6.3 Invasion risk

6.3.1 Risk, relative risk, and ranks

In Chapter 4, I discussed the difference between the ecological and economic perspectives of risk, and throughout this dissertation I have clearly delineated the differences between risk as a probability of an undesirable event occurring and relative risk as the ratio of probabilities (see, for example, Knight, 1921; Suter, 1993). But, in the ecology literature generally, and the invasion biology literature specifically, there exists a large amount of ambiguity regarding the meaning and estimation of risk (Nelson et al., In Press).

I believe the greatest concern is the use of ranked risk, that is the ordering of probabilities. In invasion biology this is used to identify locations more or less likely to become invaded by a species (Leung et al., 2004). Alternatively, lists of potential NIS are composed and then the propagule pressure is estimated for a specific location where species with large propagule pressure are assumed to pose a greater invasion risk (Cohen et al., 2007). The general assumption is that propagule pressure is correlated with invasion risk, but the link from a rank to a probability is not presented. In Chapter 2, I showed the assumptions needed, namely that the survival is the same between locations or species but unknown, in order to use relative risk measures rather than probabilities. Applied rankings of invasion risk based on propagule pressure alone fail to account for this assumption.

Even if species were pooled by similar survivorships and then ranked, it is unclear if ranks provide sufficient information to inform NIS management. First, bioeconomic applications that assess the optimal strategy of either preventing, controlling, or doing nothing, start with a probability of establishment (Leung et al., 2004; Finnoff et al., 2007). How ranks fit into bioeconomic considerations is unknown. Some would argue that providing ranks identify the species that should be targeted for risk assessment (Cohen et al., 2007). My second concern would refute that ranks are useful for even this task. Consider three scenarios (S_1, S_2, and S_3) with four species ranked by their probability of establishment, $S_1 = \{0.99, 0.98, 0.97, 0.96\}$, $S_2 = \{0.99, 0.98, 0.002, 0.001\}$, and $S_3 = \{0.004, 0.003, 0.002, 0.001\}$. Each of these scenarios has the same rank, but very different consequence. Clearly, ranks do not tell us how risky the groups of species are to invade, where in S_1 all species are likely to invade and in S_3 all species are likely not to invade. Additionally, ranks do not indicate breaks between likely and unlikely to invade species, (S_2).

Relative risk, as developed for waiting times in Chapter 2 and applied in Chapter 4, is not immune to similar criticisms. Recalculating the scenarios so the probabilities are relative to the smallest probability would yield $S_1 = \{\approx 1, \approx 1, \approx 1, 1\}$, $S_2 = \{99, 98, 2, 1\}$, and $S_3 = \{4, 3, 2, 1\}$. The magnitude of the relative risk does not map to a probability, and therefore relative risk does not indicate how risky the groups of species are to invade. However, relative risk is able to delineate breaks – an advantage over the ranked risk. Invasion biology needs to become more rigorous with regards to risk assessment (Bossenbroek et al., 2005), and it is unclear if the modeling approaches, such as gravity models, are providing the necessary information needed to perform bioeconomic assessments and make management decisions. A critical next step would be to reduce the linguistic uncertainty (Regan et al., 2002) by providing a review of risk terminology and point out inconsistencies in the literature.

6.3.2 Biotic and abiotic influences on invasion waiting time

Implicit in the models of invasion risk proposed in Chapters 2 and 5 are the influence of biotic and abiotic factors at the location the NIS are being introduced to. The invasion waiting time distributional shape (Chapters 2 and 3) is determined by the product of the propagule pressure (ϕ) and survival (p_s), where the arrival process dictates the propagule pressure but survival is a reflection of environmental conditions (see, for example, Moyle and Light, 1996). For example, Chinese mitten crabs have different survivorships for variable estuary salinity and water temperature (Anger, 1991, 2006). Herborg et al. (2007) showed that some United States shipping ports were at a lower or greater risk of invasion due to abiotic factors. My dissertation has primarily emphasized the model formulation for invasion risk into one location or assumed the survival is the same between two location when making relative risk formulations. However, including abiotic and biotic covariates into my proposed invasion risk models would be useful for tactical approaches.

With advancements in geographical information systems (GIS) and the availability of high–spatial–resolution, abiotic data, such as precipitation and temperature, researchers have made susceptibility maps that identify locations where NIS are more or less likely to establish based on an assumed positive correlation between the covariates and survival (Lodge et al., 2006). For example, Drake and Bossenbroek (2004) used GIS to identify lakes and watersheds in the western United States that are susceptible to zebra mussel invasion. The approach is called habitat matching, presence only modeling, or environmental niche modeling, and the general approach relies on using models parameterized with habitat covariates where the species is currently established (occupancy only) and then applying the model to uninvaded locations to identify susceptible locations (Lodge et al., 2006). Two common approaches are discriminant analysis (see, for example, Kolar and Lodge, 2002) and genetic algorithm rule set prediction (GARP), (see, for example, Drake and Bossenbroek, 2004). Many other approaches are available, but in general there are limitations in that the realized niche and not the fundamental niche is being modeled (Hutchinson, 1957; Drake and Bossenbroek, 2004). As a consequence, the models perform reasonably well for well defined, narrow niches but poorly for wide niches (Tsoar et al., 2007). The problem is further compounded because susceptibility is not the same as risk (Nelson et al., In Press). In order for risk to be estimated, propagule pressure must included (Chapter 2; Leung and Mandrak, In Press). As a result, the models in Chapter 2 and Chapter 5 provide a bridge from susceptibility to invasion risk if the covariates are included in the model parameters and propagule pressure is known for a particular location. When the relationship between the covariates and the survival is unknown, then relative measure may be used to compare locations with similar susceptibility (Chapter 2; Herborg et al., 2007).

In contrast to abiotic factors, the estimation of invasion risk with biotic factors at the landscape level (but see, Tsoar et al., 2007) has been virtually ignored (Lodge et al., 2006), yet many of the invasion biology hypotheses and theories have been gleaned from community ecology (Shea and Chesson, 2002). For example, (Simberloff and Von Holle, 1999) reinvented part of the facilitative model of Connell and Slayter (1977) to suggest the number of NIS present in a landscape will increase the survival of new invaders (dubbed invasional meltdown). At best, the role of species richness, be they NIS or native species, in increasing or decreasing the probability of invasion remains elusive (Shea and Chesson, 2002; Meiners et al., 2004). One of the core invasion biology hypothesis is biotic resistance (Elton, 1958), but even the large number of experiments conducted to test this hypothesis show little or no effect in preventing individual NIS from establishing (Levine et al., 2004).

Tilman (2004) proposed the idea of stochastic niche theory (SNT) to explain successful invasions by merging neutral theory (Hubbell, 2001), resource competition, and fluctuating resources (Davis et al., 2000). Tilman's SNT is that NIS establish only if a propagule can 'stochastically survive' in a new community on the limited resources available to the propagule. Species diversity itself does not exclude NIS, rather species diversity is responsible for sequestering resources such that there is a limited amount of resources available to the NIS - reducing survival. Hence diversity indirectly excludes new invaders. However, fluctuations in resources that either directly free up resources, or indirectly reduces species diversity, will facilitate invasions. The foundations of my formulation of invasion waiting times are in parallel with Tilman's formulation of the SNT. What remains unclear is how biotic interactions, particularly if they are limited to local, 'neighborhood' level interactions, can be modeled at a landscape level for use in assessing invasion risk for application in management.

6.4 Risk management and conservation biology

Invaders can reduce native species abundance and significantly impact ecosystem functioning (Parker et al., 1999; Chapin et al., 2000). In addition to the impact on the environment, substantial economic losses are accrued to the public and industry (Pimentel et al., 2000, 2005). In the few cases where the financial impact of invasive species have been assessed, it is generally acknowledged that substantially more resources should be allocated to preventing, eradicating, and slowing invasions (Leung et al., 2002; Simberloff, 2003). The process of estimating the probability of invasion, valuating the costs of an invasion and management actions, measuring the reduction in the probability of invasion due to management intervention, and deciding what action, if any, to take, is invasive species risk management (Stohlgren and Schnase, 2006).

6.4.1 Predicting invasions

In 2001, the Society for Conservation Biology (SCB) proposed a research agenda for the next decade, and one of the emerging themes was the need "to perform research to enable us to better predict which exotic species are likely to become invasive and what their effects will be" (Soule and Orians, 2001). In a following chapter on NIS, D'Antonio et al. (2001) advocated identifying critical pathways of invasion, development of quantitative techniques to evaluate invasion risk, applying risk assessment, and implementing risk management. All of these agenda items were denoted as critical developments that must be researched and applied in the next decade to stem the impacts of invasive species on biodiversity and habitat functioning.

Invasive species risk assessment forms the basis for predicting invasions (Kolar and Lodge, 2001, 2002). Predicting invasions has generally proceeded independently on three fronts, risk due to fluctuations in propagule pressure (Lockwood et al., 2005), risk inherent to the characteristics (life history) of the invader (Kolar and Lodge, 2001; Keller et al., 2007), and the invasion susceptibility of the environment due to biotic and abiotic factors (Drake and Bossenbroek, 2004).

However, the overall risk, that is the probability a population will establish in a new location, is a function of all three components - with my dissertation, I have advocated this approach.

In the context of invasive species management, I have taken a narrow research agenda relative to that advocated by members of the SCB. However, proceeding with the evaluation of management strategies and optimizing management decisions (Leung et al., 2002) are contingent on reliable risk assessments (D'Antonio et al., 2001). However, as a result of my thesis work, I believe a refinement of the SCB agenda should include two critical lines of questioning for predicting NIS: First, as I have shown, in the absence of data to estimate parameters, mathematical and biological assumptions can lead to relative risk assessments. The paucity of NIS data will continue to plague invasion biology as new potential invaders are identified for risk assessment. Consequently, is it beneficial to conduct relative risk assessments, and can current risk management strategies be adapted to include relative risk formulations? This includes model validation techniques proposed in Chapter 4. Second, how much uncertainty in the estimated probability of establishment is tolerable for accurate predictions and invasive species management? I briefly discussed this point at the end of Chapter 4, but its importance is worth iterating. As I have formulated it, the invasion process is inherently stochastic with some probability of a location becoming invaded based on parameters, such as propagule pressure and survival. But those parameters also have variability that will result in uncertainty in the probability of establishment. Poor estimates will lead to poor predictions (Breiman, 2001) and ineffective invasive species management (Stohlgren and Schnase, 2006).

6.4.2 Risk management and communicating risk

Recently, Lodge et al. (2006) provided an updated research agenda for invasive species emphasizing the need for better communication and rapid response between scientists and managers in order to slow or stop the rate of NIS introduction and spread. Many of the tools needed for invasive species risk assessment are in place (Drake and Lodge, 2006; Jerde and Lewis, 2007), but putting conservation science into practice will take educating managers about risk assessment, and educating scientists about risk management (Bossenbroek et al., 2005).

Leung et al. (2002) provide a general risk management framework that links risk assessment to an economic model, which evaluates the costs and benefits of management actions through stochastic dynamic programming. This framework requires estimating the probability of invasion, valuating the costs of an invasion and management actions, and measuring the reduction in the probability of invasion from the different management actions (Stohlgren and Schnase, 2006).

The models in Chapters 2 and 5 are risk assessments and do not answer the question, 'given this probability of establishment, what do I do?' Although my dissertation is not aimed at performing risk management analysis, the results, such as invasion waiting times, are useful for optimizing management actions (*personal communication*, Alexei Potapov). Returning to my concern about uncertainty in risk assessment, future risk management may take an 'active adaptive management' approach to biological invasions (McCarthy and Possingham, 2007).

Because invasion biology is a crisis discipline (Lodge, 1993; D'Antonio et al., 2001), decisions are made with uncertainty in the underlying model and parameters (Regan et al., 2002). Invasive species management seeks to reduce the probability of invasion and reduce the accrued costs by successful invasions for specified levels of risk. The active adaptive management strategy incorporates learning about the invasions process for which managers are responsible so that future management is improved, while balancing the costs for monitoring, intervention, and eradication (Simberloff, 2003). I believe this will form the basis for future NIS risk management based on risk assessment models.

Bibliography

Allee, W. 1938. The social life of animals. W.W. Norton and Co., New York, NY, USA.

Anger, K. 1991. Effects of temperature and salinity on the larval development on the Chinese mitten crab *Eriocheir sinensis* (Decapoda: Grapsoidea). Marine Ecology Progress Series **72**:103–110.

Anger, K. 2006. Contributions of larval biology to crustacean research: a review. Invertebrate Reproduction and Development **49**:175–205.

Bossenbroek, J., C. Kraft, and J. Nekola. 2001. Prediction of long-distance dispersal using gravity models: Zebra Mussel invasion of inland lakes. Ecological Applications **11**:1778–1788.

Bossenbroek, J., J. Ladd, B. Peter, and D. Lodge. 2007. Forecasting the expansion of Zebra Mussels in the United States. Conservation Biology **21**:800–810.

Bossenbroek, J., J. McNulty, and R. Keller. 2005. Can ecologists heat up the discussion on invasive species risk? Risk Analysis **25**:1595–1597.

Breiman, L. 2001. Statistical modeling: the two cultures. Statistical Sciences **16**:199–231.

Carlton, J. 1996. Pattern, process, and prediction in marine invasion ecology. Biological Conservation **78**:97–106.

Caswell, H. 2001. Matrix Population Models. 2nd edition. Sinauer.

Chapin, F., E. Zavaleta, V. Eviner, N. Rosamond, P. Vitousek, H. Reynolds, D. Hooper, S. Lavorel, O. Sala, S. Hobbie, M. Mack, and S. Diaz. 2000. Consequences of changing biodiversity. Nature **405**:234–242.

Cohen, J., N. Mirotchnick, and B. Leung. 2007. Thousands introduced annually: the aquarium pathway for non-indigenous plants to the St Lawrence Seaway. Frontiers in Ecology and the Environment **5**:doi:10.1890/060137.

Connell, J., and R. Slayter. 1977. Mechanisms of succession in natural communities and their role in community stability and organization. American Naturalist **111**:1119–1144.

D'Antonio, C., L. Meyerson, and J. Denslow, 2001. Conservation Biology: research priorities for the next decade, Chapter exotic species and conservation: Research needs, pages 59–80 . Island Press, Washington, D.C., USA.

Davis, M., J. Grime, and K. Thompson. 2000. Fluctuating resources in plant communities: a general theory of invasibility. Journal of Ecology **88**:528–534.

Dennis, B. 1989. Allee effects: population growth, critical density, and the chance of extinction. Natural Resource Modeling **3**:481–538.

Dennis, B. 2002. Allee effects in stochastic populations. Oikos **96**:389–401.

Drake, J., and J. Bossenbroek. 2004. The potential distribution of Zebra Mussels in the US. BioScience **54**:931–941.

Drake, J., and D. Lodge. 2004. Global hot spots of biological invasions: evaluating options for ballast-water management. Proceedings of the Royal Society of London, Series B **271**:575–580.

Drake, J., and D. Lodge. 2006. Allee effects, propagule pressure and the probability of establishment: risk analysis for biological invastions. Biological Invasions **8**:365–375.

Drury, K., J. Drake, D. Lodge, and G. Dwyer. In Press. Immigration events dispersed in space and time: factors affecting invasion success. Ecological Modelling .

Elton, C. 1958. The ecology of invasion by animals and plants. Methuen, London, UK.

Finnoff, D., J. Shogren, B. Leung, and D. Lodge. 2007. Take a risk: Preferring prevention over control of biological invaders. Ecological Economics **62**:216–222.

Floerl, O., and G. Inglis. 2005. Starting the invasion pathway: the interaction between source populations and human transport vectors. Biological Invasions **7**:589–606.

Herborg, L., C. Jerde, D. Lodge, G. Ruiz, and H. MacIsaac. 2007. Predicting the North American distribution of Chinese mitten crabs (*Eriocheir sinensis*) using measures of propagule pressure and environmental niche models. Ecological Applications **17**:663–674.

Hubbell, S. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, New Jersey.

Hutchinson, G. 1957. Concluding remarks. Cold Spring Harbor Symposia on Quantitative Biology **22**:415–427.

Jerde, C., and M. Lewis. 2007. Waiting for invasions: a framework for the arrival of non–indigenous species. American Naturalist **170**:1–9.

Johnson, L., A. Ricciardi, and J. Carlton. 2001. Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. Ecological Applications **11**:1789–1799.

Keller, R., J. Drake, and D. Lodge. 2007. Fecundity as a basis for risk assessment of nonindigenous freshwater molluscs. Conservation Biology **21**:191–200.

Knight, F. 1921. Risk, Uncertainty, and Profit. Houghton Mifflin, Boston, Mass.

Kolar, C., and D. Lodge. 2001. Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution **16**:199–204.

Kolar, C., and D. Lodge. 2002. Ecological predictions and risk assessment for alien fishes in North America. Science **298**:1233–1236.

Kot, M., M. Lewis, and P. van den Driessche. 1996. Dispersal data and the spread of invading organisms. Ecology **77**:2027–2042.

Leung, B., J. Bossenbroek, and D. Lodge. 2006. Boats, pathways, and aquatic biological invasions: estimating dispersal potential with gravity models. Biological Invasions **8**:241–254.

Leung, B., J. Drake, and D. Lodge. 2004. Predicting invasions: propagule pressure and the gravity of Allee effects. Ecology **85**:1651–1660.

Leung, B., D. Lodge, D. Finnoff, J. Shogren, M. Lewis, and G. Lamberti. 2002. An ounce of prevention or a pound of cure: bioeconomic risk analysis of invasive species. Proceedings of the Royal Society of London, B **269**:2407–2413.

Leung, B., and N. Mandrak. In Press. The risk of establishment of aquatic invasive species: joining invasibility and propagule pressure. Proceedings of the Royal Society, B .

Levine, J., P. Adler, and S. Yelenik. 2004. A meta–analysis of biotic resistance to exotic plant invasions. Ecology Letters **7**:975–989.

Lewis, M., and P. Kareiva. 1993. Allee dynamics and the spread of invading organisms. Theoretical Population Biology **43**:141–158.

Lockwood, J., P. Cassey, and T. Blackburn. 2005. The role of propagule pressure in explaining species invasions. Trends in Ecology and Evolution **20**:223–228.

Lodge, D. 1993. Biological Invasions: lessons for ecology. Trends in Ecology and Evolution **8**:133–137.

Lodge, D., S. Williams, H. MacIsaac, K. Hayes, B. Leung, S. Reichard, R. Mack, P. Moyle, M. Smith, D. Andow, J. Carlton, and M. McMichael. 2006. Biological invasions: recommendations for U.S. policy and management. Ecological Applications **16**:20352054.

MacIsaac, H., J. Borbely, J. Muirhead, and P. Graniero. 2004. Backcasting and forecasting biological invasions of inland lakes. Ecological Applications **14**:773–783.

McCarthy, M., and H. Possingham. 2007. Active adaptive management for conservation. Conservation Biology **21**:956–963.

Meiners, S., M. Cadenasso, and S. Pickett. 2004. Beyond biodiversity: individualistic controls of invasion in a self-assembled community. Ecology Letters **7**:121–126.

Moyle, P., and T. Light. 1996. Fish invasions in California: Do abiotic factors determine success? Ecology **77**:1666–1670.

Muirhead, J., 2007. Forecasting and validating the establishment of aquatic nonindigenous species. Ph.D. thesis, The University of Windsor, Windsor, Ontario, Canada.

Nelson, W., A. Potapov, M. Lewis, A. Hundsdorfer, and F. He. In Press. Balancing ecological complexity in predictive models: a reassessment of risk models in the mountain pine beetle system. Journal of Applied Ecology .

Parker, I., D. Simberloff, W. Lonsdale, K. Goodell, M. Wonham, P. Kareiva, M. Williamson, B. Von Holle, P. Moyle, J. Byers, and L. Goldwasser. 1999. Impact: Toward a framework for understanding the ecological effects of invaders. Biological Invasions **1**:3–19.

Pimentel, D., L. Lach, R. Zunign, and D. Morrison. 2000. Environmental and economic costs of non-indiginous species in the United States. BioScience **50**:53–67.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien–invasive species in the United States. Ecological Economics **52**:273–288.

Puth, L., and D. Post. 2005. Studying invasion: have we missed the boat? Ecology Letters **8**:715–721.

Regan, H., M. Colyvan, and M. Burgman. 2002. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications **12**:618–628.

Sakai, A., F. Allendorf, J. Holt, D. Lodge, J. Molofsky, K. With, S. Baughman, R. Cabin, J. Cohen, N. Ellstrand, D. McCauley, P. O'Neil, I. Parker, J. Thompson, and S. Weller. 2001. The population biology of invasive species. Annual Reviews of Ecology and Systematics **32**:305–32.

Schneider, D., C. Ellis, and K. Cummings. 1998. A transportation model assessment of the risk to native mussel communities from Zebra Mussel spread. Conservation Biology **12**:788–800.

Shea, K., and P. Chesson. 2002. Community ecology theory as a framework for biological invasions. Trends in Ecology and Evolution **17**:170–176.

Simberloff, D. 2003. Eradication – preventing invasions at the outset. Weed Science **51**:247–253.

Simberloff, D., and B. Von Holle. 1999. Positive interactions of nonindigenous species: Invasional meltdown? Biological Invasions **1**:21–32.

Soule, M., and G. Orians. 2001. Conservation Biology: research priorities for the next decade. Island Press, Washington, D.C., USA.

Stohlgren, T., and J. Schnase. 2006. Risk Analysis for biological hazards: what we need to know about invasive species. Risk Analysis **26**:163–173.

Suter, G. 1993. Ecological Risk Assessment. Lewis Publishers, Chelsea, Michigan.

Taylor, C., and A. Hastings. 2005. Allee effects in biological invasions. Ecology Letters **8**:895–908.

Tilman, D. 2004. Niche tradeoffs, neutrality, and community structure: A stochastic theory of resource competition, invasion, and community assesmbly. Proceedings of the National Academy of Science, U.S.A. **101**:10854–10861.

Tsoar, A., O. Allouche, D. Steinitz, D. Rotem, and R. Kadmon. 2007. A comparative evaluation of presence only methods for modelling species distribution. Diversity and Distributions **13**:397–405.

Vermeij, G. 1996. An agenda for invasion biology. Biological Conservation **78**:3–9.

Von Holle, B., and D. Simberloff. 2005. Ecological resistance to biological invasion overwhelmed by propagule pressure. Ecology **86**:3212–3218.

Wonham, M., S. Bailey, H. MacIsaac, and M. Lewis. 2005*a*. Modelling the invasion risk of diapausing organisms transported in ballast sediments. Canadian Journal of Fisheries and Aquatic Science **63**:2386–2398.

Wonham, M., J. Carlton, G. Ruiz, and L. Smith. 2000. Fish and ships: relating dispersal frequency to success in biological invasions. Marine Biology **136**:1111–1121.

Wonham, M., M. Lewis, and H. MacIsaac. 2005*b*. Minimizing invasion risk by reducing propagule pressure: a model for ballast water exchange. Frontiers in Ecology and the Environment **3**:473–478.

Wonham, M., M. Lewis, J. Renclawowicz, and P. van den Driessche. 2006. Transmission assumptions generate conflicting predictions in hostvector disease models: a case study in West Nile virus. Ecology Letters **9**:706–725.

Appendix A

Simplifications of the hierarchical model

The source (s) population is modeled as a Poisson random variable, N_s, with parameter λ,

$$\begin{aligned} N_s &\sim \text{Poisson}(\lambda), \\ \text{Prob}(N_s = n_s) &= \frac{e^{-\lambda}\lambda^{n_s}}{n_s!}. \end{aligned}$$

The dispersal pool (dp) is modeled as binomial random variable, N_{dp}, with parameters p_t the individual probability of being transported and N_s a realization from the source population. Thus the dispersal pool distribution is a conditional distribution and forms a hierarchical probability model,

$$\begin{aligned} N_{dp} \mid N_s &\sim \text{Binomial}(N_s, p_t), \\ \text{Prob}(N_{dp} = n_{dp} \mid N_s = n_s) &= \binom{n_s}{n_{dp}} p_t^{n_{dp}} (1 - p_t)^{n_s - n_{dp}} \end{aligned}$$

For purposes of inference we are interested in the marginal distribution of N_{dp}. Using conditional probability we can find this distribution,

$$\begin{aligned}
\text{Prob}(N_{dp} = n_{dp}) &= \sum_{n_s=0}^{\infty} \text{Prob}(N_{dp} = n_{dp} \ \mid \ N_s = n_s)\ \text{Prob}(N_s = n_s) \\
&= \sum_{n_s=0}^{\infty} \binom{n_s}{n_{dp}} p_t^{n_{dp}} (1 - p_t)^{n_s - n_{dp}}\ \frac{e^{-\lambda}\lambda^{n_s}}{n_s!} \\
&= \frac{(\lambda p_t)^{n_{dp}}\ e^{-\lambda}}{n_{dp}!} \ \sum_{n_s=n_{dp}}^{\infty} \frac{((1 - p_t)\lambda)^{n_s - n_{dp}}}{(n_s - n_{dp})!} \\
&= \frac{(\lambda p_t)^{n_{dp}}\ e^{-\lambda}}{n_{dp}!}\ e^{(1-p_t)\lambda} \\
&= \frac{(\lambda p_t)^{n_{dp}}\ e^{-(\lambda p_t)}}{n_{dp}!} \\
&\sim \text{Poisson}(\lambda p_t)
\end{aligned}$$

This same simplification can be used to find the marginal distributions of N_d and N_E.

Appendix B

Order of invasion derived distribution

The waiting time for location A to become invaded is geometrically distributed,

$$\text{Prob}(T_A = t_A) = \gamma_A(1-\gamma_A)^{t_A-1}.$$

Similarly, the waiting time for location B to become invaded is geometrically distributed,

$$\text{Prob}(T_B = t_B) = \gamma_B(1-\gamma_B)^{t_B-1}.$$

We are interested in the transformation $T_D = T_A - T_B$. This results in the statement $T_B = T_A - T_D$. We can find the distribution of T_D by using the joint distribution method and deriving the marginal distribution of T_D.

The joint distribution, assuming independence, of T_A and T_B, is,

$$f_{T_A,T_B}(t_A,t_B) = f_{T_A}(t_A)f_{T_B}(t_B) = \gamma_A(1-\gamma_A)^{t_A-1}\gamma_B(1-\gamma_B)^{t_B-1}.$$

Inserting the transformation $T_A - T_D$ for T_B yields,

$$f_{T_A,T_B}(t_A,t_B) = f_{T_A,T_A-T_D}(t_A,t_A-t_D) = \gamma_A(1-\gamma_A)^{t_A-1}\gamma_B(1-\gamma_B)^{t_A-t_D-1}.$$

Summing over all values of t_A leads to the marginal of T_D.

$$\begin{aligned} f_{T_D}(t_D) &= \sum_{t_A=t_D+1}^{\infty} f_{T_A,T_A-T_D}(t_A, t_A-t_D) \\ &= \sum_{t_A=t_D+1}^{\infty} \gamma_A(1-\gamma_A)^{t_A-1}\gamma_B(1-\gamma_B)^{t_A-t_D-1} \\ &= \frac{\gamma_A\gamma_B(1-\gamma_B)^{t_D}((-1+\gamma_A)(-1+\gamma_B))^{t_D}}{\gamma_A+\gamma_B-\gamma_A\gamma_B} \end{aligned}$$

The probability of location B being invaded before location A is the sum for all positive, real numbers $t_D = 1, 2, ...\infty$ of the transformation, T_D.

$$\text{Prob}(B \text{ before A}) = \sum_{\mathrm{t_D=1}}^{\infty} \frac{\mathrm{\gamma_A\gamma_B(1-\gamma_B)^{t_D}((-1+\gamma_A)(-1+\gamma_B))^{t_D}}}{\mathrm{\gamma_A+\gamma_B-\gamma_A\gamma_B}} = \frac{\mathrm{\gamma_B-\gamma_A\gamma_B}}{\mathrm{\gamma_A+\gamma_B-\gamma_A\gamma_B}}$$

The probability of location A and location B being invaded at the same time step is the sum of all $t_A = t_B = 1, 2, ...\infty$ of the joint probability,

$$\text{Prob(A and B)} = \sum_{\mathrm{t_A=t_B=1}}^{\infty} \mathrm{\gamma_A(1-\gamma_A)^{t_A-1}\gamma_B(1-\gamma_B)^{t_B-1}} = \frac{\mathrm{\gamma_A\gamma_B}}{\mathrm{\gamma_A+\gamma_B-\gamma_A\gamma_B}}.$$

The probability of location A being invaded before location B is $1 -$ Prob(B before A) – Prob(A and B),

$$1 - \frac{\gamma_B-\gamma_A\gamma_B}{\gamma_A+\gamma_B-\gamma_A\gamma_B} - \frac{\gamma_A\gamma_B}{\gamma_A+\gamma_B-\gamma_A\gamma_B} = \frac{\gamma_A-\gamma_A\gamma_B}{\gamma_A+\gamma_B-\gamma_A\gamma_B}.$$

These order of invasion probabilities are useful when the absolute probabilities of establishment (γ_A and γ_B) are known. However, as mentioned, γ_A and γ_B are likely difficult to estimate. Alternatively, with the assumption of small γ_A and γ_B, the order of invasion probabilities can be estimated using the relative risk, $R = \frac{\gamma_A}{\gamma_B}$ and eliminating higher order terms (h.o.t.). The ordered invasion equations can be reformulated as,

$$\text{Prob(B before A)} = \frac{\gamma_B-\gamma_A\gamma_B}{\gamma_A+\gamma_B-\gamma_A\gamma_B} = \frac{\gamma_B}{\gamma_A+\gamma_B} + \text{h.o.t.} \approx \frac{1}{1+\mathrm{R}},$$

$$\text{Prob(A and B)} = \frac{\gamma_A\gamma_B}{\gamma_A+\gamma_B-\gamma_A\gamma_B} = 0 + \text{h.o.t.} \approx \ 0,$$

$$\text{Prob(A before B)} = \frac{\gamma_A - \gamma_A\gamma_B}{\gamma_A + \gamma_B - \gamma_A\gamma_B} = \frac{\gamma_A}{\gamma_A + \gamma_B} + \text{h.o.t.} \approx \frac{\text{R}}{1 + \text{R}}.$$

If γ_A or γ_B are large, the ordered probabilities of invasion estimated from R will be poor. This can be observed in the scentless chamomile example.

Appendix C

Formulation of the waiting time prediction

The question posed is, “What is the probability of observing $r^\star$ lakes of a group of N references lakes invaded by time t?”

For a single lake the waiting time will follow the probability mass function (PMF) of the geometric distribution,

$$\underbrace{\text{Pr}(T = t)}_{\substack{\text{Probability of}\\ \text{being invaded}\\ \text{at time t}}} = \underbrace{p_u}_{\substack{\text{Probability of}\\ \text{being invaded}}} \underbrace{(1 - p_u)^{t-1}}_{\substack{\text{Probability of}\\ \text{being uninvaded}\\ \text{until time t}}}, \tag{C.1}$$

where p_u is the probability a lake becomes invaded and $t = 1, 2, 3 \ldots$. The probability of a lake being invaded by time is the Cumulative Mass Function (CMF) of the geometric distribution,

$$\text{Pr}(T \leq t) = \underbrace{1 - (1 - p_u)^t}_{\substack{\text{Probability of}\\ \text{being invaded}\\ \text{by time t}}}. \tag{C.2}$$

To consider multiple lakes in the reference group, the probability of being invaded by time t needs to be considered for N lakes. If we assume the probability of becoming invaded is the same for all lakes, then we can formulate the probability of observing $r^\star$ lakes invaded by time t from a group of N reference lakes as a binomially distributed random variable,

$$\underbrace{\Pr(R = r^\star)}_{\substack{\text{Probability of}\\ \text{r}^\star\text{ lakes}\\ \text{invaded}}} = \underbrace{\binom{N}{r^\star}}_{\substack{\text{No. of}\\ \text{ways}\\ \text{r}^\star\text{ arises}}} \underbrace{(1-(1-p_u)^t)^{r^\star}}_{\substack{\text{Probability of}\\ \text{r}^\star\text{ invaded}\\ \text{lakes}}} \underbrace{((1-p_u)^t)^{N-r^\star}}_{\substack{\text{Probability of}\\ \text{N}-\text{r}^\star\text{ uninvaded}\\ \text{lakes}}} \tag{C.3}$$

When $N = 1$ this formulation reduces to the CMF of the geometric distribution.

A second question of biological interest is, "what is the probability that r lakes become invaded from a reference group of N lakes at time t while another lake (i) remains uninvaded?" The $\star$ notation is dropped to denote the difference in the random variables.

The probability a lake i becoming invaded is p_i. Then the probability the lake remains uninvaded is the complement $1 - p_i$. Assuming the invasion of this lake is i.i.d., the probability of remaining uninvaded at time t is $(1 - p_i)^t$. Again assuming the invasion of this lake and the lakes of the reference group are independent of each other, then the probability of observing r lakes of N reference lakes becoming invaded by time t while the target lake remains uninvaded is,

$$\underbrace{\Pr(R = r)}_{\substack{\text{Probability of}\\ \text{r lakes}\\ \text{invaded}}} = \underbrace{\binom{N}{r}}_{\substack{\text{No. of}\\ \text{ways}\\ \text{r arises}}} \underbrace{(1-(1-p_u)^t)^{r}}_{\substack{\text{Probability of}\\ \text{r invaded}\\ \text{lakes}}} \underbrace{((1-p_u)^t)^{N-r}}_{\substack{\text{Probability of}\\ \text{N}-\text{r uninvaded}\\ \text{lakes}}} \underbrace{(1-p_i)^t}_{\substack{\text{Probability of}\\ \text{H.S.Truman}\\ \text{uninvaded}}}. \tag{C.4}$$

If the probabilities of any of the reference lakes or the target lake were known, then this second formulation is unnecessary. Instead, the frequency with which similar lakes, such as the first formulation could be used to assess the waiting time directly. However, from gravity scores the actual probability of invasion is unknown and only relative waiting times are available. This is demonstrated in the main body of the text.

Appendix D

Formulation of the hierarchical probability model of establishment

D.1 Model formulation

We take the approach of modeling individual processes, such as survival, fertilization, and fecundity, using discrete probability mass functions, and then generalize the emerging population properties with statistics such as the expected geometric growth rate and reproductive value. Throughout we assume individuals act independently and are identically distributed (i.i.d.). For consistency, italicized, capital letters are random variables (Table 5.2), italicized, lower-case letters are the possible values of random variables, and italicized, Greek letters, along with the letters p and n_0, are parameters (Table 5.1).

D.1.1 Arrival and initial survival

Some fixed and known number of individuals, n_0, are introduced into a population. Of this population, each individual has some probability p of surviving to reproductive maturity. The number of individuals of n_0 that survive can modeled as a binomial random variable S, where

$$\Pr(S = s) = \binom{n_0}{s} p^s (1-p)^{n_0 - s}. \tag{D.1}$$

If the population does not experience mortality between arrival and reproductive maturity, then the initial survival step may be skipped and the random variable S would then be replaced with the parameter n_0. Additionally, if p is small and n_0

is large, then S can be modeled as a Poisson distributed random variable following the Poisson approximation of the binomial distribution (Casella and Berger, 2002). Hereafter we use the Poisson approximation such that,

$$\Pr(S = s) = \frac{e^{-n_0 p}(n_0 p)^s}{s!}. \tag{D.2}$$

D.1.2 Demographics

A sexual population of S individuals is the sum of males (M) and females (F) where the probability of being female is the parameter, ψ. Because the population consists of only reproductively mature males and females, the sex ratio (e.g. males to females = 4:1) can be directly interpreted as the probability of being female (e.g. ψ=0.2). This leads to a random variable of the number of females that is conditional on the number of sexually mature individuals in the population,

$$\Pr(F = f \mid S = s) = \binom{s}{f}\psi^f(1-\psi)^{s-f}. \tag{D.3}$$

The distribution of the number of females (F) is,

$$\begin{aligned}\Pr(F = f) &= \sum_S \binom{s}{f}\psi^f(1-\psi)^{s-f}\frac{e^{-n_0 p}(n_0 p)^s}{s!} &\text{(D.4)}\\ &= \frac{e^{-n_0 p\psi}(n_0 p\psi)^f}{f!}. &\text{(D.5)}\end{aligned}$$

Similarly, the marginal distribution of the males (M) is,

$$\begin{aligned}\Pr(M = m) &= \sum_S \binom{s}{m}(1-\psi)^m(\psi)^{s-m}\frac{e^{-n_0 p}(n_0 p)^s}{s!} &\text{(D.6)}\\ &= \frac{e^{-n_0 p(1-\psi)}(n_0 p(1-\psi))^m}{m!}. &\text{(D.7)}\end{aligned}$$

Alternatively, the process of separating surviving males and females for small n_0 can be formulated as a multinomial joint distribution of fatalities, D, with probability $1 - p$, surviving females, F, with probability $p\psi$, and surviving males, M, with probability, $p(1 - \psi)$.

D.1.3 Fertilization

There are F females and M males in the system. In a simple case of one female and one male, we could assume there is a probability, γ, that the male fertilizes the female. This would result in a Bernoulli trial for the random variable K,

$$\Pr(K = k) = \gamma^k (1-\gamma)^{1-k}. \tag{D.8}$$

where if k=1, then the female is fertilized, and if k=0, then the female is not fertilized.

Now if we assume that each of M males has an equal chance of fertilizing a female (i.e. $\gamma_i, \ldots, \gamma_M = \gamma$), then the distribution of the number of males that could have successfully fertilized the female, Y, is binomially distributed,

$$\Pr(Y = y \mid M = m) = \binom{m}{y} \gamma^y (1-\gamma)^{m-y}. \tag{D.9}$$

Again, using conditional probability, we are able to reduce to the marginal of Y, which is

$$\begin{aligned}
\Pr(Y = y) &= \sum_M \binom{m}{y} \gamma^y (1-\gamma)^{m-y} \frac{e^{-n_0 p(1-\psi)}(n_0 p(1-\psi)^m}{m!} && \text{(D.10)}\\
&= \frac{e^{-n_0 \gamma p(1-\psi)}(n_0 \gamma p(1-\psi))^y}{y!} && \text{(D.11)}\\
Y &\sim \text{Poisson}(n_0 \gamma p(1-\psi)). && \text{(D.12)}
\end{aligned}$$

But, we are not interested in the number of males that fertilized a single female but rather the probability at least one male fertilized a female. This leads to the probability statement

$$\Pr(y \geq 1) = 1 - \Pr(y = 0) = 1 - e^{-n_0 \gamma p(1-\psi)}. \tag{D.13}$$

There are now F females in the system. Each female has probability $1 - e^{-n_0 \gamma p(1-\psi)}$ of being fertilized by M males. We now seek the number of fertilized females. This can be treated as a the random variable Q, which is

$$\Pr(Q = q \mid F = f) = \binom{f}{q}(1 - e^{-n_0\gamma p(1-\psi)})^q(1 - (1 - e^{-n_0\gamma p(1-\psi)}))^{f-q}. \tag{D.14}$$

Again, this can be reduced to the distribution of Q,

$$\Pr(Q = q) = \frac{e^{-n_0 p\psi(1-e^{-n_0\gamma p(1-\psi)})}(n_0 p\psi(1 - e^{-n_0\gamma p(1-\psi)}))^q}{q!} \tag{D.15}$$

$$Q \sim \text{Poisson}(n_0 p\psi(1 - e^{-n_0\gamma p(1-\psi)})). \tag{D.16}$$

D.1.4 Survival, fecundity, and the next population size

Next, the fertilized females must survive to the point of giving birth. Given a probability of surviving to birthing (v), the number of fertilized females surviving to birthing (H) is a binomial survival process such that,

$$H \mid Q \sim \text{binomial}(Q, v) \tag{D.17}$$

$$\Pr(H = h \mid Q = q) = \binom{q}{h}(v)^h(1 - v)^{q-h}, \tag{D.18}$$

where the distribution of H is,

$$\Pr(H = h) = \frac{e^{-n_0 vp\psi(1-e^{-n_0\gamma p(1-\psi)})}(n_0 vp\psi(1 - e^{-n_0\gamma p(1-\psi)}))^h}{h!} \tag{D.19}$$

$$H \sim \text{Poisson}(n_0 vp\psi(1 - e^{-n_0\gamma p(1-\psi)})). \tag{D.20}$$

The surviving fertilized females, H, now produce offspring. To estimate the population size from the offspring at the next time step (N), we have H realizations of a fecundity random variable, B, from a poisson distribution. This formulation is a compound poisson process (CPP; a.k.a. random sum (Chatfield and Theobald, 1973)),

$$N \mid H, B = \sum_{i=1}^{H} B_i. \tag{D.21}$$

where,

$$B \sim \text{Poisson}(\beta) \tag{D.22}$$

$$\Pr(B = b) = \frac{e^{-\beta}\beta^{b}}{b!} \tag{D.23}$$

and β is the expected number of offspring per surviving fertilized female.

Bibliography

Casella, G., and R. Berger. 2002. Statistical Inference. 2nd edition. Duxbury, Pacific Grove, CA.

Chatfield, C., and C. Theobald. 1973. Mixtures and random sums. The Statistician **22**:281–287.

Appendix E

Recursive formulation and proof for estimating probabilities from CPP probability generating function

E.1 Formulation

The generating function of the compound Poisson process is,

$$G(t) = \mathrm{E}[t^N] = \sum_{n=0}^{\infty} \Pr(N = n)t^n = e^{(e^{(t-1)\beta}-1)(n_0 vp\psi(1-e^{-n_0\gamma p(1-\psi)}))}, \tag{E.1}$$

If we let $k = n_0 vp\psi(1 - e^{-n_0\gamma p(1-\psi)})$, then the generating function is

$$G(t) = e^{(e^{(t-1)\beta}-1)k} = e^{-k}e^{ke^{-\beta}e^{\beta t}}. \tag{E.2}$$

If we let $c = n_0 pv\psi(1 - e^{(-n_0\gamma p(1-\psi))})e^{-\beta} = ke^{-\beta}$ then,

$$G(t) = e^{-k}e^{ce^{\beta t}}, \tag{E.3}$$

with first derivative,

$$G^{(1)}(t) = e^{-k}e^{ce^{\beta t}}c\beta e^{\beta t} = c\beta e^{\beta t}G(t). \tag{E.4}$$

The probability mass of $N = n$ is recovered with,

$$\Pr(N = n) = \frac{G^{(n)}(0)}{n!}. \tag{E.5}$$

The probability that $N = 0$ is,

$$\Pr(N = 0) = \frac{G^{(0)}(0)}{0!} = G(0). \tag{E.6}$$

The derivatives of $G(0)$ in the probability mass function yield,

$$\Pr(N=1) = c\beta(\frac{\Pr(N=0)}{\beta^0 0!}), \tag{E.7}$$

$$\Pr(N=2) = \frac{c\beta^2}{2}(\frac{\Pr(N=0)}{\beta^0 1!} + \frac{\Pr(N=1)}{\beta^1 0!}), \tag{E.8}$$

$$\Pr(N=3) = \frac{c\beta^3}{3}(\frac{\Pr(N=0)}{\beta^0 2!} + \frac{\Pr(N=1)}{\beta^1 1!} + \frac{\Pr(N=2)}{\beta^2 0!}), \tag{E.9}$$

$$\Pr(N=4) = \frac{c\beta^4}{4}(\frac{\Pr(N=0)}{\beta^0 3!} + \frac{\Pr(N=1)}{\beta^1 2!} + \frac{\Pr(N=2)}{\beta^2 1!} + \frac{\Pr(N=3)}{\beta^3 0!}). \tag{E.10}$$

The formulation of the first five probability mass formulations by differentiation leads us to predict the recursive relationship,

$$\Pr(N=n) = \begin{cases} G(0) & \text{when } n=0, \\ \frac{c\,\beta^n}{n}\sum_{i=0}^{n-1}\frac{\Pr(N=i)}{(n-1-i)!\beta^i} & \text{when } n>0. \end{cases} \tag{E.11}$$

In Section E.2, we prove that the n^{th} derivative of the generating function $G(t)$ can be written in terms of the lower derivatives as follows. Suppose the following holds,

$$G^{(n)}(t) = c\beta e^{\beta t}\sum_{i=0}^{n-1}\binom{n-1}{i}G^{(i)}(t)\beta^{n-1-i}. \tag{E.12}$$

Then,

$$\begin{aligned}\Pr(N=n) = \frac{G^{(n)}(0)}{n!} &= \frac{c\beta}{n}\sum_{i=0}^{n-1}\binom{n-1}{i}G^{(i)}(t)\beta^{n-1-i} \\ &= \frac{c\beta^n}{n}\sum_{i=0}^{n-1}\frac{(n-1)!}{(n-1-i)!i!}\frac{1}{(n-1)!}\frac{G^{(i)}(t)}{\beta^i} = \frac{c\beta^n}{n}\sum_{i=0}^{n-1}\frac{\frac{G^{(i)}(0)}{i!}}{(n-1-i)!\beta^i} \\ &= \frac{c\beta^n}{n}\sum_{i=0}^{n-1}\frac{\Pr(N=i)}{(n-1-i)!\beta^i},\end{aligned} \tag{E.13}$$

and thus or predicted recursive relationship holds.

E.2 Proof by induction

For the recursive relationship in Equation E.11 to hold, we must prove Equation E.12. We provide the proof using the inductive steps:

1. Provide a base case
2. State the inductive assumption
3. State the inductive step
4. Proof by induction.

E.2.1 Provide the base case:

Equation E.4 is the basis such that,

$$\frac{dG}{dt} = G^{(1)}(t) = c\beta e^{\beta t} G^{(0)}(t). \tag{E.14}$$

E.2.2 State the inductive assumption

Assume,

$$G^{(k)}(t) = c\beta e^{\beta t} \sum_{i=0}^{k-1} \binom{k-1}{i} G^{(i)}(t) \beta^{k-1-i}. \tag{E.15}$$

E.2.3 State the inductive step

We would like to show that the following equation holds,

$$G^{(k+1)}(t) = c\beta e^{\beta t} \sum_{i=0}^{k} \binom{k}{i} G^{(i)}(t) \beta^{k-i}. \tag{E.16}$$

E.2.4 Proof by induction

$$\begin{aligned} G^{(k+1)}(t) &= \frac{d}{dt} G^{(k)}(t) \\ &= c\beta^2 e^{\beta t} \sum_{i=0}^{k-1} \binom{k-1}{i} G^{(i)}(t) \beta^{k-1-i} + c\beta e^{\beta t} \sum_{i=0}^{k-1} \binom{k-1}{i} G^{(i+1)}(t) \beta^{k-1-i}. \end{aligned} \tag{E.17}$$

by the product rule. Grouping terms yields,

$$G^{(k+1)}(t) = c\beta e^{\beta t}\Bigg(\sum_{i=0}^{k-1}\binom{k-1}{i}G^{(i)}(t)\beta^{k-i} + \sum_{i=0}^{k-1}\binom{k-1}{i}G^{(i+1)}(t)\beta^{k-1-i}\Bigg). \tag{E.18}$$

Expanding the equation yields,

$$\begin{aligned} G^{(k+1)}(t) =& c\beta e^{\beta t}\Bigg(\binom{k-1}{0}G^{(0)}(t)\beta^{k} \\ &+ \sum_{i=1}^{k-1}\binom{k-1}{i}G^{(i)}(t)\beta^{k-i} \\ &+ \sum_{i=0}^{k-2}\binom{k-1}{i}G^{(i+1)}(t)\beta^{k-1-i} \\ &+ \binom{k-1}{k-1}G^{(k)}(t)\beta^{0}\Bigg). \end{aligned} \tag{E.19}$$

Changing the index of the right summation ($j = i + 1$),

$$\begin{aligned} G^{(k+1)}(t) =& c\beta e^{\beta t}\Bigg(\binom{k-1}{0}G^{(0)}(t)\beta^{k} \\ &+ \sum_{i=1}^{k-1}\binom{k-1}{i}G^{(i)}(t)\beta^{k-i} \\ &+ \sum_{j=1}^{k-1}\binom{k-1}{j-1}G^{(j)}(t)\beta^{k-j} \\ &+ \binom{k-1}{k-1}G^{(k)}(t)\beta^{0}\Bigg). \end{aligned} \tag{E.20}$$

Changing to common index and merging the sums results in,

Merging the sums,

$$\begin{aligned} G^{(k+1)}(t) =& c\beta e^{\beta t}\Bigg(\binom{k-1}{0}G^{(0)}(t)\beta^{k} \\ &+ \sum_{i=1}^{k-1}\left\{\binom{k-1}{i} + \binom{k-1}{i-1}\right\}G^{(i)}(t)\beta^{k-i} \\ &+ \binom{k-1}{k-1}G^{(k)}(t)\beta^{0}\Bigg). \end{aligned} \tag{E.21}$$

NOTE:

$$\begin{aligned}&\binom{k-1}{i}+\binom{k-1}{i-1}=\frac{(k-1)!}{i!(k-1-i)!}+\frac{(k-1)!}{(i-1)!(k-i)!}=\frac{(k-1)!(k-1)}{i!(k-i)!}+\frac{(k-1)!i}{i!(k-i)!}\\&=\frac{(k-1)!(k-1)+(k-1)!i!}{i!(k-i)!}=\frac{(k-1)!(k-i+i)}{i!(k-i)!}\\&=\frac{k!}{i!(k-i)!}=\binom{k}{i}\end{aligned} \tag{E.22}$$

Thus

$$\begin{aligned}G^{(k+1)}(t)=&c\beta e^{\beta t}\Bigg(\binom{k-1}{0}G^{(0)}(t)\beta^k\\&+\sum_{i=1}^{k-1}\binom{k}{i}G^{(i)}(t)\beta^{k-i}\\&+\binom{k-1}{k-1}G^{(k)}(t)\beta^0\Bigg).\end{aligned} \tag{E.23}$$

Merging the terms results in,

$$G^{(k+1)}(t)=c\beta e^{\beta t}\left(\sum_{i=0}^{k}\binom{k}{i}G^{(i)}(t)\beta^{k-i}\right). \tag{E.24}$$

By induction, Equation E.12 holds for all $n > 0$, q. e. d.

Of all the things I’ve lost,
I miss my mind the most.

–Mark Twain